九州大学学術情報リポジトリ
Kyushu University Institutional Repository

Quantitative and theoretical study on HIV-1 co-infection during cell free infection under heterogeneity of target cell population

伊藤, 悠介

https://hdl.handle.net/2324/4474957

出版情報：Kyushu University, 2020, 博士（理学）, 課程博士
バージョン：
権利関係：

Ph.D. Thesis

Quantitative and theoretical study on HIV-1 co-infection during cell free infection under heterogeneity of target cell population

Yusuke Ito

January, 2021

Preface

Viral infections have suffered from a great number of people so far (Barre-Sinoussi et al., 2013; Scheel and Rice, 2013; Wu et al., 2020). In development of molecular and cellular biology, many aspects of viral infections have been revealed (Dang et al., 2004; Gao et al., 2020; Lohmann et al., 1999). Additionally, to eradicate viral infections, enough data on drug treatment are obtained (Koizumi et al., 2017; Ohashi et al., 2020; Shen et al., 2008). However, there are still some difficulties in understanding, quantifying and predicting viral infections, using only experimental approaches (Dang et al., 2004; Ho et al., 1995).

On the other hand, in the mathematical biology area, lots of theory on viral infections (e.g., virus dynamics) have been developed (Iwami et al., 2013; Nowak and May, 2000; Perelson et al., 1997; Iwami et al., 2017). Furthermore, these theories have been widely applied to quantify and predict the understanding of infectious disease and drug efficacy (Perelson et al., 1997). In addition, the other mathematical biology theories (e.g., negative binomial distribution) were utilized in ecology area and others (Iwasa, 1998).

For the above reasons, in viral infection area, quantitative and theoretical approaches (e.g., virus dynamics) have been expected and contributed to explore the mechanism, dynamics and quantification on viral infections. Thus, I've mainly studied viral infections quantitatively and theoretically.

In the following, I summarize the contents of two chapters as my Ph.D. thesis.

Chapter 1: Number of infection events per cell during HIV-1 cell-free infection

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, I build on these notions and provide a more detailed and extensive quantitative framework. I 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, I took into account the different susceptibility of the target cells as a continuous distribution. Interestingly, I 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. This study was published in Scientific Reports (Ito et al., 2017).

Chapter 2: Dynamics of HIV-1 coinfection in different susceptible target cell populations during cell-free infection

HIV-1 mutations rapidly accumulate through genetic recombination events, which require the infection of a single cell by two virions (coinfection). Accumulation of mutations in the viral population may lead to immune escape and high-level drug resistance. The existence of cell subpopulations characterized by different susceptibility to HIV-1 infection has been proposed as an important parameter driving coinfection (Q. Dang et al., Proc. Natl. Acad. Sci. USA 2004:101(2):632-7). While the mechanism and the quantification of HIV-1 coinfection have been recently investigated by mathematical models, the detailed dynamics of this process during cell-free infection remains elusive. In this study, we constructed ordinary differential equations considering the heterogeneity of target cell populations during cell-free infection in cell culture, and reproduced the cell culture experimental data. Our mathematical analyses showed that the

presence of two differently susceptible target cell subpopulations could explain our experimental datasets, while increasing the number of subpopulations did not improve the fitting. In addition, we quantitatively demonstrated that cells infected by multiple viruses mainly accumulated from one cell subpopulation under cell-free infection conditions. In particular, the frequency of infection events in the more susceptible subpopulation was 6.11-higher than that from the other subpopulation, and 98.3% of coinfected cells emerged from the more susceptible subpopulation. Our mathematical-experimental approach is able to extract such a quantitative information, and can be easily applied to other virus infections. This study was published in Journal of Theoretical Biology (Ito et al., 2018) .

Besides the above main works, I published a study on Simian Human Immunodeficiency Vius as the first author (Hara et al., 2019). I also show the collaborated works I joined. For instance, I've studied and published the other papers related to the above chapters as a co-author: T. Oda et al., 2020 (Oda et al., 2020), K. Kim et al., 2020 (Kim et al., 2020). In addition, I've contributed to collaborated researches on the drug efficacy: H. Ohashi et al., 2020 (Ohashi et al., 2020).

In the end of this preface, this thesis was submitted to the faculty of the Graduate School in partial fulfillment of the requirements for the degree Doctor of Philosophy in Science Kyushu University.

Reference of Preface

Barre–Sinoussi, F., Ross, A.L., and Delfraissy, J.F. (2013). Past, present and future: 30 years of HIV research. Nat Rev Microbiol *11*, 877–883.

Dang, Q., Chen, J.B., Unutmaz, D., Coffin, J.M., Pathak, V.K., Powell, D., KewalRamani, V.N., Maldarelli, F., and Hu, W.S. (2004). Nonrandom HIV–1 infection and double infection via direct and cell–mediated pathways. P Natl Acad Sci USA *101*, 632–637.

Gao, Y., Yan, L., Huang, Y., Liu, F., Zhao, Y., Cao, L., Wang, T., Sun, Q., Ming, Z., Zhang, L., *et al.* (2020).

Structure of the RNA-dependent RNA polymerase from COVID-19 virus. Science *368*, 779-782.

Hara, A., Iwanami, S., Ito, Y., Miura, T., Nakaoka, S., and Iwami, S. (2019). Revealing uninfected and infected target cell dynamics from peripheral blood data in highly and less pathogenic simian/human immunodeficiency virus infected Rhesus macaque. Journal of theoretical biology *479*, 29-36.

Ho, D.D., Neumann, A.U., Perelson, A.S., Chen, W., Leonard, J.M., and Markowitz, M. (1995). Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. Nature *373*, 123-126.

Ito, Y., Remion, A., Tauzin, A., Ejima, K., Nakaoka, S., Iwasa, Y., Iwami, S., and Mammano, F. (2017). Number of infection events per cell during HIV-1 cell-free infection. Sci Rep-Uk *7*.

Ito, Y., Tauzin, A., Remion, A., Ejima, K., Mammano, F., and Iwami, S. (2018). Dynamics of HIV-1 coinfection in different susceptible target cell populations during cell-free infection. Journal of theoretical biology *455*, 39-46.

Iwami, S., Koizumi, Y., Ikeda, H., and Kakizoe, Y. (2013). Quantification of viral infection dynamics in animal experiments. Front Microbiol *4*, 264.

Kim, K.S., Ejima, K., Ito, Y., Iwanami, S., Ohashi, H., Koizumi, Y., Asai, Y., Nakaoka, S., Watashi, K., Thompson, R.N., *et al.* (2020). Modelling SARS-CoV-2 Dynamics: Implications for Therapy. medRxiv, 2020.2003.2023.20040493.

Koizumi, Y., Ohashi, H., Nakajima, S., Tanaka, Y., Wakita, T., Perelson, A.S., Iwami, S., and Watashi, K. (2017). Quantifying antiviral activity optimizes drug combinations against hepatitis C virus infection. Proc Natl Acad Sci U S A *114*, 1922-1927.

Lohmann, V., Korner, F., Koch, J., Herian, U., Theilmann, L., and Bartenschlager, R. (1999). Replication of subgenomic hepatitis C virus RNAs in a hepatoma cell line. Science *285*, 110-113.

Nowak, M., and May, R.M. (2000). Virus dynamics: mathematical principles of immunology and virology: mathematical principles of immunology and virology (Oxford University Press, UK).

Oda, T., Kim, K.S., Fujita, Y., Ito, Y., Miura, T., and Iwami, S. (2020). Quantifying antiviral effects against simian/human immunodeficiency virus induced by host immune response. Journal of theoretical biology *509*, 110493.

Ohashi, H., Watashi, K., Saso, W., Shionoya, K., Iwanami, S., Hirokawa, T., Shirai, T., Kanaya, S., Ito, Y., Kim, K.S., *et al.* (2020). Multidrug treatment with nelfinavir and cepharanthine against COVID-19. bioRxiv, 2020.2004.2014.039925.

Perelson, A.S., Essunger, P., Cao, Y., Vesanen, M., Hurley, A., Saksela, K., Markowitz, M., and Ho, D.D. (1997). Decay characteristics of HIV-1-infected compartments during combination therapy. Nature *387*, 188-191.

Scheel, T.K., and Rice, C.M. (2013). Understanding the hepatitis C virus life cycle paves the way for highly effective therapies. Nat Med *19*, 837-849.

Shen, L., Peterson, S., Sedaghat, A.R., McMahon, M.A., Callender, M., Zhang, H., Zhou, Y., Pitt, E., Anderson, K.S., Acosta, E.P., *et al.* (2008). Dose-response curve slope sets class-specific limits on

inhibitory potential of anti–HIV drugs. Nat Med *14*, 762–766.

Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G., Hu, Y., Tao, Z. W., Tian, J. H., Pei, Y. Y., *et al.* (2020). Author Correction: A new coronavirus associated with human respiratory disease in China. Nature *580*, E7.

Yoh Iwasa (1998). 数理生物学入門: 生物社会のダイナミックスを探る (共立出版).

Shingo Iwami, Kei Sato, and Yasuhiro Takeuchi (2017). ウイルス感染と常微分方程式 (共立出版).

Acknowledgement

I’d like to thank my superviser, Prof. Shingo Iwami (Kyushu University), who supported me and gave many incredible insights. I’m thankful to Prof. Yoh Iwasa (Kwansei-Gakuin University) who were the professor in Kyushu university when I was in master course, and provided me lots of academical comments. Also, I’m grateful to Prof. Akiko Satake (Kyushu University) for very kind supports. I’d like to give special thanks to all senior and junior mathematical biology laboratory members and staffs who kindly supported me so far. In addition, collaborators are always quite helpful to me and I couldn’t complete my works and thesis without their cooperation. Finally, I’d love to thank my family and good company.

Some works in this thesis were financially supported by a Research Fellowship for Young Scientists (DC2) by Japan Society for the Promotion of Science [a] and Qdai-jump Research Program (QR program) of Kyushu University Foundation [b].

Reference

[a] https://kaken.nii.ac.jp/grant/KAKENHI-PROJECT-19J12155/

[b] https://airimaq.kyushu-u.ac.jp/upload_file/editor_files/H30_qr_selected_project.pdf

Contents

Chapter 1

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

The study of this chapter, done in collaboration with Dr Azaria Remion, Dr. Alexandra Tauzin, Dr. Keisuke Ejima, Prof. Shinji Nakaoka, Prof. Yoh Iwasa, Prof. Shingo Iwami and Prof. Fabrizio Mammano, was published in Scientific Reports 7:6559 in July 2017

Introduction

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 (Burke, 1997a)). By mixing the viral genomes, in one step, recombination creates new variants whose adaptation to the environment may exceed those of the parental viruses (Charpentier et al., 2006; Nora et al., 2007). This process could participate in the unfavourable prognosis of patients infected by two strains of HIV, known as double infection (Sagar et al., 2003). Although the majority of HIV-infected lymphocytes in the peripheral blood of patients carry only one viral genome copy (Josefsson et al., 2011; Josefsson et al., 2013), the epidemiologic spread of circulating recombinant forms (CRF) of HIV-1 demonstrates that recombination, and thus double infections, take place in infected patients (Allen and Altfeld, 2003a). It was also shown that recombination in HIV-infected patients may rescue defective viral genomes that carry drug resistance mutations (Donahue et al., 2013; Quan et al., 2009a).

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 (Chen et al., 2007; Jolly and Sattentau, 2004). 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 (Del Portillo et al., 2011). The genomes transmitted by one infected cell via the cell-mediated path- way, 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 (Bregnard et al., 2012; Chen et al., 2005; Dang et al., 2004; Haqqani et al., 2015; Remion et al., 2016). Two previous studies proposed that differences in cell susceptibilities could justify the experimental observation of double infection frequencies that are higher than expected for independent infection events (Dang et al., 2004; Law et al., 2016). 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 (Law et al., 2016). 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 (Dang et al., 2004). Here, by considering susceptibility as a continuous variable, I expand on those original reports, and provide a more detailed quantitative framework. I 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, I show that the number of infection events per cell follows a negative-binomial distribution. I also quantified the increase in the double and multiple infection events as a function of the amount of inoculated virus, and I 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.

This part was done by Ms. Azaria Remion, Mr. Alexandra Tauzin, Prof. Fabrizio Mammano. 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 penicil- lin and 100 µg/ml streptomycin). MT4R5 cells (Amara et al., 2003) 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% CO2. 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 (Imbeault et al., 2009; Levy et al., 2004a). To prevent virus spread in culture, these constructs by deleting 1.3 kb of the env gene between the KpnI and BglII sites were conducted (Remion et al., 2016). To complement these proviral constructs, 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 SRα promoter (SV40-early promoter and LTR from HTLV-I) was used (Remion et al., 2016).

Preparation of virus stocks, infection, and datasets.

This part was performed by Ms. Azaria Remion, Mr. Alexandra Tauzin, Prof. Fabrizio Mammano. Stocks of viruses expressing either GFP or HSA were prepared by transfecting sub-confluent 293-T cells in T75 flasks by JetPei (Polyplus Inc. Illkrich, 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 × 105 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 PerpCy5.5 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 72h 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 Care 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 $e^{-\beta sV}$, the probability of a target cell being infected is $1-e^{-\beta sV}$ (Haqqani et al., 2015). I 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 $\bar{V}_{HSA}$ and $\bar{V}_{GFP}$, 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), I assumed $\bar{V}_{HSA} = V_{HSA}$ and $\bar{V}_{GFP} = V_{GFP}$ for 3.12 µl of inoculated HIV-1 expressing HSA and GFP, respectively, and $\bar{V}_{HSA} = V_{HSA} \times r_*$, and $\bar{V}_{GFP} = V_{GFP} \times r_*$, for the other amounts of inoculated HIV-1. I estimated the distribution of the following 11 parameters (θ): the shape parameter, p, the composed parameters, $\beta V_{HSA}/q$ and $\beta V_{GFP}/q$, 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 $200\mu l$ of the inoculated HIV-1 (Table 1) (i.e., $\theta = \{p, \beta V_{HSA}/q, \beta V_{GFP}/$

$q, r_{6.25},\ r_{12.5}, r_{25}, r_{37},\ r_{50}, r_{100}, r_{200}\}$). 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:

$$F_{A,co}(\theta) = \int_0^\infty \left(1 - e^{-\beta s \bar{V}_{HSA}}\right) e^{-\beta s \bar{V}_{GFP}} \frac{q^p s^{p-1}}{\Gamma(p)} e^{-qs} ds,$$

$$F_{B,co}(\theta) = \int_0^\infty \left(1 - e^{-\beta s \bar{V}_{HSA}}\right)\left(1 - e^{-\beta s \bar{V}_{GFP}}\right) \frac{q^p s^{p-1}}{\Gamma(p)} e^{-qs} ds,$$

$$F_{C,co}(\theta) = \int_0^\infty e^{-\beta s \bar{V}_{HSA}}\, e^{-\beta s \bar{V}_{GFP}} \frac{q^p s^{p-1}}{\Gamma(p)} e^{-qs} ds,$$

$$F_{D,co}(\theta) = \int_0^\infty e^{-\beta s \bar{V}_{HSA}} \left(1 - e^{-\beta s \bar{V}_{GFP}}\right) \frac{q^p s^{p-1}}{\Gamma(p)} e^{-qs} ds.$$

I further calculated and simplified those equations as follows:

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

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

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

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

Notably, in the experiments with only HSA-expressing HIV-1 (i.e., single HSA HIV-1 experiments), I 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, I 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, I employed likelihood estimation to obtain an optimal set of parameter values. If I assume that the datasets used follow the Gaussian distribution with the mean μ_i and the variance σ^2(i.e., $N(\mu_i, \sigma^2)$, the corresponding likelihood function is given as follows:

$$L(\boldsymbol{\rho}; \boldsymbol{x}) = \prod_{i=1}^{N_{data}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\mu_i - x_i)^2}{2\sigma^2}} = \left(\prod_{i=1}^{N_{data}} \frac{1}{\sqrt{2\pi\sigma^2}}\right) e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{N_{data}}(\mu_i - x_i)^2}.$$

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

$$\mathrm{SSR}(\boldsymbol{\theta}) = \sum_{r=1}^{10}\left\{\left(F_{A,r} - \tilde{F}_{A,r}\right)^2 + \left(F_{C,r} - \tilde{F}_{C,r}\right)^2\right\} + \sum_{g=1}^{10}\left\{\left(F_{C,g} - \tilde{F}_{C,g}\right)^2 + \left(F_{D,g} - \tilde{F}_{D,g}\right)^2\right\}$$

$$+ \sum_{co=1}^{18}\left\{\left(F_{A,co} - \tilde{F}_{A,co}\right)^2 + \left(F_{B,co} - \tilde{F}_{B,co}\right)^2 + \left(F_{C,co} - \tilde{F}_{C,co}\right)^2 + \left(F_{D,co} - \tilde{F}_{D,co}\right)^2\right\}.$$

Here, θ 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., $r = 1, \ldots, 10$ correspond to 3.12, 6.25, 12.5, 25, 37, 50, 75, 100, 100 and 200 µl, respectively, of single HSA HIV-1 experiments; $g = 1, \ldots, 10$ correspond to 3.12, 6.25, 12.5, 25, 50, 50, 75, 100, 100 and 200 µl, respectively, of single GFP HIV-1 experiments; $co = 1, \ldots, 18$ correspond to 25 and 25, 25 and 50, 50 and 25, 37 and 50, 50 and 50, 37 and 75, 75 and 50, 75 and 75, 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 µl and 100 and 100 µl, respectively, of double HSA and GFP HIV-1 experiments). If the variance σ^2 is constant, then likelihood estimation is completely determined by the SSR between theoretical values and measurements (i.e., $\sum(\boldsymbol{\mu} - \boldsymbol{x})^2$). However, to take into account for possible variations, I 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 per- form MCMC sampling by "delayed rejection and adaptive Metropolis algorithm" (Soetaert and Petzoldt, 2010). 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 σ^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 (Soetaert and Petzoldt, 2010). 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 (Bregnard et al., 2012; Dang et al., 2004; Del Portillo et al., 2011; Haqqani et al., 2015; Remion et al., 2016). Here, I developed a novel mathematical model explicitly considering the heterogeneity of target cells susceptibility to infection. In our model, I define the following parameters: V is the amount of effective virus for infection events, β 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 (Bregnard et al., 2012; Del Portillo et al., 2011; Haqqani et al., 2015):

$$f(X = n\,;\,\beta sV) = \frac{(\beta sV)^n\, e^{-\beta sV}}{n!}. \tag{5}$$

Additionally, to consider the heterogeneity of target cell susceptibility (Dang et al., 2004; Remion et al., 2016), I assumed that the susceptibility parameter, s, obeys the following Gamma distribution (MacDonald, 2008):

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

It is well known that Gamma distribution can approximate any one-peak distribution and reproduce a variety of biological phenomena (Bliss and Fisher, 1953; Kakizoe et al., 2015; MacDonald, 2008). Here $p > 0$ and $q > 0$ are the shape and rate parameters, and $\Gamma(*)$ is the gamma function. In a previous study (Dang et al., 2004), 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). I 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 (Chen et al., 2005; Kabat et al., 1994; Platt et al., 1998). Using Eqs (5, 6), I calculated the probability density function for a cell being infected by n HIV-1 in a heterogeneous target cell population:

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

$$= (\beta V)^n q^p \frac{1}{n!\,\Gamma(p)} \int_0^\infty s^{n+p-1}\, e^{-(\beta V+q)s} ds,$$

$$= \frac{\Gamma(n+p)}{n!\,\Gamma(p)} \left(\frac{q}{\beta V+q}\right)^p \left(\frac{\beta V}{\beta V+q}\right)^n,$$

$$= \binom{n+p-1}{n} \left(\frac{1}{1+\frac{\beta V}{q}}\right)^p \left(\frac{\frac{\beta V}{q}}{1+\frac{\beta V}{q}}\right)^n . \quad (7)$$

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 $\beta pV/q$ and variance $\beta pV/q(1 + \beta V/q)$.

I estimated the parameters in Eq. (7) by fitting Eqs (1–4) to the experimental measured frequencies of quad- rants 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 μl 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 (Chen et al., 2005; Dang et al., 2004).

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 (Chen et al., 2005; Dang et al., 2004; Remion et al., 2016). The odds of HSA-positive cells being GFP-positive can be calculated by

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

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

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

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 (Chen et al., 2005; Dang et al., 2004; Remion et al., 2016). To study whether or not our novel model quantitatively reproduces this important property of HIV-1 coinfection, I derived the theoretical odds ratio, OR_M, from Eqs (1–4) as follows:

$$OR_M=\frac{F_B(\boldsymbol{\theta})F_C(\boldsymbol{\theta})}{F_A(\boldsymbol{\theta})F_D(\boldsymbol{\theta})}$$

$$=1+\frac{\dfrac{1}{\left(1+\dfrac{\beta}{q}(\bar{V}_{\mathrm{HSA}}+\bar{V}_{\mathrm{GFP}})\right)^p}-\dfrac{1}{\left(1+\dfrac{\beta}{q}\bar{V}_{\mathrm{HSA}}\right)^p\left(1+\dfrac{\beta}{q}\bar{V}_{\mathrm{GFP}}\right)^p}}{\left\{\dfrac{1}{\left(1+\dfrac{\beta}{q}\bar{V}_{\mathrm{GFP}}\right)^p}-\dfrac{1}{\left(1+\dfrac{\beta}{q}(\bar{V}_{\mathrm{HSA}}+\bar{V}_{\mathrm{GFP}})\right)^p}\right\}\left\{\dfrac{1}{\left(1+\dfrac{\beta}{q}\bar{V}_{\mathrm{HSA}}\right)^p}-\dfrac{1}{\left(1+\dfrac{\beta}{q}(\bar{V}_{\mathrm{HSA}}+\bar{V}_{\mathrm{GFP}})\right)^p}\right\}}.$$

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, I 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 (Del Portillo et al., 2011; Haqqani et al., 2015), 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 (Del Portillo et al., 2011). As I derived $F_{quadrant,r,g,co}(r = 1, \dots, 10)(g = 1, \dots, 10)(co = 1, \dots, 18)$ in the MATERIALS AND METHODS, I 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 fol- lowing equation, in which n_{HSA} and n_{GFP} are the number of infection events with HSA and GFP virions, respectively:

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

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 μl 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 pre- viously estimated (21% in ref. (Del Portillo et al., 2011)) 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, I 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 ▲, ◆, ●, and ■ 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 μl 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, I could estimate the mean number of infection events per infected cell during cell-free HIV-1 infection. In Fig. 4c, I 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 observa- tion of proviral copy number in infected cells measured by fluorescence in situ hybridization (Del Portillo et al., 2011).

Discussion

In this study, I 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 (Dang et al., 2004; Law et al., 2016). In those reports, cell susceptibility was either considered as a binary feature (cells were either susceptible or non-susceptible) (Law et al., 2016), or a limited number of subpopulations characterized by distinct susceptibility levels were considered (Dang et al., 2004). Here, by considering susceptibility as a continuous variable, I 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, I 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 (Chen et al., 2005; Dang et al., 2004; Remion et al., 2016). 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 (Kakizoe et al., 2015; Pinilla et al., 2012). 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

I estimated a composed parameter ($\beta\bar{V}_{\mathrm{HSA}}/q$ or $\beta\bar{V}_{\mathrm{GFP}}/q$ in Table 1) to express the effective virus dose for a given volume of different viruses (e.g., 3.12 µl 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, I 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 µl 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 (Bregnard et al., 2012; Chen et al., 2005; Dang et al., 2004; Haqqani et al., 2015; Remion et al., 2016), while the Poisson distribution predicts an $OR_M = 1$. Finally, having derived the values of the relevant parameters, 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 discrimi- nation. In Fig. 4b, I 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 (Del Portillo et al., 2011). 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 (Bregnard et al., 2012). I 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 (Remion et al., 2016).

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+ T cells from the spleen of HIV-infected individuals carry on average 3.2 HIV proviral copies (Dixit and Perelson, 2004; Jung et al., 2002). Also, the number of multiply infected cells was found to correlate with the square of the overall number of infected cells (Dixit and Perelson, 2005; Levy et al., 2004a; Wodarz and Levy, 2011) in homogeneous target cell populations.

In agreement with previous reports (Del Portillo et al., 2011), I 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.

Figure

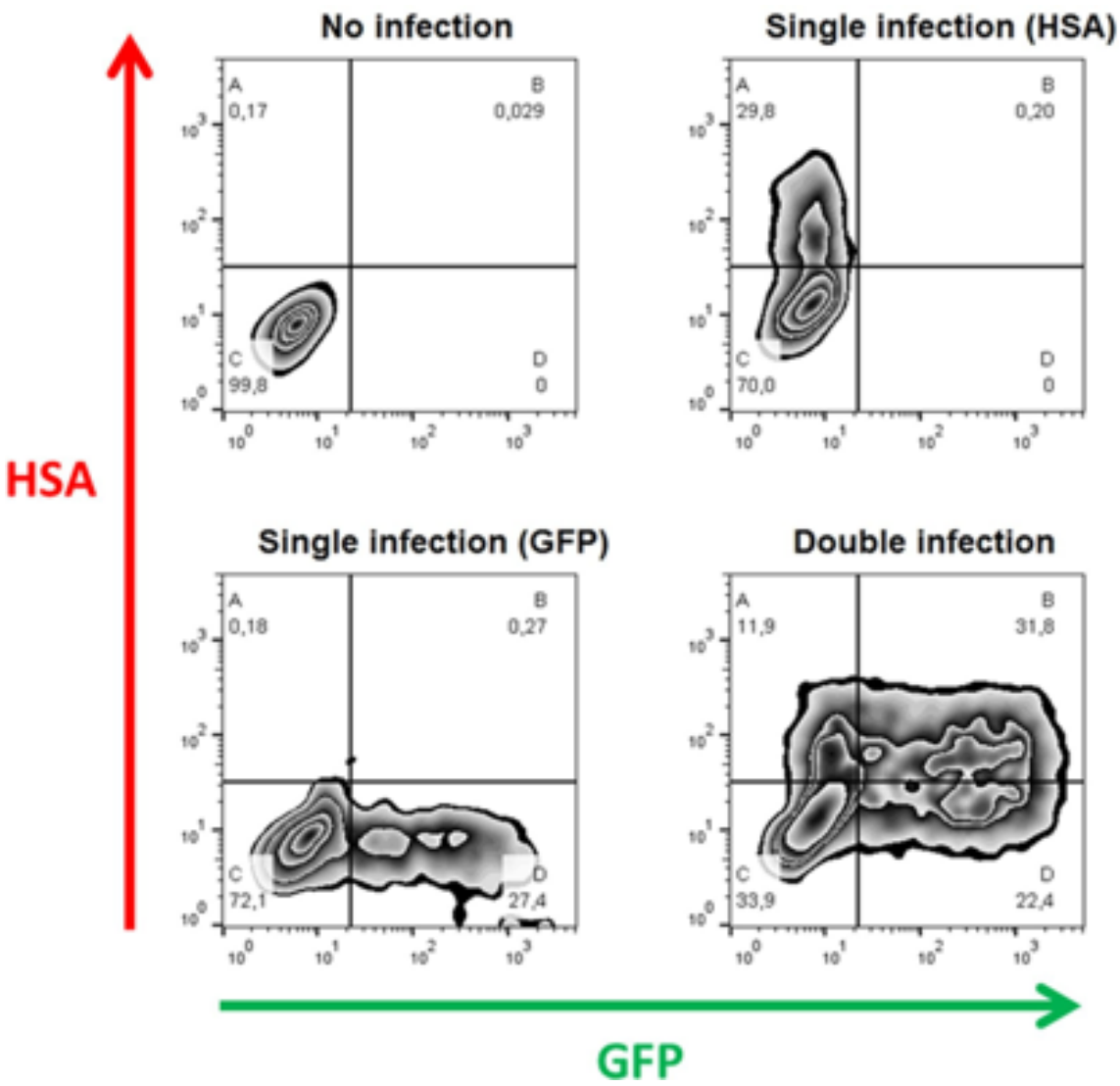


Figure1: Flow cytometry analysis of single and double HIV-1 infection. Panels represent the following conditions, clockwise starting from the upper left panel: no infection; infection by the HSA virus; coinfection with HSA and GFP viruses, infection by the GFP virus. In each panel, the quadrants correspond to HSA+ (A); HSA+GFP+ (B); uninfected cells (C); and GFP+ (D). The percentage of cells in each quadrant is indicated under the letter identifying the quadrant.

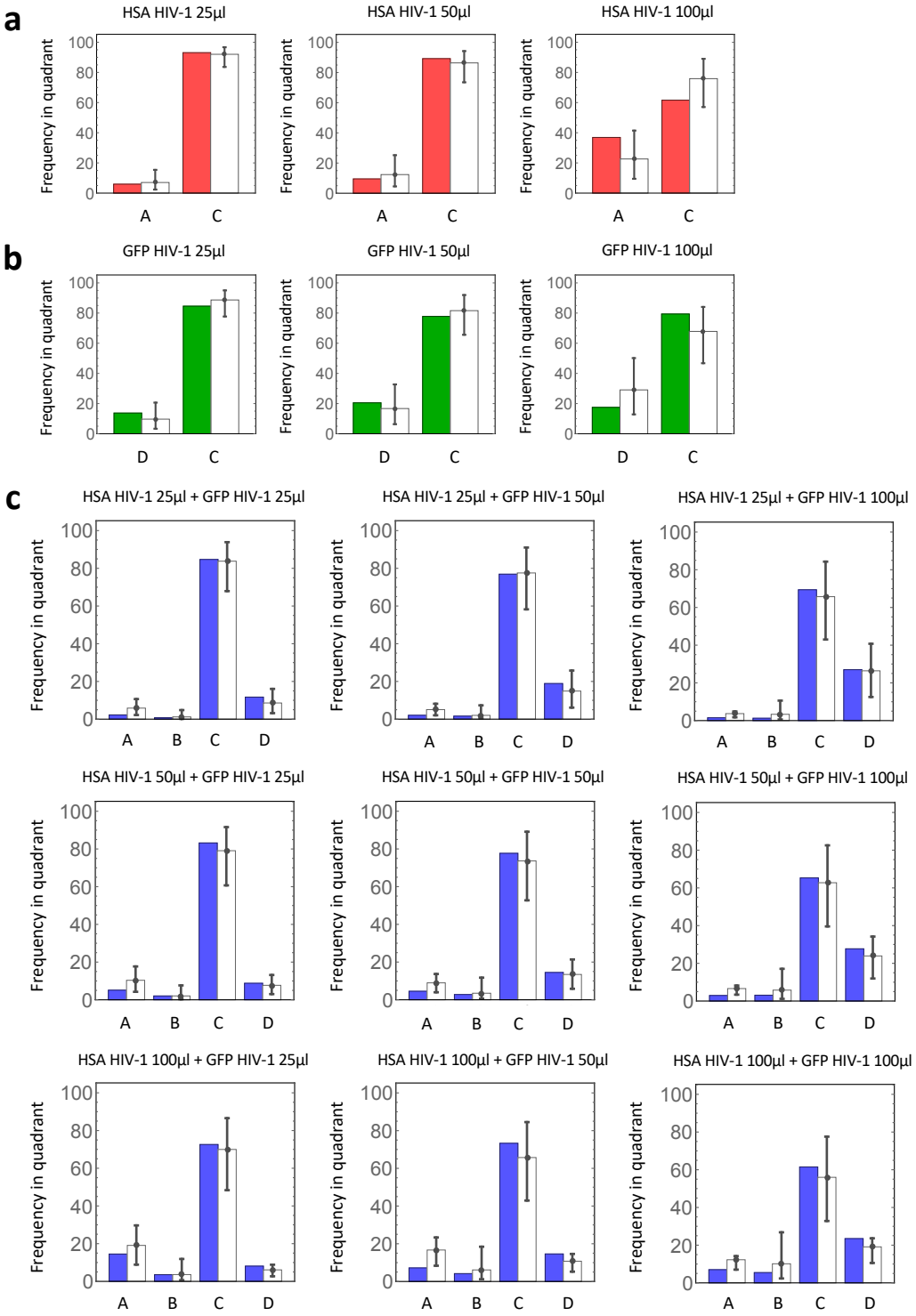


Figure 2. Frequency of single infection and coinfection: (**a**) The experimental and theoretical frequencies of quadrants A (i.e., HSA-positive) and C (i.e., HSA-negative) in three independent experiments using only HSA HIV-1 are shown by red and white bars, respectively. (**b**) The experimental and 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).

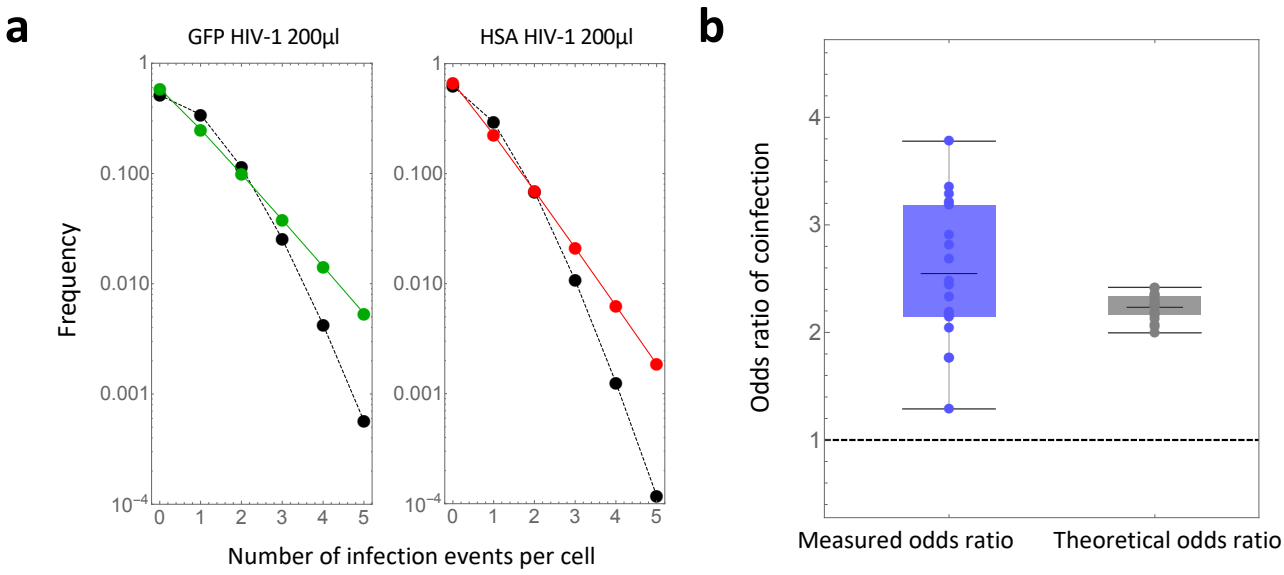


Figure 3. Frequency of multiple infection events per cell: (**a**) The expected negative-binomial distributions of the number of infection events per cell in 200 µl in GFP and HSA HIV-1 single experiments are shown in green and red curves, respectively. These curves were drawn using the mean derived from MCMC parameter inferences (Table 1). The black curves represent the expected Poisson distribution with the mean of the Gamma-distributed susceptibility parameter, (i.e., the target cell population is assumed to have homogeneous susceptibility). (**b**) The experimental odds ratio (ORE) and theoretical odds ratio (ORM, 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., ORM = 1: random HIV-1 infection).

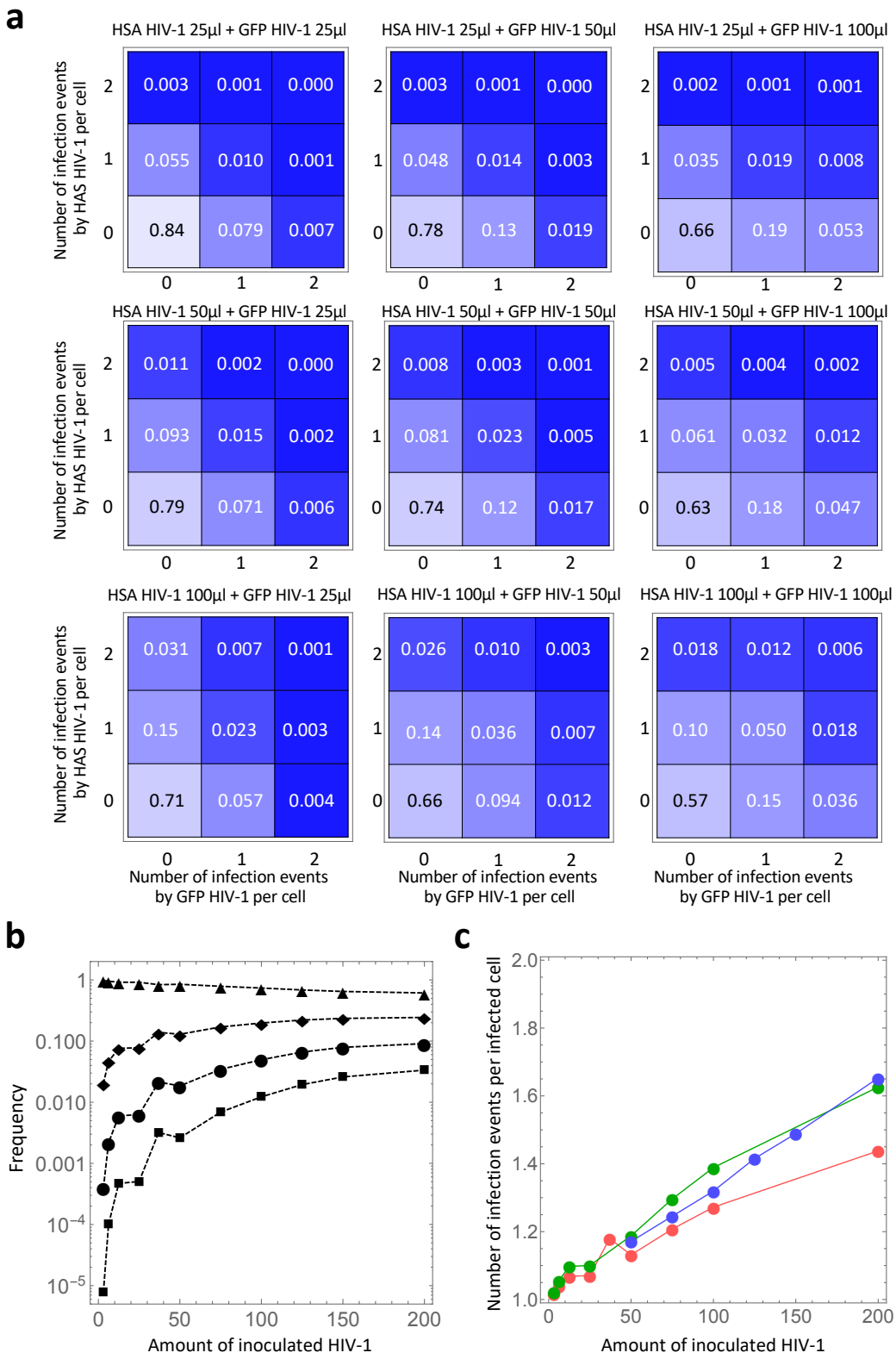


Figure 4. Quantitative analyses of multiple infection: (**a**) The distribution of the number of infection events per cell in double HIV-1 infection experiments with nine different combinations of virus amounts are shown. The number in each square is the estimated frequency of the corresponding infection 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 ▲, ◆, ●, and ■ 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).

Table

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

Parameters	Estimated values (mean)	95%CI
p	1.176	0.878 - 1.512
$\beta\bar{V}_{\mathrm{HSA}}/q$	1.484×10^{-2}	1.034×10^{-2} - 1.963×10^{-2}
$\beta\bar{V}_{\mathrm{GFP}}/q$	2.115×10^{-2}	1.486×10^{-2} - 2.796×10^{-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

Supplementary Note

Calculation of theoretical odds ratio

I derived the theoretical odds ratio, OR_M, by Eqs. (1–4) as follows:

$$OR_M = \frac{F_B(\boldsymbol{\theta})F_C(\boldsymbol{\theta})}{F_A(\boldsymbol{\theta})F_D(\boldsymbol{\theta})}$$

$$= 1 + \frac{\dfrac{1}{\left(1+\frac{\beta}{q}(\bar{V}_{\mathrm{HSA}}+\bar{V}_{\mathrm{GFP}})\right)^{p}} - \dfrac{1}{\left(1+\frac{\beta}{q}\bar{V}_{\mathrm{HSA}}\right)^{p}\left(1+\frac{\beta}{q}\bar{V}_{\mathrm{GFP}}\right)^{p}}}{\left\{\dfrac{1}{\left(1+\frac{\beta}{q}\bar{V}_{\mathrm{GFP}}\right)^{p}} - \dfrac{1}{\left(1+\frac{\beta}{q}(\bar{V}_{\mathrm{HSA}}+\bar{V}_{\mathrm{GFP}})\right)^{p}}\right\}\left\{\dfrac{1}{\left(1+\frac{\beta}{q}\bar{V}_{\mathrm{HSA}}\right)^{p}} - \dfrac{1}{\left(1+\frac{\beta}{q}(\bar{V}_{\mathrm{HSA}}+\bar{V}_{\mathrm{GFP}})\right)^{p}}\right\}}.$$

Because $\beta > 0,\ p > 0,\ 1/q > 0,\ \bar{V}_{\mathrm{HSA}} > 0,\ \bar{V}_{\mathrm{GFP}} > 0$ and

$$0 < 1 + \frac{\beta}{q}(\bar{V}_{\mathrm{HSA}} + \bar{V}_{\mathrm{GFP}}) < \left(1+\frac{\beta}{q}\bar{V}_{\mathrm{HSA}}\right)\left(1+\frac{\beta}{q}\bar{V}_{\mathrm{GFP}}\right),$$

I obtained the following relationship of

$$0 < \left(1+\frac{\beta}{q}(\bar{V}_{\mathrm{HSA}}+\bar{V}_{\mathrm{GFP}})\right)^{p} < \left(1+\frac{\beta}{q}\bar{V}_{\mathrm{HSA}}\right)^{p}\left(1+\frac{\beta}{q}\bar{V}_{\mathrm{GFP}}\right)^{p}.$$

Thus, it is shown that

$$\frac{1}{\left(1+\frac{\beta}{q}(\bar{V}_{\mathrm{HSA}}+\bar{V}_{\mathrm{GFP}})\right)^{p}} - \frac{1}{\left(1+\frac{\beta}{q}\bar{V}_{\mathrm{HSA}}\right)^{p}\left(1+\frac{\beta}{q}\bar{V}_{\mathrm{GFP}}\right)^{p}} > 0\,.$$

Hence, the numerator in the second term of OR_M is always positive. Similarly, since

$$0 < 1 + \frac{\beta}{q}\bar{V}_{\mathrm{GFP}} < 1 + \frac{\beta}{q}(\bar{V}_{\mathrm{HSA}} + \bar{V}_{\mathrm{GFP}}),$$

I obtained

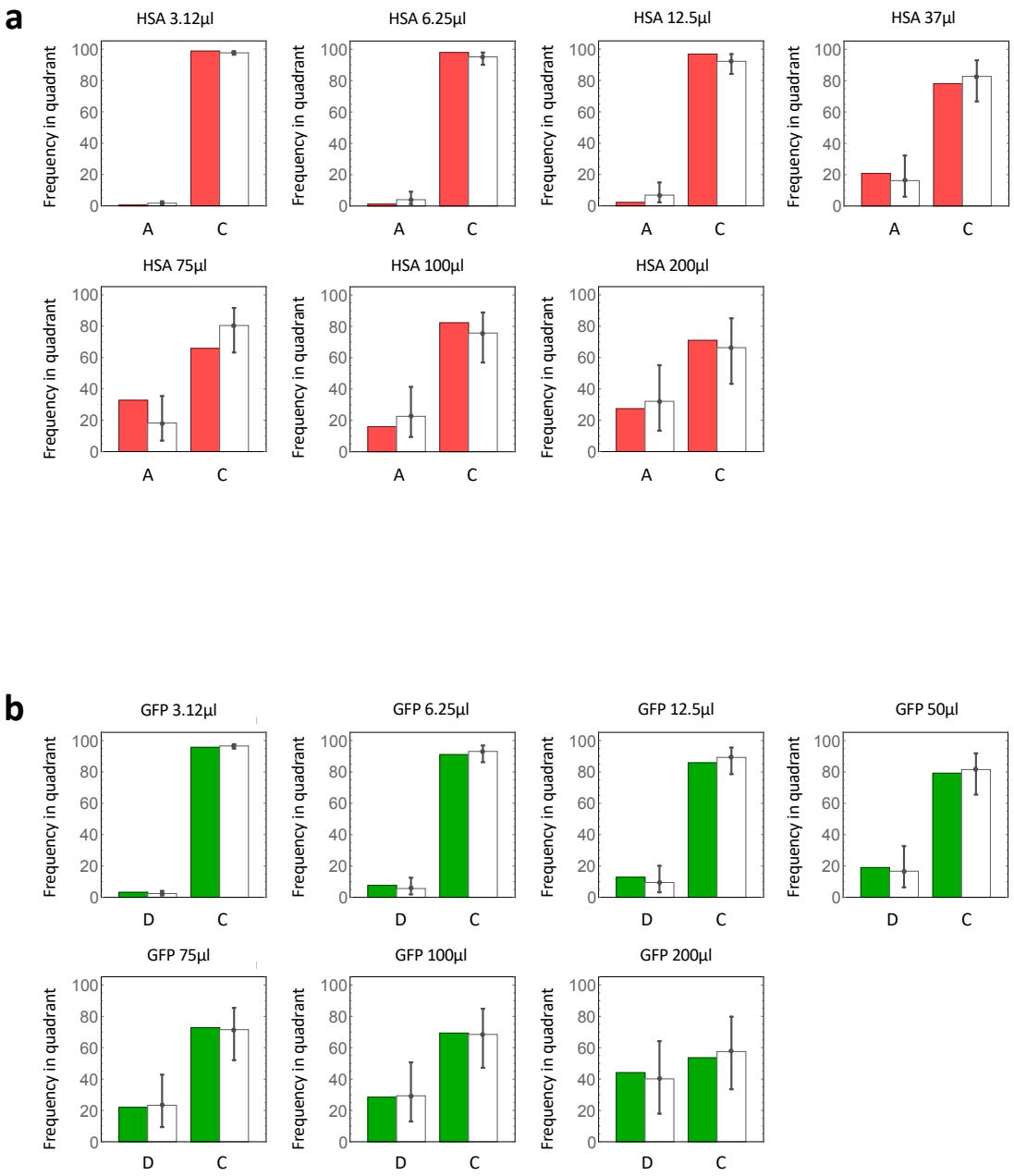
$$0 < \left(1+\frac{\beta}{q}\bar{V}_{\mathrm{GFP}}\right)^{p} < \left(1+\frac{\beta}{q}(\bar{V}_{\mathrm{HSA}}+\bar{V}_{\mathrm{GFP}})\right)^{p},$$

this leads to

$$\frac{1}{\left(1+\frac{\beta}{q}\bar{V}_{\mathrm{GFP}}\right)^{p}}-\frac{1}{\left(1+\frac{\beta}{q}(\bar{V}_{\mathrm{HSA}}+\bar{V}_{\mathrm{GFP}})\right)^{p}}>0.$$

This inequality explains why the denominator in the second term of OR_M is always positive. Therefore, I was able to show that $OR_M > 1$ holds.

Supplementary Figure



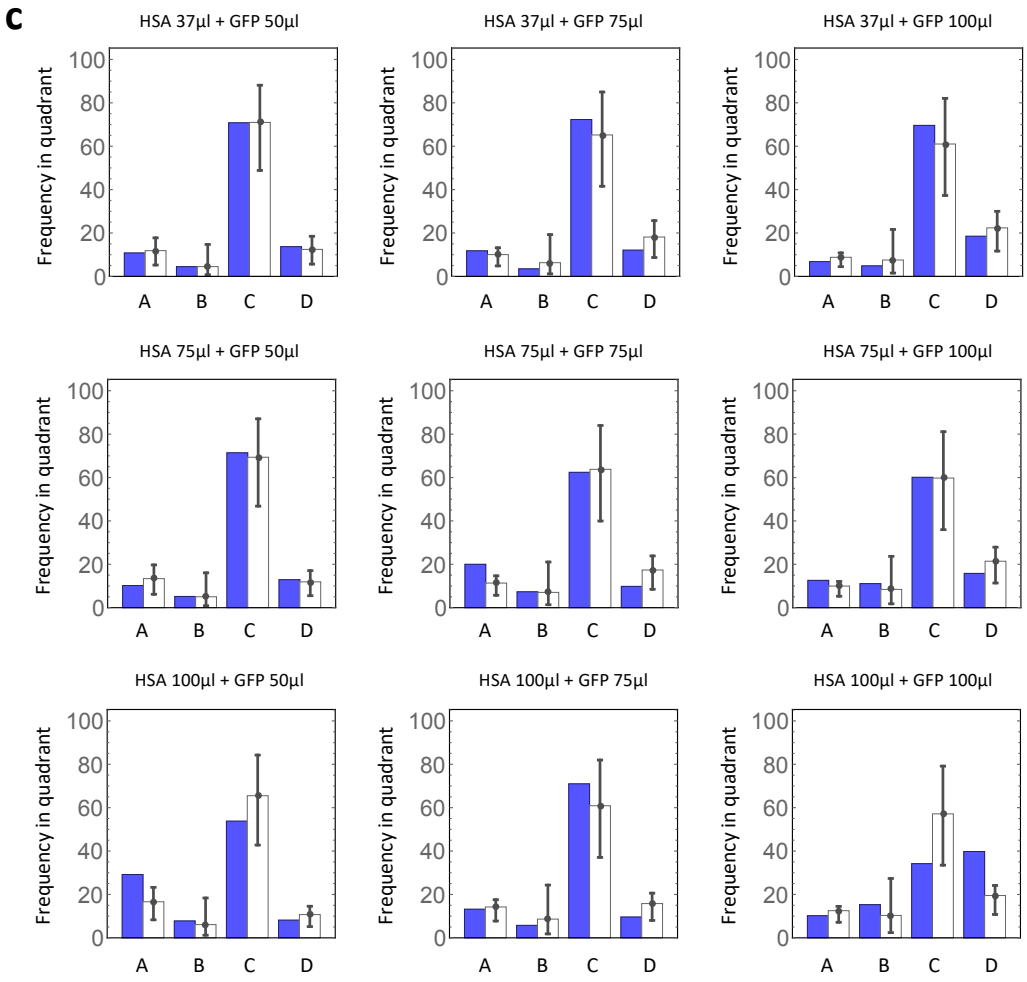
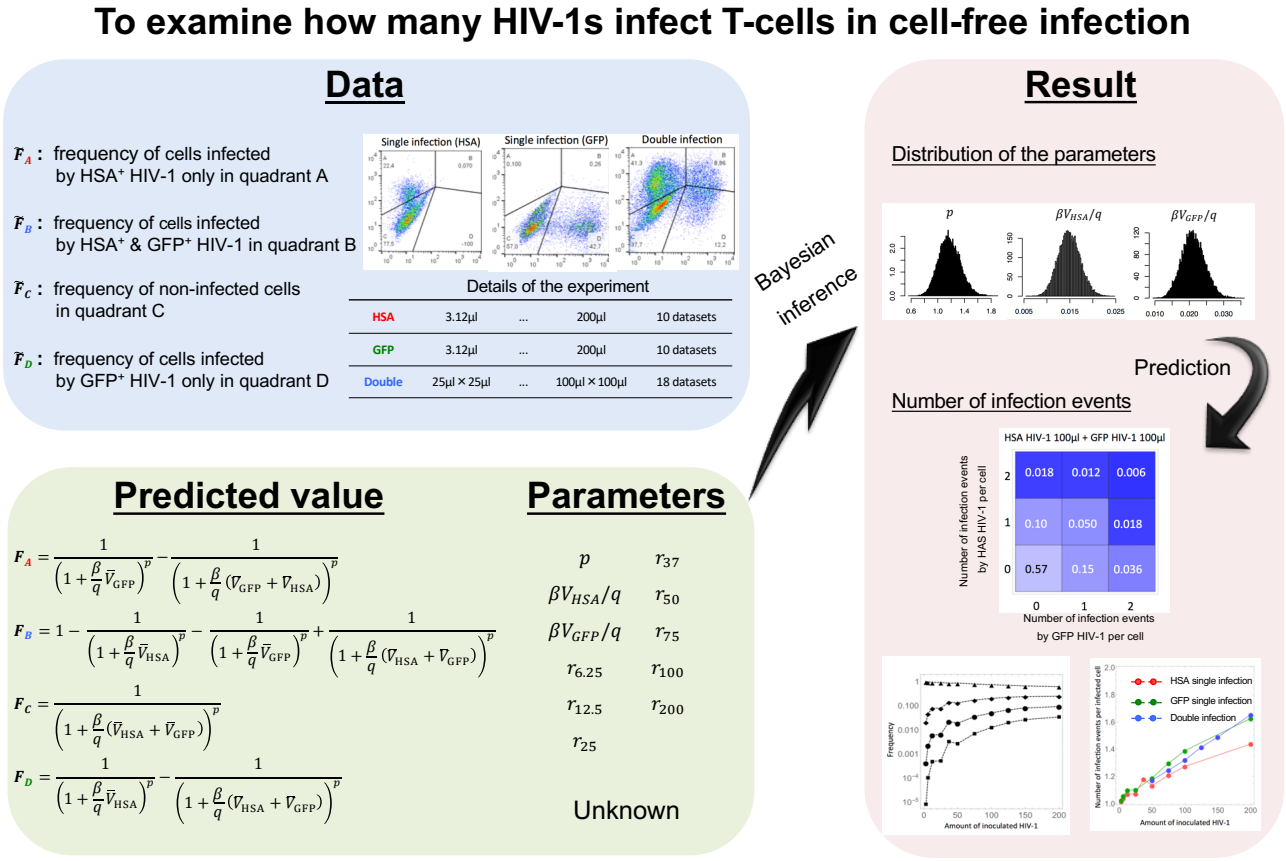


Figure S1 | Frequency of single infection and coinfection in other experiment: (**a**) The experimental and theoretical frequencies of quadrants A (i.e., HSA-positive) and C (i.e., HSA- negative) in seven independent experiments using only HSA HIV-1 are shown by red and white bars, respectively. (**b**) 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, B (i.e., positive both for HSA and GFP), C (i.e., negative both for HSA and GFP), and D in double HIV-1 experiments are shown by blue and white bars, respectively, with different combinations of inoculated viral amount. Note that each error bar represents 95% credible interval obtained from Markov Chain Monte Carlo(MCMC) parameter inferences.



FigureS2 | The outline of our study: (**Blue area**) The frequency of cells in each quadrant were taken by FACS analyses in 3 independent experiments. (**Green area**) I calculated the predicted frequencies corresponding to the measurements, and the 11 parameters are unknown. (**Red area**) To estimate the distribution of the all parameters, the bayesian method are applied. Using the values, the calculations on the number of infection events were done(see Fig4).

Reference

Allen, T.M., and Altfeld, M. (2003a). HIV-1 superinfection. J Allergy Clin Immun *112*, 829-835.

Amara, A., Vidy, A., Boulla, G., Mollier, K., Garcia-Perez, J., Alcami, J., Blanpain, C., Parmentier, M., Virelizier, J.L., Charneau, P., *et al.* (2003). G protein-dependent CCR5 signaling is not required for efficient infection of primary T lymphocytes and macrophages by R5 human immunodeficiency virus type I isolates. J Virol *77*, 2550-2558.

Bliss, C.I., and Fisher, R.A. (1953). Fitting the Negative Binomial Distribution to Biological Data - Note on the Efficient Fitting of the Negative Binomial. Biometrics *9*, 176-200.

Bregnard, C., Pacini, G., Danos, O., and Basmaciogullari, S. (2012). Suboptimal Provirus Expression Explains Apparent Nonrandom Cell Coinfection with HIV-1. J Virol *86*, 8810-8820.

Burke, D.S. (1997a). Recombination in HIV: an important viral evolutionary strategy. Emerg Infect Dis *3*, 253-259.

Charpentier, C., Nora, T., Tenaillon, O., Clavel, F., and Hance, A.J. (2006). Extensive recombination among human immunodeficiency virus type 1 quasispecies makes an important contribution to viral diversity in individual patients. J Virol *80*, 2472-2482.

Chen, J.B., Dang, Q., Unutmaz, D., Pathak, V.K., Maldarelli, F., Powell, D., and Hu, W.S. (2005). 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.

Chen, P., Hubner, W., Spinelli, M.A., and Chen, B.K. (2007). 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.

Dang, Q., Chen, J.B., Unutmaz, D., Coffin, J.M., Pathak, V.K., Powell, D., KewalRamani, V.N., Maldarelli, F., and Hu, W.S. (2004). Nonrandom HIV-1 infection and double infection via direct and cell-mediated pathways. P Natl Acad Sci USA *101*, 632-637.

Del Portillo, A., Tripodi, J., Najfeld, V., Wodarz, D., Levy, D.N., and Chen, B.K. (2011). Multiploid Inheritance of HIV-1 during Cell-to-Cell Infection. J Virol *85*, 7169-7176.

Dixit, N.M., and Perelson, A.S. (2004). Multiplicity of human immunodeficiency virus infections in lymphoid tissue. J Virol *78*, 8942-8945.

Dixit, N.M., and Perelson, A.S. (2005). HIV dynamics with multiple infections of target cells. P Natl Acad Sci USA *102*, 8198-8203.

Donahue, D.A., Bastarache, S.M., Sloan, R.D., and Wainberg, M.A. (2013). 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.

Haqqani, A.A., Marek, S.L., Kumar, J., Davenport, M., Wang, H., and Tilton, J.C. (2015). Central memory CD4+T cells are preferential targets of double infection by HIV-1. Virol J *12*.

Imbeault, M., Lodge, R., Ouellet, M., and Tremblay, M.J. (2009). 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.

Jolly, C., and Sattentau, Q.J. (2004). Retroviral spread by induction of virological synapses. Traffic *5*, 643-650.

Josefsson, L., King, M.S., Makitalo, B., Brannstrom, J., Shao, W., Maldarelli, F., Kearney, M.F., Hu, W.S., Chen, J.B., Gaines, H., *et al.* (2011). Majority of CD4(+) T cells from peripheral blood of HIV-1-infected individuals contain only one HIV DNA molecule. P Natl Acad Sci USA *108*, 11199-11204.

Josefsson, L., Palmer, S., Faria, N.R., Lemey, P., Casazza, J., Ambrozak, D., Kearney, M., Shao, W., Kottilil, S., Sneller, M., *et al.* (2013). Single Cell Analysis of Lymph Node Tissue from HIV-1 Infected Patients Reveals that the Majority of CD4(+) T-cells Contain One HIV-1 DNA Molecule. Plos Pathog *9*.

Jung, A., Maier, R., Vartanian, J.P., Bocharov, G., Jung, V., Fischer, U., Meese, E., Wain-Hobson, S., and Meyerhans, A. (2002). Recombination - Multiply infected spleen cells in HIV patients. Nature *418*, 144-144.

Kabat, D., Kozak, S.L., Wehrly, K., and Chesebro, B. (1994). Differences in Cd4 Dependence for Infectivity of Laboratory-Adapted and Primary Patient Isolates of Human-Immunodeficiency-Virus Type-1. J Virol *68*, 2570-2577.

Kakizoe, Y., Nakaoka, S., Beauchemin, C.A.A., Morita, S., Mori, H., Igarashi, T., Aihara, K., Miura, T., and Iwami, S. (2015). A method to determine the duration of the eclipse phase for in vitro infection with a highly pathogenic SHIV strain. Sci Rep-Uk *5*.

Law, K.M., Komarova, N.L., Yewdall, A.W., Lee, R.K., Herrera, O.L., Wodarz, D., and Chen, B.K. (2016). In Vivo HIV-1 Cell-to-Cell Transmission Promotes Multicopy Micro-compartmentalized Infection. Cell Rep *15*, 2771-2783.

Levy, D.N., Aldrovandi, G.M., Kutsch, O., and Shaw, G.M. (2004a). Dynamics of HIV-1 recombination in its natural target cells. P Natl Acad Sci USA *101*, 4204-4209.

MacDonald, N., Cannings, C. & Hoppensteadt, F. C (2008). Biological delay systems: linear stability theory. Cambridge University Press.

Nora, T., Charpentier, C., Tenaillon, O., Hoede, C., Clavel, F., and Hance, A.J. (2007). Contribution of recombination to the evolution of human immunodeficiency viruses expressing resistance to antiretroviral treatment. J Virol *81*, 7620-7628.

Pinilla, L.T., Holder, B.P., Abed, Y., Boivin, G., and Beauchemin, C.A.A. (2012). 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.

Platt, E.J., Wehrly, K., Kuhmann, S.E., Chesebro, B., and Kabat, D. (1998). Effects of CCR5 and CD4 cell surface concentrations on infections by macrophagetropic isolates of human immunodeficiency virus type 1. J Virol *72*, 2855-2864.

Quan, Y., Liang, C., Brenner, B.G., and Wainberg, M.A. (2009a). 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.

Remion, A., Delord, M., Hance, A.J., Saragosti, S., and Mammano, F. (2016). Kinetics of the establishment of HIV-1 viral interference and comprehensive analysis of the contribution of viral genes. Virology *487*, 59-67.

Sagar, M., Lavreys, L., Baeten, J.M., Richardson, B.A., Mandaliya, K., Chohan, B.H., Kreiss, J.K., and Overbaugh, J. (2003). Infection with multiple human immunodeficiency virus type 1 variants is associated with faster disease progression. J Virol *77*, 12921-12926.

Soetaert, K., and Petzoldt, T. (2010). Inverse Modelling, Sensitivity and Monte Carlo Analysis in R Using Package

FME. J Stat Softw *33*.

Wodarz, D., and Levy, D.N. (2011). 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.

Chapter 2

Dynamics of HIV-1 coinfection

in different susceptible target cell populations during cell-free infection

The study of this chapter, done in collaboration with Dr. Alexandra Tauzin, Dr Azaria Remion, Dr. Keisuke Ejima, Prof. Shingo Iwami and Prof. Fabrizio Mammano, was published in Journal of Theoretical Biology 455: 39-46, October 2018

Introduction

Human Immunodeficiency Virus type-I (HIV-1) infection remains one of the leading public-health concerns since the early 1980s (Barre-Sinoussi et al., 2013; Fauci, 2003). The major hurdles to the control of HIV-1 infection are viral escape from immune responses and high-level drug resistance to antiretroviral therapy (Chang et al., 2013). These phenomena are caused by the rapid production and accumulations of HIV-1 mutations over time in HIV-1 patients, driving viral evolution and escape (Gottlieb et al., 2004; Nora et al., 2007). One of the most efficient mechanisms favoring the accumulations of mutation consists in the recombination events, which allow to combine in a single step portions of the viral genomes that evolved independently (Burke, 1997b; Jung et al., 2002). Recombination events require the infection of a single cell by two virions. Double infections can take place within a short time frame (coinfection) or with variable delay (super-infection) (Law et al., 2016; Remion et al., 2016). Double-infected cells produce virus particles that may carry two different RNA molecules, which can recombine during reverse transcription and produce a chimeric genome (Redd et al., 2013).

Genetic recombination is a relevant phenomenon in infected patients, and several reports have described patients infected by multiple HIV-1 variants (Allen and Altfeld, 2003b; Powell et al., 2009; Ronen et al., 2013; Smith et al., 2005). Although most infected circulating cells in patients harbor a single viral genome, double-infected cells are consistently detected (Josefsson et al., 2011; Josefsson et al., 2013). Genetic recombination in coinfected cells constantly happens over time (Fraser, 2005; Simon-Loriere and Holmes, 2011), and double infections by HIV-1 was shown to lead to the emergence of multidrug resistant viruses (Donahue et al., 2013; Quan et al., 2009b).

Interestingly, it was previously shown in tissue culture that HIV-1 coinfection is not a random event, and it takes place with higher frequency than expected for independent events (Dang et al., 2004). Dang et al., suggested that the heterogeneity of target cells in terms of susceptibility to infection, is largely responsible for the nonrandom distribution of coinfection events *in vitro* (Dang et al., 2004). In their study, for simplicity, five subpopulations with a gradient of susceptibility to the infection were considered (Dang et al., 2004). The concept

of heterogeneity of target cells has been widely supported by subsequent publications (Chen et al., 2005; Ito et al., 2017; Remion et al., 2016; van der Kuyl and Cornelissen, 2007). One aspect that has not been sufficiently investigated, is the time-evolution of coinfection. To understand such non-random multiple infection events from the point of view of "virus dynamics", the kinetics of HIV-1 coinfection during the course of the infection should be quantified. In a recent report by Law et al., (Law et al., 2016), a simple ordinary differential equation model has been proposed and the frequency of HIV-1 coinfection *in vivo* was investigated. Their mathematical analysis showed that HIV-1 coinfection is indeed non-random, and target cell population must have different susceptibilities to explain their experimental data *in vivo*. Here, to further expand their approach, considering target cell heterogeneity and number of infection events, and to investigate the time-evolution of HIV-1 coinfection in detail, we developed a novel mathematical model, which captures a large dataset from *in vitro* experiments, and characterized HIV-1 cell-free infection dynamics.

Results

Mathematical model considering target cell heterogeneity and number of infection events

Our model extended the well-parameterized basic model of viral dynamics (e.g., (Perelson, 2002)) by incorporating both target cell heterogeneity and number of infection events. N target cell subpopulations with different susceptibility were assumed. Dynamics of the target cell subpopulation i ($\in \{1,\ldots,N\}$) and the viral dy- namics in cell culture were modelled as follows:

$$\frac{dT_i}{dt} = -\beta s_i V_R T_i - \beta s_i V_G T_i$$

$$\frac{dI_i^{R_1}}{dt} = \beta s_i V_R T_i - \beta s_i V_R I_i^{R_1} - \beta s_i V_G I_i^{R_1},$$

$$\frac{dI_i^{R_2}}{dt} = \beta s_i V_R I_i^{R_1} - \beta s_i V_G I_i^{R_2},$$

$$\frac{dI_i^{G_1}}{dt} = \beta s_i V_G T_i - \beta s_i V_R I_i^{G_1} - \beta s_i V_G I_i^{G_1},$$

$$\frac{dI_i^{G_2}}{dt} = \beta s_i V_G I_i^{G_1} - \beta s_i V_R I_i^{G_2},$$

$$\frac{dI_i^{R_1G_1}}{dt} = \beta s_i V_R I_i^{G_1} + \beta s_i V_G I_i^{R_1} - \beta s_i V_R I_i^{R_1G_1} - \beta s_i V_G I_i^{R_1G_1},$$

$$\frac{dI_i^{R_2G_1}}{dt} = \beta s_i V_G I_i^{R_2} + \beta s_i V_R I_i^{R_1G_1} - \beta s_i V_G I_i^{R_2G_1},$$

$$\frac{dI_i^{R_1G_2}}{dt} = \beta s_i V_R I_i^{G_2} + \beta s_i V_G I_i^{R_1G_1} - \beta s_i V_R I_i^{R_1G_2},$$

$$\frac{dI_i^{R_2G_2}}{dt} = \beta s_i V_R I_i^{R_1G_2} + \beta s_i V_G I_i^{R_2G_1},$$

$$\frac{dV_R}{dt} = -cV_R,$$

$$\frac{dV_G}{dt} = -cV_G. \qquad (1)$$

Here T_i is the number of uninfected (i.e., susceptible) target cell subpopulation i. The subpopulation i was further divided into 9 sub-groups by the state of infection: $I_i^{R_j}$ and $I_i^{G_k}$ are infected cells either by HSA HIV-1 (symbolized by R) or GFP HIV-1 (symbolized by G), $I_i^{R_jG_k}$ is coinfected cells, where the subscripts j and k correspond to the number of infection events either by HSA or by GFP HIV-1, respectively (see Fig. 1). V_R and

V_G are the amount of HSA and GFP HIV-1, respectively. The *de novo* infection, $\beta s_i V_R + \beta s_i V_G$ is assumed linearly dependent on the number of viral populations, V_R and V_G. β is the infection rate and s_i is the relative susceptibility for subpopulation i. For sim- plicity, we ordered the target cell subpopulations by susceptibility (i.e., $1 = s_1 > s_2 > \cdots > s_i$ for $i \geq 2$). That is, $\beta(= \beta s_1)$ is the net infection rate of the most susceptible target cell population T_1 for HIV-1 and, for example, βs_2 is the reduced infection rate of the 2nd susceptible populations T_2. Since these HSA and GFP HIV-1 could not infect their target cells multiple times (i.e., single cycle experiments), we only considered a viral clearance rate c but not virus production. As we previously estimated in (Ito et al., 2017), more than 99.9% of infection events were largely occupied by 1 and/or 2 infections by either HSA or GFP HIV-1 even in the 100 and 100 µl of double HSA and GFP HIV-1 experiment. Therefore, we neglected the case which $j, k > 3$ in Eq.(1) for analysing our experimental datasets, and assumed that more than one infection events did not occur simultaneously (i.e., $\beta^2 = 0$). Based on our previous findings, the death rate of HIV-1 infected cell is negligible because cells were exposed to the virus only during 2 hours in our experiments (Ito et al., 2017; Remion et al., 2016).

Heterogeneity of target cell populations

To characterize the heterogeneity of target cell population we quantified the goodness-of-fit between the candidate models with different number of cell subpopulations, and the experimental data (see Methods for our data fitting). To this end, we calculated the Akaike Information Criteria (AIC) for each fit using

$$AIC = N_{pts} \ln\left(\frac{SSR(\widehat{\boldsymbol{\theta}})}{N_{pts}}\right) + 2N_{par},$$

and compared them for the different number of susceptible subpopulations in **Fig2**. Here N_{pts} and N_{par} represent the number of data points and the number of parameters, respectively. The sum of the residuals $SSR(\widehat{\boldsymbol{\theta}})$ is explained in **Methods**. Interestingly, complete homogeneity of target cells (i.e., the number of susceptible target cell population is equal to "1") yields a very poor fit (high AIC) to our cell culture experimental datasets (see **Fig2**). Although the very best fit (smallest AIC) was obtained with two susceptible subpopulations (i.e., $N = 2$), higher dimensional models having more than three subpopulations provide an

adequate description of the data. We summarized our best-fitted parameter values (i.e., estimated parameter values for $N = 2$) in **Table1**. A set of representative data fitting is shown in **Fig3**. In these panels, the colored and white bars represent experimental and predicted frequencies for each quadrant, respectively. In both single (**Fig3a** and **b**) and double (**Fig3c**) HIV-1 infection experiments, our mathematical model reproduces all experimental datasets well. An independent set of data and the corresponding analysis is shown in **FigS1** in **Appendix A**, both for single and double infection experiments. In addition, our mathematical model with estimated parameters successfully reproduces an important index on non-randomness of HIV-1 coinfection, i.e., odd ratio (Dang et al., 2004; Del Portillo et al., 2011; Ito et al., 2017; Remion et al., 2016), as well (see **FigS2** in **Appendix B**).

Dynamics of HIV-1 coinfection during cell-free infection

The FACS analysis itself could not extract a time-evolution of HIV-1 coinfection (see **Methods**). This is because conventional FACS analysis cannot distinguish infected cells having different number of infection events by the same virus (e.g., $I_i^{R_1}$ and $I_i^{R_2}$; $I_i^{R_1G_1}$ and $I_i^{R_2G_2}$). In this study, using the mathematical model with our estimated parameter values, we reconstructed the frequency of HIV-1 coinfection dynamics during the cell-free infection in cell culture. For example, as described in **Fig4a-d**, we calculated the frequency dynamics of HSA HIV-1 infected (i.e., quadrant A), double infected (quadrant B), uninfected susceptible target cells (quadrant C) and GFP HIV-1 infected (quadrant D) cells in the 100 and 100 μl of double HSA and GFP HIV-1 experiment, respectively. We found that the frequency of infection in the 1st susceptible target cell population was 6.11-fold higher than that of the 2nd susceptible cell population, when cells were exposed to the virus inoculum for 2 hours. In other words, cells from the more susceptible population were almost selectively infected among the total target cell population (c.f., (Ito et al., 2017)). Furthermore, in **Fig4e**, we calculated the time-evolution of the number of infection events. Interestingly, multiple infection events rapidly occur upon exposure of cells to the virus, and they accumulate in up to 10% of all target cells within hours.

Contribution of differently susceptible target cell populations to coinfection

The more susceptible a target cell is, the more likely it will be infected by HIV-1. To investigate how target cell heterogeneity affects the HIV-1 coinfection frequency, we quantified the contribution of different target cell subpopulation to the event. The contribution of T_i at 2 hours post exposure to the virus is defined as follows:

$$\frac{I_i^{R_2}(2) + I_i^{G_2}(2) + \sum_{j=1}^{N}\sum_{k=1}^{N} I_i^{R_jG_k}(2)}{\sum_{i=1}^{N} I_i^{R_2}(2) + \sum_{i=1}^{N} I_i^{G_2}(2) + \sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2} I_i^{R_jG_k}(2)}.$$

Interestingly, in **Fig5**, we revealed that 98.3% of coinfection events emerged within the most susceptible target cell subpopulation (i.e., 1st susceptible target cell population: T_1) on average based on the mathematical model with our estimated parameter values. This demonstrates that certain susceptible target cell subpopulations might be a driving force of HIV-1 recombination which may lead to drug resistance and immune escape.

Methods

HIV-1 coinfection experiment

The proviral constructs 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. To limit infections to a single cycle, we used Env-defective constructs, pseudotyped by HIV Env glycoproteins expressed in trans. Stocks of viruses expressing either GFP or HSA were prepared by transfecting sub-confluent 293-T cells in T75 flasks by JetPei (Polyplus Inc. Illkrich, France), following the manufacturer's instructions. Subsequently, MT4R5 cells in a 96-well plate were exposed to two-fold dilutions of one virus (single infection) or of both viruses at the same time (coinfection), using different ratios of the two viruses (as detailed in the results). Finally, flow cytometry data were acquired using a FACS-Calibur instrument (Becton Dickinson, Le Pont-de Claix, France) with CellQuest software and were analysed using FlowJo software (Treestar, Ashland, OR, USA). Note that the detailed description of our experiment can be found in our previous publications (Ito et al., 2017; Remion et al., 2016).

Data fitting

In coinfection experiments, the expression of the two reporter proteins (GFP and HSA) allows to separate infected cells in different quadrants of the FACS analysis: A: HSA-positive (HSA^+), B: both HSA-positive and GFP-positive (HSA^+GFP^+), C: both HSA-negative and GFP-negative (HSA^-GFP^-), and D: GFP-positive (GFP^+) (**Fig1a**). Since these data are measured as frequencies, we normalized the total cell number (i.e., $\sum_{i=1}^{N} T_i(t) + \sum_{i=1}^{N}\sum_{j=1}^{2} I_i^{R_j}(t) + \sum_{i=1}^{N}\sum_{k=1}^{2} I_i^{G_k}(t) + \sum_{i=1}^{N}\sum_{j=1}^{2}\sum_{k=1}^{2} I_i^{R_jG_k}(t) = 1$), and calculated the expected values corresponding to each quadrant using Eq.(1) at 2 hours post exposure to the virus. Note that FACS analysis does not allow to identify the number of infection events per cell, as cells infected by one or multiple virions carrying the same reporter gene will be indistinguishable. Here, for considering the different amounts of inoculated HIV-1, we assumed that $\overline{V_R}(0) = V_R(0)$ and $\overline{V_G}(0) = V_G(0)$ for $3.12\,\mu l$ of HIV-1

supernatant (our smallest inoculum), while $\overline{V_R}(0) = V_R(0) \times r_*$ and $\overline{V_G}(0) = V_G(0) \times r_*$ for the other inoculated amounts in Eq.(1). Therefore, the set of parameters n Eq.(1), $\boldsymbol{\theta} = (\beta, c, s_i, V_R(0), V_G(0), r_*)$ was estimated by fitting: infection rate β, viral clearance rate c, reduction rate on the target cell susceptibility s_i, HSA-positive initial value $V_R(0)$, GFP-positive initial value $V_G(0)$, the scaling parameters r_* for $6.25, 12.5, 25, 37, 50, 75, 100, \text{and } 200\ \mu l$ of the inoculated HIV-1 (see **Table1**). Note that $s_1 = 1$ and number of differently susceptible target cell subpopulations, N, is determined by the AIC values (see **Results**).

To estimate those parameters in Eq.(1), we assumed the dataset follows the Gaussian distribution with mean μ_i and common variance σ^2, and used the following likelihood function:

$$L(\boldsymbol{\mu}, \sigma^2; \boldsymbol{x}) = \prod \prod \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(\boldsymbol{\mu}-\boldsymbol{x})^2}{2\sigma^2}} = \left(\prod \prod \frac{1}{\sqrt{2\pi\sigma^2}}\right) e^{-\frac{SSR(\boldsymbol{\theta})}{2\sigma^2}}.$$

Here, $\boldsymbol{\mu}$, $\boldsymbol{x}$ and $\boldsymbol{\theta}$ represent the set of the expected values by Eq.(1), the experimental measurements and the set of parameters in Eq.(1), respectively. The sum of the residuals, $SSR(\boldsymbol{\theta})$, is defined as

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

where $F_{*,\#}$ and $\tilde{F}_{*,\#}$ are the expected values by Eq.(1) and the experimental measurements at 2 hours post exposure to the virus, respectively. Here $*$ and $\#$ correspond to one of the quadrants (i.e., A, B, C, D) and one of the experiments with different amounts of inoculated HIV-1 (i.e., $r = 1, \ldots, 10$ correspond to 3.12, 6.25, 12.5, 25, 37, 50, 75, 100, 100 and 200 µl, respectively, of single HSA HIV-1 experiments; $g = 1, \ldots, 10$ correspond to 3.12, 6.25, 12.5, 25, 50, 50, 75, 100, 100 and 200 µl, respectively, of single GFP HIV-1 experiments; $co = 1, \ldots, 18$ correspond to 25 and 25, 25 and 50, 50 and 25, 37 and 50, 50 and 50, 37 and 75, 75 and 50, 75 and 75, 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 µl and 100 and 100 µl, respectively of double HSA and GFP HIV-1 experiments).

To consider the uncertainty of the variance as well as parameters in Eq. (1), we employed a bayesian inference approach to estimate the parameters. We applied MCMC method sampling by "delayed rejection and adaptive Metropolis algorithm" to the parameter estimation using the R FME package (R version 3.4.1) (Soetaert and Petzoldt, 2010). In this package, the error between experimental values and predicted values is assumed to follow a Gaussian distribution with mean 0 and the variance σ^2. The additional assumption is that the inverse of the variance (i.e., $1/\sigma^2$) follows a Gamma distribution, and the prior distribution of the parameters follows a Gaussian distribution. We set 300,000 iterations, while the first 200,000 chains are discarded as burn-in-sample. Finally, the convergence of the posterior distributions of the all parameters is confirmed by checking the "traceplot" and "histgram". The 95% credible interval is the range from 2.5% to 97.5%, and each mean represents one of the posterior distribution (**Table1**).

Discussion

HIV mutation and recombination are driving forces of HIV evolution, especially for immune evasion and drug resistance *in vivo* (Gottlieb et al., 2004; Nora et al., 2007; Price et al., 1997). HIV recombination has been well studied both *in vitro* and *in vivo* (Cromer et al., 2016; Law et al., 2016; Levy et al., 2004b; Schlub et al., 2010). An essential factor for viral recombination is coinfection, and therefore it has been investigated extensively (Dang et al., 2004; Del Portillo et al., 2011; Dixit and Perelson, 2005; Ito et al., 2017; Remion et al., 2016). Taking advantage of mathematical modeling and its computational simulations, the dynamics of viral recombination and coinfection has been gradually revealed (Dixit and Perelson, 2005; Giorgi et al., 2013; Keele et al., 2008), although those studies did not take into account the heterogeneity of the target cell population.

To further our understanding of the coinfection (and multiple infection) dynamics for HIV-1 infection, in particular, in the context of different susceptible target cell populations, we propose the novel mathematical model including target cell heterogeneity (i.e., described as i) and explicitly counting the number of infection events in infected cells (i.e., j and k) (**Fig1**). By our experimental-theoretical approach, we characterized the heterogeneity of target cell population, and found that the presence of two different susceptible subpopulations is enough to reproduce our cell culture experimental datasets for multiple infections, although models having three or more subpopulations also might be acceptable (**Fig2** and **Fig3**). For example, in (Dang et al., 2004), it was reported that a simulation model considering 5 different susceptible subpopulations explained the nonrandom HIV-1 infection in cell culture. They did not estimate the number of subpopulations from experimental datasets, but our results further support their simple assumption.

In addition, based on our estimated parameter values, we were able to extract the dynamics of HIV-1 coinfection during cell-free infection in cell culture (**Fig4**). Our simulation explained that multiple infection events rapidly occur and accumulate in up to 10% of all target cells within hours. This implies that there might be massive multiple infection events, which enhance viral recombination, during an exponential viral growth phase in patients. In fact, it has been reported that *in vivo* the estimated template switching rate is close to *in vitro* estimated rate found in primary T lymphocytes (Cromer et al., 2016). Furthermore, our important finding

is that around 98% of the infection by more than 1 virus has emerged from the most susceptible target cell subpopulation (**Fig5**). This demonstrates that certain target cell populations play a dominant role as major coinfection source, as compared to the whole target cell population. Since highly susceptible target cells are selectively infected and removed from the population due to apoptosis or virus lysis, one can hypothesize that the coinfection rate and its associated viral recombination rate reach a maximum during the exponential viral growth phase, a possibility that it will be worth examining in future studies. This is indeed a testable hypothesis both experimentally and mathematically.

HIV-1 coinfection is an important strategy to promote the genetic recombination followed by viral evolution, with consequences for viral escape from the immune or pharmacological pressures. As we previously estimated in (Iwami et al., 2015), the cell-to-cell infection accounts for approximately 60% of HIV-1 infection events in cell culture and therefore this is a major infection mode. Despite its lower infection efficiency, we show here that cell-free HIV transmission results in frequent coinfection events. In the context of genetic recombination, cell-free infection may have an advantage over cell-to-cell transmission, in that it is more likely to bring together viruses that have different genotypic and phenotypic properties. Indeed, cell-free viruses produced by different cells, may carry advantageous traits due to their exposure to different selective forces. Our study, by exploring the conditions that favor coinfection and its kinetics, help elucidating this process that participates in virus escape.

Figure

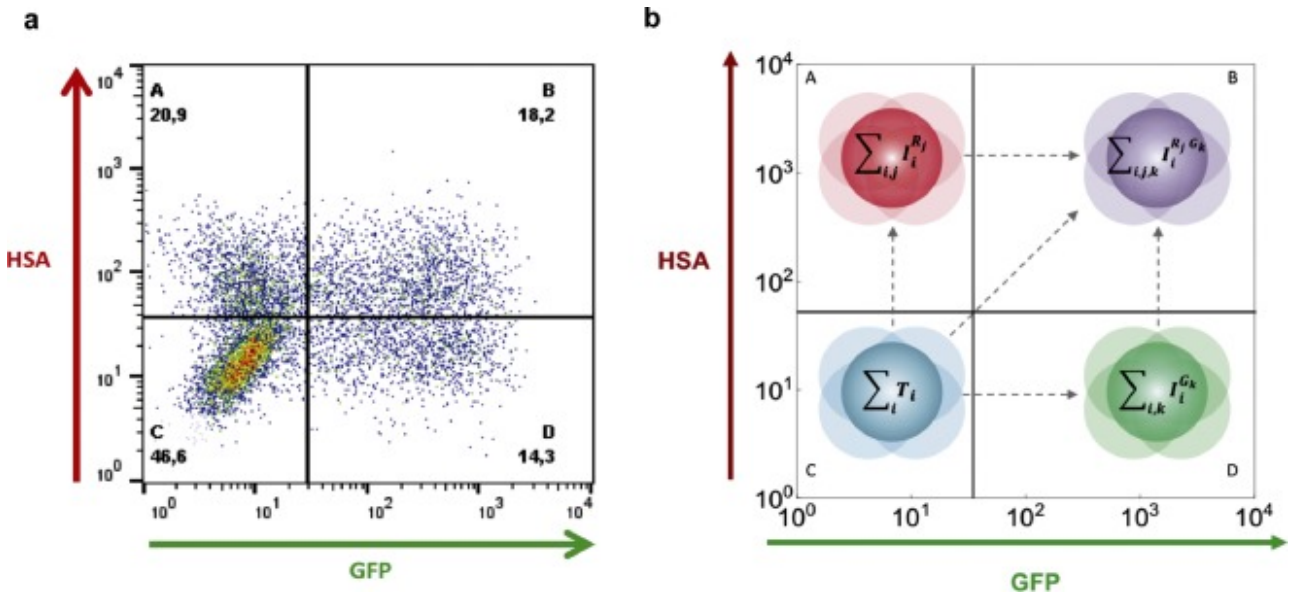


Figure 1 | Flow cytometry analysis of HIV-1 double infection and its mathematical modelling. (a) The panel shows a representative flow cytometry result of double infection with HSA and GFP viruses: the quadrants correspond to HSA^+ cells (A); HSA^+GFP^+ cells (B); uninfected susceptible target cells (C); and GFP^+ cells (D). The colors of dots represents the density of the above cells. The percentage of cells in each quadrant is indicated under the letter identifying the quadrant. **(b)** Each quadrant, i.e., (A), (B), (C) and (D) is defined as $\sum_{i,j} I_i^{R_j}$, $\sum_i T_i$, $\sum_{i,k} I_i^{G_k}$ and $\sum_{i,j,k} I_i^{R_j G_k}$, respectively, in our mathematical model. Here the indices i represents the class of target cell susceptibility, and j,k correspond to the number of infection events either by HSA virus or by GFP virus, respectively.

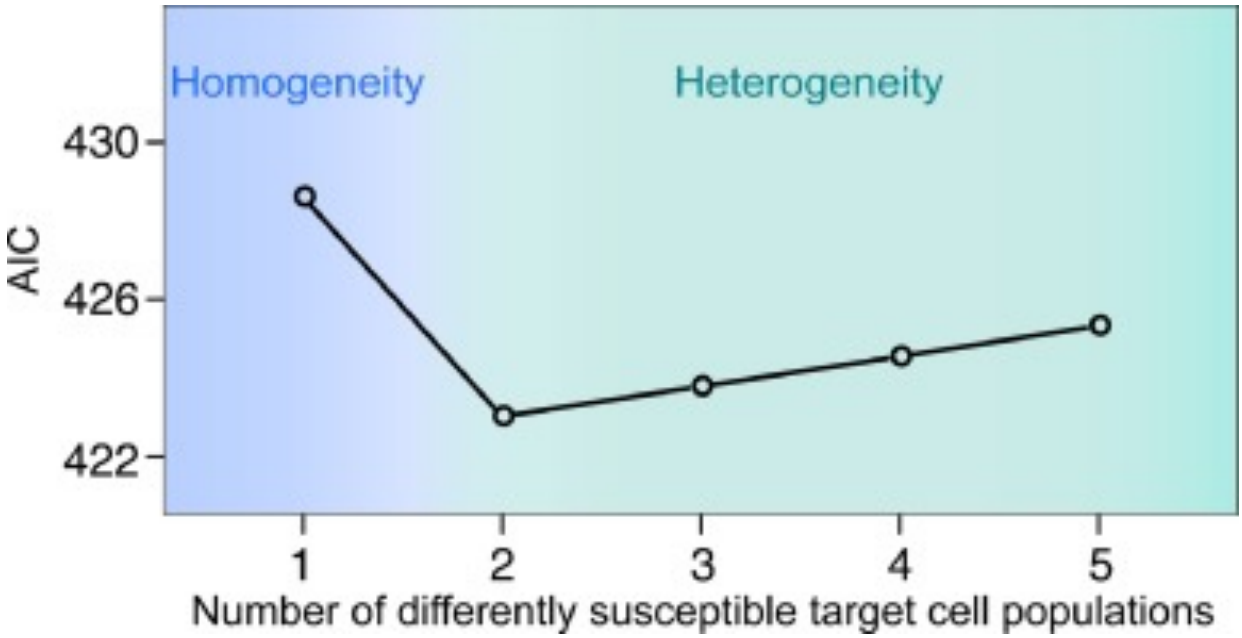


Figure 2 | Target cell heterogeneity based on AIC calculation. For each number of different susceptible target cell populations, AIC values were calculated in Eq.(1), and compared. The mean value of SSR derived from Markov Chain Monte Carlo (MCMC) inferences was used to calculate the AIC value.

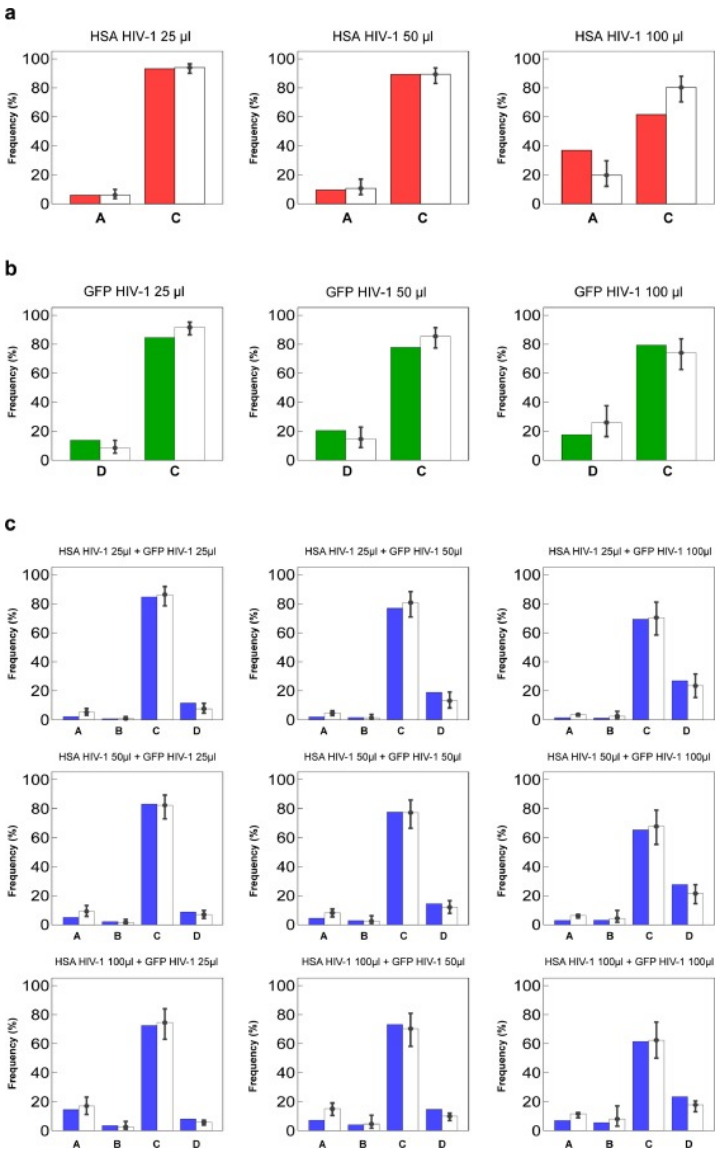


Figure 3 | Frequency of single infection and double infection. (a) The experimental and predicted frequencies of quadrants A (i.e., HSA^+) and C (i.e., HSA^-) at 2 hours post exposure to the virus in three independent experiments using only HSA HIV-1 are shown by red and white bars, respectively. **(b)** The experimental and predicted frequencies of quadrants D (i.e., GFP^+) and C (i.e., GFP^-) in single GFP HIV-1 experiments are shown by green and white bars, respectively. **(c)** The experimental and predicted frequencies of quadrants A (i.e., HSA^+), B (i.e., HSA^+GFP^+), C (i.e., HSA^-GFP^-), and D (i.e., GFP^+) in double HIV-1 experiments are shown by blue and white bars, respectively. Two independent experiments were performed with the nine indicated combinations of HSA and GFP HIV-1 using the three different amounts of each virus used in single experiments. Note that each error bar represents the 95% credible interval obtained from MCMC inferences (**Table 1**).

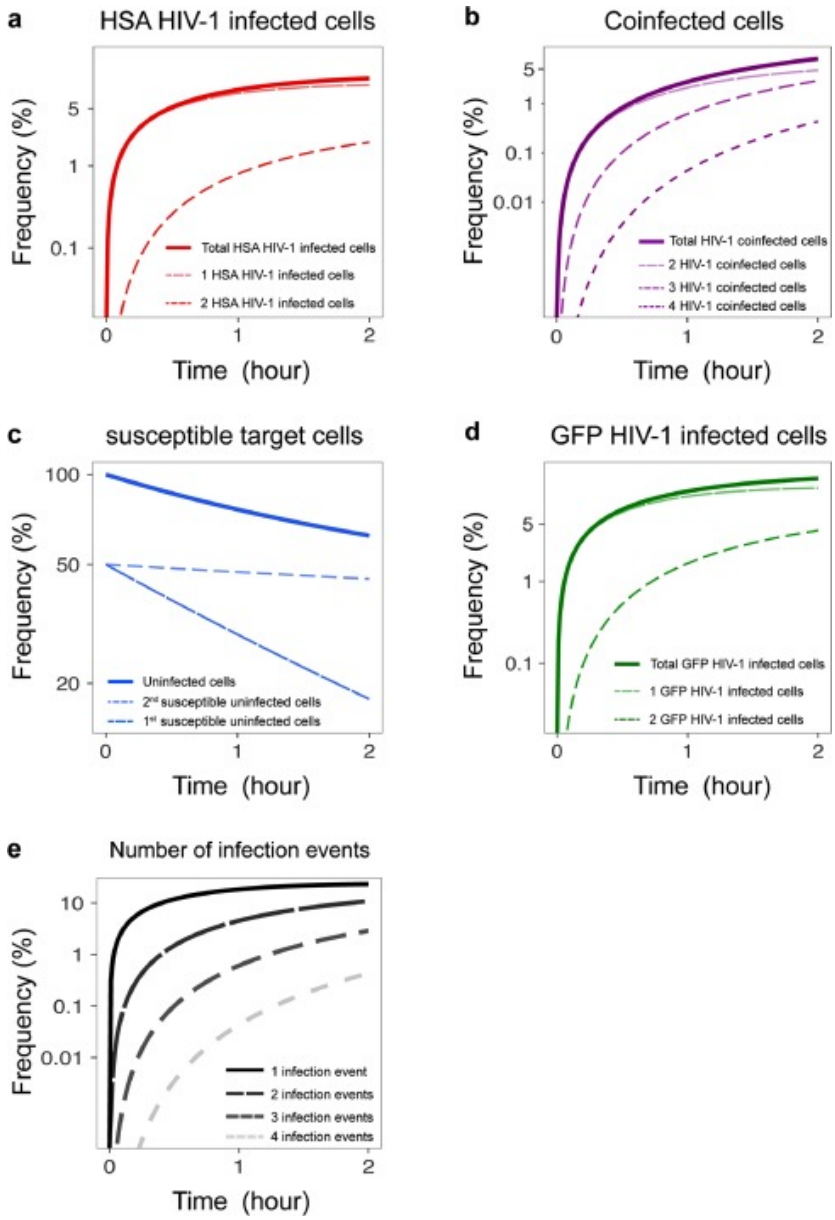


Figure 4 | Dynamics of HIV-1 coinfection during cell-free infection. The predicted frequency dynamics of HSA-infected (i.e., quadrant A), double infected (quadrant B), uninfected susceptible target (quadrant C) and GFP-infected (quadrant D) cells were calculated in **(a)**, **(b)**, **(c)** and **(d)**, respectively, using the mean estimated parameter values derived from MCMC inferences in 100 and 100 μl of double HSA and GFP HIV-1 experiment. The solid and dashed lines in **(a)**, **(b)** and **(d)** correspond to the dynamics of total HIV-1 infection and each dynamics of 1, 2, 3 and 4 (co)infection event(s) denoted as in panels, respectively. In **(c)**, the solid and dashed lines represent the dynamics of total and each uninfected susceptible cell population, respectively. In **(e)**, the dynamics of 1, 2, 3 and 4 infection event(s) were separately calculated.

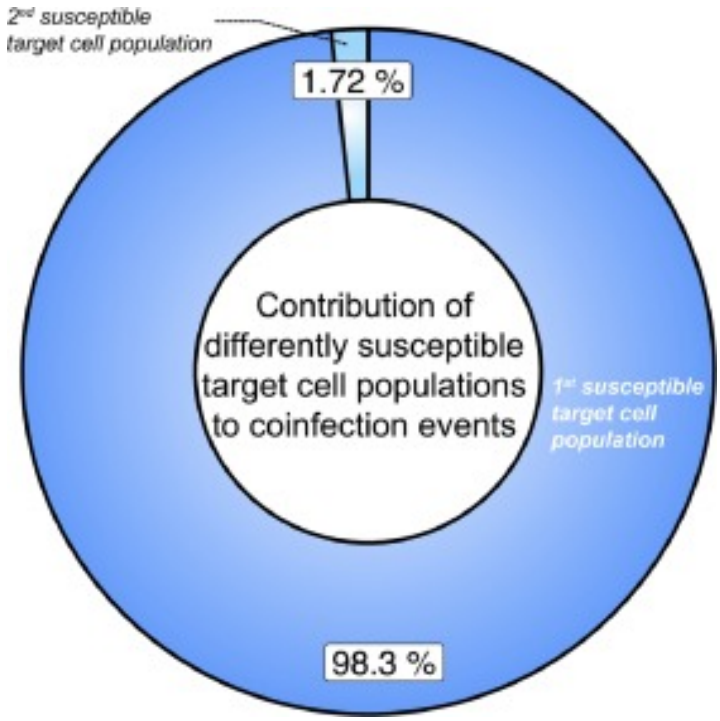


Figure 5 | Contribution of differently susceptible target cell populations to coinfection events. In 18 double HSA and GFP HIV-1 experiments**,** 98.28% and 1.72% of coinfected cells, on average, were generated from 1st and 2nd susceptible target cell population at 2 hours post exposure to the virus, respectively.

Table

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

Parameter name	Parameters	Unit	Estimated mean values	95% credible interval
Number of differently susceptible target cell subpopulations	N	-	2	-
Infection rate	β	(virion/μl)$^{-1}$·(hour)$^{-1}$	2.709×10^{-5}	$(1.818 - 3.836) \times 10^{-5}$
Viral clearance rate	c	(hour)$^{-1}$	0.3060	$0.08060 - 0.5347$
Reduction rate on target cell susceptibility	s_2	-	0.1304	$0.1075 - 0.1566$
Initial HSA HIV-1 viral amount	$\overline{V_R}(0)$	(RNA copies/μl)	0.8505×10^3	$(0.5967 - 1.200) \times 10^3$
Initial GFP HIV-1 viral amount	$\overline{V_G}(0)$	(RNA copies/μl)	1.205×10^3	$(0.8383 - 1.704) \times 10^3$
Increase rate at $6.12\mu l$ of inoculated virus	$r_{6.25}$	-	2.311	$1.735 - 2.958$
Increase rate at $12.5\mu l$ of inoculated virus	$r_{12.5}$	-	4.011	$3.138 - 4.913$
Increase rate at $25\mu l$ of inoculated virus	r_{25}	-	4.040	$3.247 - 4.874$
Increase rate at $37\mu l$ of inoculated virus	r_{37}	-	10.27	$8.195 - 12.43$
Increase rate at $50\mu l$ of inoculated virus	r_{50}	-	7.425	$6.005 - 8.938$
Increase rate at $75\mu l$ of inoculated virus	r_{75}	-	11.86	$9.512 - 14.26$
Increase rate at $100\mu l$ of inoculated virus	r_{100}	-	15.03	$12.13 - 18.08$
Increase rate at $200\mu l$ of inoculated virus	r_{200}	-	24.24	$19.59 - 29.11$

Appendix

Appendix A

In addition to the representative experimental datasets in **Fig3**, we used the other datasets from 7 single HSA HIV-1 infection, 7 single GFP HIV-1 infection and 9 double HIV-1 infection in our simultaneous fitting for our parameter estimation. As described in **FigS1**, our mathematical model fits well to datasets independent of inoculated viral amounts.

Appendix B

HIV-1 coinfection events were demonstrated to occur more frequently than would be expected for independent infection events and do not follow a random distribution (Chen et al., 2005; Dang et al., 2004; Ito et al., 2017; Remion et al., 2016). If the coinfection occurred more frequently than that expected from random events, then the expected odds ratio would be more than 1. As we reported in (Ito et al., 2017), the experimental odds ratio, $F_{B,\#}F_{C,\#}/F_{A,\#}F_{D,\#}$, was estimated to be more than 1 independent of inoculated viral amount. To quantitatively further support our model prediction, we calculated the theoretical odds ratio, $\tilde{F}_{B,\#}\tilde{F}_{C,\#}/\tilde{F}_{A,\#}\tilde{F}_{D,\#}$, from our mathematical model with our estimated parameters in **Table1**. In **FigS2**, we compared the odds ratio measured by our experiments with the theoretical odds ratio. Thus, our novel model quantitatively reproduces the odds ratio and captures the known property of coinfection during cell-free HIV-1 infection in cell culture.

Appendix Figure

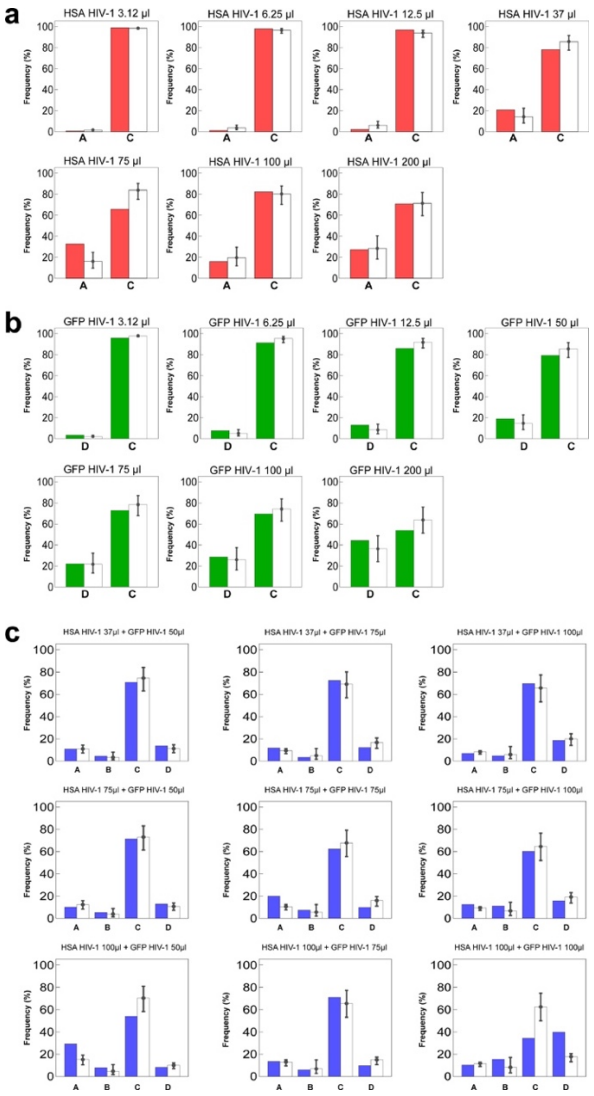


Figure S1 | Frequency of single infection and double infection in other experiments: **(a)** The experimental and predicted frequencies of quadrants A (i.e., HSA^+) and C (i.e., HSA^-) at 2 hours post exposure to the virus in seven independent experiments using only HSA HIV-1 are shown by red and white bars, respectively. **(b)** The experimental and predicted frequencies of quadrants D (i.e., GFP^+) and C (i.e., GFP^-) in single GFP HIV-1 experiments are shown by green and white bars, respectively. **(c)** The experimental and predicted frequencies of quadrants A (i.e., HSA^+), B (i.e., HSA^+GFP^+), C (i.e., HSA^-GFP^-), and D (i.e., GFP^+) in double HIV-1 experiments are shown by blue and white bars, respectively, with different combinations of inoculated viral amount. Note that each error bar represents 95% credible interval obtained from Markov Chain Monte Carlo (MCMC) parameter inferences.

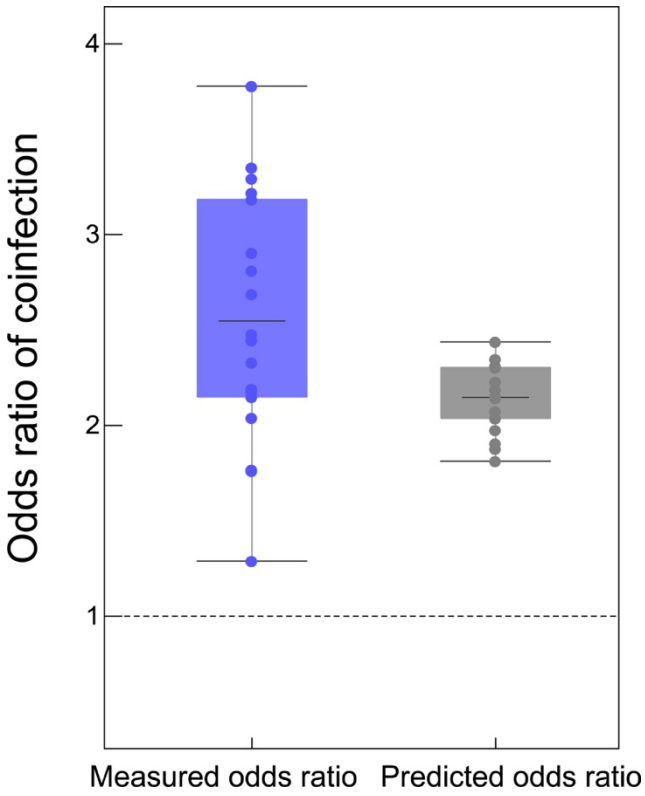


Figure S2 | Comparison of measured odds ratio with predicted odds ratio: The experimental odds ratio and predicted odds ration are shown in blue and black box plots, respectively, using 18 HIV-1 double infection experiment datasets. The dotted line corresponds to the odds ratio of 1 predicted by a simple homogeneous target cell model (i.e., Eq.(1) with $i = 1$). Note that the predicted odds ratios were calculated by the mean value of posterior distribution obtained from MCMC method.

Reference

Allen, T.M., and Altfeld, M. (2003b). HIV-1 superinfection. J Allergy Clin Immunol *112*, 829-835; quiz 836.

Barre-Sinoussi, F., Ross, A.L., and Delfraissy, J.F. (2013). Past, present and future: 30 years of HIV research. Nat Rev Microbiol *11*, 877-883.

Burke, D.S. (1997b). Recombination in HIV: An important viral evolutionary strategy. Emerging Infectious Diseases *3*, 253-259.

Chang, C.C., Crane, M., Zhou, J.L., Mina, M., Post, J.J., Cameron, B.A., Lloyd, A.R., Jaworowski, A., French, M.A., and Lewin, S.R. (2013). HIV and co-infections. Immunol Rev *254*, 114-142.

Chen, J.B., Dang, Q., Unutmaz, D., Pathak, V.K., Maldarelli, F., Powell, D., and Hu, W.S. (2005). 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.

Cromer, D., Grimm, A.J., Schlub, T.E., Mak, J., and Davenport, M.P. (2016). Estimating the in-vivo HIV template switching and recombination rate. AIDS (London, England) *30*, 185-192.

Dang, Q., Chen, J.B., Unutmaz, D., Coffin, J.M., Pathak, V.K., Powell, D., KewalRamani, V.N., Maldarelli, F., and Hu, W.S. (2004). Nonrandom HIV-1 infection and double infection via direct and cell-mediated pathways. P Natl Acad Sci USA *101*, 632-637.

Del Portillo, A., Tripodi, J., Najfeld, V., Wodarz, D., Levy, D.N., and Chen, B.K. (2011). Multiploid Inheritance of HIV-1 during Cell-to-Cell Infection. J Virol *85*, 7169-7176.

Dixit, N.M., and Perelson, A.S. (2005). HIV dynamics with multiple infections of target cells. P Natl Acad Sci USA *102*, 8198-8203.

Donahue, D.A., Bastarache, S.M., Sloan, R.D., and Wainberg, M.A. (2013). 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.

Fauci, A.S. (2003). HIV and AIDS: 20 years of science. Nat Med *9*, 839-843.

Fraser, C. (2005). HIV recombination: what is the impact on antiretroviral therapy? J R Soc Interface *2*, 489-503.

Giorgi, E.E., Korber, B.T., Perelson, A.S., and Bhattacharya, T. (2013). Modeling sequence evolution in HIV-1 infection with recombination. Journal of theoretical biology *329*, 82-93.

Gottlieb, G.S., Nickle, D.C., Jensen, M.A., Wong, K.G., Grobler, J., Li, F.S., Liu, S.L., Rademeyer, C., Learn, G.H., Karim, S.S.A., *et al.* (2004). Dual HIV-1 infection associated with rapid disease progression. Lancet *363*, 619-622.

Ito, Y., Remion, A., Tauzin, A., Ejima, K., Nakaoka, S., Iwasa, Y., Iwami, S., and Mammano, F. (2017). Number of infection events per cell during HIV-1 cell-free infection. Sci Rep-Uk *7*.

Iwami, S., Takeuchi, J.S., Nakaoka, S., Mammano, F., Clavel, F., Inaba, H., Kobayashi, T., Misawa, N., Aihara, K., Koyanagi, Y., *et al.* (2015). Cell-to-cell infection by HIV contributes over half of virus infection. eLife *4*.

Josefsson, L., King, M.S., Makitalo, B., Brannstrom, J., Shao, W., Maldarelli, F., Kearney, M.F., Hu, W.S., Chen, J.B., Gaines, H., *et al.* (2011). Majority of CD4(+) T cells from peripheral blood of HIV-1-infected individuals contain only one HIV DNA molecule. P Natl Acad Sci USA *108*, 11199-11204.

Josefsson, L., Palmer, S., Faria, N.R., Lemey, P., Casazza, J., Ambrozak, D., Kearney, M., Shao, W., Kottilil, S., Sneller,

M., *et al.* (2013). Single Cell Analysis of Lymph Node Tissue from HIV-1 Infected Patients Reveals that the Majority of CD4(+) T-cells Contain One HIV-1 DNA Molecule. Plos Pathog *9*.

Jung, A., Maier, R., Vartanian, J.P., Bocharov, G., Jung, V., Fischer, U., Meese, E., Wain-Hobson, S., and Meyerhans, A. (2002). Recombination - Multiply infected spleen cells in HIV patients. Nature *418*, 144-144.

Keele, B.F., Giorgi, E.E., Salazar-Gonzalez, J.F., Decker, J.M., Pham, K.T., Salazar, M.G., Sun, C., Grayson, T., Wang, S., Li, H., *et al.* (2008). Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. Proc Natl Acad Sci U S A *105*, 7552-7557.

Law, K.M., Komarova, N.L., Yewdall, A.W., Lee, R.K., Herrera, O.L., Wodarz, D., and Chen, B.K. (2016). In Vivo HIV-1 Cell-to-Cell Transmission Promotes Multicopy Micro-compartmentalized Infection. Cell Rep *15*, 2771-2783.

Levy, D.N., Aldrovandi, G.M., Kutsch, O., and Shaw, G.M. (2004b). Dynamics of HIV-1 recombination in its natural target cells. Proc Natl Acad Sci U S A *101*, 4204-4209.

Nora, T., Charpentier, C., Tenaillon, O., Hoede, C., Clavel, F., and Hance, A.J. (2007). Contribution of recombination to the evolution of human immunodeficiency viruses expressing resistance to antiretroviral treatment. J Virol *81*, 7620-7628.

Perelson, A.S. (2002). Modelling viral and immune system dynamics. Nat Rev Immunol *2*, 28-36.

Powell, R.L., Urbanski, M.M., Burda, S., Kinge, T., and Nyambi, P.N. (2009). High frequency of HIV-1 dual infections among HIV-positive individuals in Cameroon, West Central Africa. JAIDS Journal of Acquired Immune Deficiency Syndromes *50*, 84-92.

Price, D.A., Goulder, P.J., Klenerman, P., Sewell, A.K., Easterbrook, P.J., Troop, M., Bangham, C.R., and Phillips, R.E. (1997). Positive selection of HIV-1 cytotoxic T lymphocyte escape variants during primary infection. Proc Natl Acad Sci U S A *94*, 1890-1895.

Quan, Y., Liang, C., Brenner, B.G., and Wainberg, M.A. (2009b). 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. The Journal of infectious diseases *200*, 1479-1483.

Redd, A.D., Quinn, T.C., and Tobian, A.A.R. (2013). Frequency and implications of HIV superinfection. Lancet Infect Dis *13*, 622-628.

Remion, A., Delord, M., Hance, A.J., Saragosti, S., and Mammano, F. (2016). Kinetics of the establishment of HIV-1 viral interference and comprehensive analysis of the contribution of viral genes. Virology *487*, 59-67.

Ronen, K., McCoy, C.O., Matsen, F.A., Boyd, D.F., Emery, S., Odem-Davis, K., Jaoko, W., Mandaliya, K., McClelland, R.S., and Richardson, B.A. (2013). HIV-1 superinfection occurs less frequently than initial infection in a cohort of high-risk Kenyan women. Plos Pathog *9*, e1003593.

Schlub, T.E., Smyth, R.P., Grimm, A.J., Mak, J., and Davenport, M.P. (2010). Accurately measuring recombination between closely related HIV-1 genomes. PLoS computational biology *6*, e1000766.

Simon-Loriere, E., and Holmes, E.C. (2011). Why do RNA viruses recombine? Nat Rev Microbiol *9*, 617-626.

Smith, D.M., Richman, D.D., and Little, S.J. (2005). HIV superinfection. J Infect Dis *192*, 438-444.

Soetaert, K., and Petzoldt, T. (2010). Inverse Modelling, Sensitivity and Monte Carlo Analysis in R Using Package FME. J Stat Softw *33*.

van der Kuyl, A.C., and Cornelissen, M. (2007). Identifying HIV-1 dual infections. Retrovirology *4*.