Clustering of activated CD8 T cells around malaria-infected hepatocytes is rapid and is driven by antigen-specific T cells

Reka K. Kelemen[1,2], Harshana Rajakaruna[3], Ian A. Cockburn[4] and Vitaly V. Ganusov[2,3,*]

[1] *Institute of Science and Technology, 3400 Klosterneuburg, Vienna, Austria*
[2] *Genome Science and Technology (GST) program, University of Tennessee, Knoxville, TN 37996, USA*
[3] *Department of Microbiology, University of Tennessee, Knoxville, TN 37996, USA*
[4] *Department of Immunology and Infectious Disease, John Curtin School of Medical Research The Australian National University, Canberra, ACT, 2601, Australia*
[*] *Corresponding author: vitaly.ganusov@gmail.com*

May 27, 2019

Abstract

Copyrighted by the AAI (abstract was published as a part of AAI annual meeting).

1 Introduction

Malaria is a life-threatening disease that is a result of red blood cell (erythrocyte) destruction by eukaryotic parasites of the *Plasmodium* genus. The majority of deaths (estimated to be about 500,000 annually) are among children, who have not yet developed immunity to the pathogen [1, 2]. There are five species that infect humans: *P. falciparum*, *P. vivax*, *P. malariae*, *P. ovale*, and *P. knowlesi* [3]. Three species of malaria parasites that are used as animal models for human malaria in mice are *P. yeolii*, *P. berghei*, and *P. chabauidi* [4]. While there are similarities and differences in replication and pathogenesis of Plasmodium species in humans and mice, in this paper we focus solely on infection of mice with Plasmodium parasites.

The infection of the host is started by a mosquito, the vector between mammalian hosts, injecting the sporozoite form of parasites into the skin. Studies have estimated that the initial number of sporozoites entering the host is as low as 10-50 [5, 6], of which only a fraction succeed to migrate to the liver to start an infection of hepatocytes by forming liver stages [7–9]. This liver stage of infection lasts for approximately 6.5 days in humans and about 2 days in mice [10–13]. Because liver stage is asymptomatic, removal of all liver stages prevents clinical symptoms of malaria and thus is highly desirable feature of an effective vaccine. Indeed, previous studies have shown that memory CD8 T cells are required for protection against a challenge with a relatively large number of sporozoites [14, 15] and that vaccination that induces exclusively memory CD8 T cells of a single specificity can mediate sterilizing protection against a sporozoite challenge [16–23]. Antibodies and CD4 T cells may also contribute to protection in some circumstances, for example, following inoculation of sporozoites by mosquitoes in the skin [24, 25]. Given that mouse liver contains about $1 - 2 \times 10^8$

hepatocytes [26–28] and only a tiny proportion of these are infected the ability of memory CD8 T cells of a single specificity to locate and eliminate all liver stages within 48 hours is remarkable. Yet, specific mechanisms by which T cells achieve such an efficiency remain poorly defined.

Recent studies utilizing fluorescently labeled sporozoites and activated Plasmodium-specific CD8 T cells and intravital microscopy revealed clustering of CD8 T cells near the parasite in the mouse livers whereby multiple T cells were located in close proximity ($\leq 40\ \mu$m) of some liver stages [23, 29–31]. Interestingly, we observed that clustering of T cells near the parasite results in a higher chances of parasite's death suggesting that clusters may increase the efficiency at which T cells eliminate the infection. Recent *in vivo* studies also found that the killing of virus-infected cells occurs faster when several T cells are near an infected cell [32] that is consistent with a previous report estimating that killing of targets *in vivo* follows the law of mass-action [33] (meaning that the rate of killing is directly proportional to the concentration of the killers and targets).

Clustering of T cells around Plasmodium liver stages in mice was not uniform as the majority of parasites had no T cells around them (at 6 hours after T cell transfer), while some parasites were surrounded by 20-25 T cells [29]. We have developed three alternative mathematical models aimed at explaining this observed variability in cluster formation and by fitting the models to a subset of the data concluded that the data are best explained by a model in which formation of clusters is driven by a positive feedback loop — clusters of a large size attract more T cells to the site of infection [29]. Analysis of T cell movement in the liver suggested that there may be a bias towards the infection site [34]. Additional experiments revealed a significant correlation between the rate at which new T cells locate the infection site and the number of T cells found in the cluster and independence of the per capita rate at which T cells leave the cluster from the cluster size — both observations were consistent with the "density-dependent" recruitment model.

Yet, our previous analysis did not investigate several other important issues of cluster formation. In particular, we did not fully determine the role of the environment in the formation of clusters around Plasmodium liver stages. For example, the observed correlation between entry rate into a cluster and cluster size could simply arise because some parasites may accidentally "attract" more T cells (e.g., due to higher induced inflammation or a higher blood flow rate, [35, 36]). In addition, we have observed that transfer of activated T cells, specific to Plasmodium, and T cells of irrelevant specificity resulted in co-clustering of T cells of two types [29]; however, whether "non-specific" T cells contributed to the formation of clusters was not determined. Finally, our previous analyses did not determine the kinetics of the cluster formation by assuming that cluster size reaches a steady state by the time of imaging. In this paper we re-analyzed some of the previously published data to further define mechanisms by which clusters of activated CD8 T cells around Plasmodium liver stages are formed.

2 Materials and methods

2.1 Data

In our analyses we used data from previously published work [29, 37]. Data were generated using an experimental system with Py-specific TCR transgenic T cells (PyTCR). For most of our analyses, data were from experiments involving infection of Balb/c mice with variable doses of Plasmodium yoelii (Py) sporozoites, expressing GFP; location of T cells around GFP-expressing liver stages was

then visualized using spinning disk confocal microscopy [29]. Following our previous work we consider T cells located within 40 μm distance from the parasite as being close enough to recognize the infection; thus, all T cells within 40 μm from the parasite are called to form a "cluster". How well the length of 40 μm represents the size of hepatocytes in mice remains unclear. 2D images of mouse hepatocytes suggested the diameter of 40-80 μm [38]; however, measurements of the total volume of mouse hepatocytes of about $V_h = 10^4$ μm^3 [39] suggest a radius $r_h = \sqrt[3]{3V_h/(4\pi)} \approx 13$ μm for a "spherical" hepatocyte or cube edge length $r_h = \sqrt[3]{V_h} \approx 22$ μm for a "cubical" hepatocyte. A classical textbook on human liver anatomy cites the human hepatocyte volume of $10^4 - 6 \times 10^4$ μm^3, corresponding to a cube edge of about 40 μm [35, p. 13]. Despite these inconsistent estimates we consider T cells within 40 μm from the parasite to be close enough for recognition of the parasite.

Clustering of CD8 T cells around the parasite was measured in several alternative experiments. In one set of experiments, clustering of CD8 T cells around Py liver stages was performed by immunizing mice with radiation-attenuated Py sporozoites (RAS). In another set of experiments, clustering of T cells was observed following transfer of activated T cell receptor (TCR) transgenic CD8 T cells, specific to epitope located in Py circumsporozoite (CS) protein ($CS_{280\text{-}288}$: SYVPSAEQI) denoted as PyTCR cells, or of TCR transgenic CD8 T cells, specific to an epitope in chicken ovalbumin ($OVA_{257\text{-}264}$: SIINFEKL) denoted as OT1. Following infection with Py, PyTCR recognize the infection while OT1 cells serve as a control (non-specific to Py) T cells. PyTCR and OT1 cells were activated in similar *in vivo* experiments (using Vaccinia virus expressing CS or OVA epitopes) [29]. Experiments involving co-transfer of PyTCR and OT1 cells were done in CB6 mice (F1 cross of Balb/c and C57BL/6 mice) [29].

In summary, the following datasets were used in the analysis:

1. Dataset #1. fluorescently labeled PyTCR cells ('PyTCR alone') and PyTCR cells pre-treated with pertussis toxin ('PyTCR+PT') and transferred into mice infected with Py. This dataset was published before [29] but was not analyzed with mathematical models.

2. Dataset #2: naive or RAS-immunized mice that were infected with Py-GFP and clustering around the parasite was imaged using anti-CD8 antibody. This dataset was published before [29] and was only analyzed with models based on T cell-intrinsic clustering mechanisms (see below).

3. Dataset #3: fluorescently labeled PyTCR and OT1 cells transferred alone into individual mice ('PyTCR alone' and 'OT1 alone') or as 1:1 mixture ('PyTCR mix' and 'OT1 mix') to Py-infected mice. This dataset was published before [29] but was analyzed assuming that clustering of PyTCR and OT1 T cells was independent.

4. Dataset #4: fluorescently labeled PyTCR cells transferred into Py-infected mice and imaged at two different time points after T cell transfer. This dataset was generated for a previous study [29] but was not analyzed before with the use of mathematical models.

All datasets are made available as an online supplement to this paper to facilitate further independent analyses.

2.2 Mathematical models

2.2.1 Basic mathematical model for clustering of one cell type

Previously we proposed a standard "birth-death" model to describe formation of clusters around Plasmodium liver stages [29]. To be comprehensive, we describe this modeling framework here but then extend it to consider environmental variability in cluster sizes, co-clustering of different T cell specificies, and kinetics of cluster formation. This modeling framework assumes that infection of hepatocytes by Plasmodium sporozoites occurs independently, i.e., there is no interactions between sporozoites infecting different hepatocytes. This assumption of independence is likely to be justified given that in our experiments i) in general $\sim 10^5$ sporozoites are injected i.v. into mice, ii) only a fraction of these is expected to reach the liver [7–9, 40, 41], and iii) mouse liver contains $1-2\times 10^8$ hepatocytes [26–28]. Because in general in our experiments the number of Plasmodium-specific T cells exceeds the number of liver stages by 10 to 30 fold, we assume that clustering of T cells around one parasite does not interfere or compete with T cells clustering around another parasite. In the model describing formation of clusters around Plasmodium liver stages by T cells of a single specificity we denote $P_k(t)$ as the probability to observe k T cells around the parasite at time t with $k = 0, 1, 2, \ldots k_{\max}$. Increase in cluster size occurs at the "birth" (or entry) rate λ_k ($k = 0, 1, 2, \ldots, k_{\max}$) and decline in cluster size occurs due to "death" (or exit) rate μ_k ($k = 1, 2, \ldots, k_{\max}$). The mathematical model describing the change in the probability $P_k(t)$ with time is given by the system of differential equations:

$$\frac{\mathrm{d}P_0(t)}{\mathrm{d}t} = -\lambda_0 P_0(t) + \mu_1 P_1(t), \tag{1}$$

$$\frac{\mathrm{d}P_k(t)}{\mathrm{d}t} = -(\lambda_k + \mu_k)P_k(t) + \mu_{k+1}P_{k+1}(t) + \lambda_{k-1}P_{k-1}(t), \qquad k \geq 1. \tag{2}$$

By assuming different specific forms for the T cell entry (λ_k) and exit (μ_k) rates (e.g., see Figure 1 and below) the model can be solved numerically and fitted to the data using maximum likelihood method (see below). For some analyses we made a simplifying assumption that the distribution of cluster sizes reaches a steady state, and the steady state values for the probability to observe k CD8 T cells near a given liver stage P_k^* is given by

$$P_k^* = P_0 \frac{\prod_{i=0}^{k-1} \lambda_i}{\prod_{i=1}^{k} \mu_i}, \tag{3}$$

where P_0^* is found by normalizing the total probability to one. By assuming steady state solutions it is in general impossible to estimate individual values for the rates of T cell entry into the cluster and exit from the cluster but we can estimate the ratio of the entry and exit rates, which we define as the relative entry rate $\theta_k = \lambda_k/\mu_k$.

Mechanisms explaining the clustering of T cells around Plasmodium parasites in the liver can be broadly divided into two categories: T cell-intrinsic and T cell-extrinsic. In the T cell-intrinsic mechanism, the formation of clusters is driven exclusively by T cells and thus this mechanism ignores any potential differences in the variability in local liver environment. In the T cell-extrinsic mechanism, formation of clusters is driven exclusively due to variability in the liver environment near

individual parasites, for example, due to a higher blood flow to some liver stages or a higher degree of inflammation that individual parasites may induce [35, 36]. It is possible that ultimately both mechanisms may contribute to the cluster formation.

2.2.2 Sub-models assuming T cell-intrinsic clustering mechanisms

We consider several alternative models of how T cells may mediate formation of clusters around Plasmodium liver stages in mice (Figure 1A-C). Some of these models have been presented in our previous publication [29] and here are presented again for completeness. Our simplest random entry/exit (**Poisson model**) assumes that entry into the cluster and exit from the cluster occur randomly, i.e., $\lambda_k = \lambda_0$ and $\mu_k = k\mu$ where λ_0 and μ are constants (Figure 1A). Solving eqn. (3), the probability to observe k T cells around a parasite according to this random entry/exit model is then given by the Poisson distribution:

$$P_k^* = P_0^* \frac{\prod_{i=0}^{k-1} \lambda_i}{\prod_{i=1}^{k} \mu_i} = P_0^* \frac{\lambda_0^k}{\mu^k k!} = \frac{\theta_0^k}{k!} e^{-\theta_0}, \tag{4}$$

where $\theta_0 = \lambda_0/\mu$.

We have shown previously that the Poisson model is often unable to describe distribution of cluster sizes of Plasmodium-specific CD8 T cells in the liver [29]. One potential mechanism proposed to describe formation of large clusters is a "retention" model in which T cells which recognize the infection, are retained near the parasite. One version of such a model is a **density-independent (DI) exit model** (Figure 1B) in which the rate of T cell exit from a cluster declines with the number of T cells in the cluster, i.e., $\mu_k = k\mu/k = \mu$ for $k > 0$ and $\lambda_k = \lambda_0$ for all k. Solving eqn. (3), the probability to observe k T cells around a parasite according to the DI exit model is given by a geometric distribution:

$$P_k^* = P_0^* \frac{\prod_{i=0}^{k-1} \lambda_i}{\prod_{i=1}^{k} \mu_i} = P_0^* \frac{\lambda_0^k}{\mu^k} = (1 - \theta_0)\theta_0^k, \tag{5}$$

where $\theta_0 = \lambda_0/\mu$. There are other ways in which the total rate of T cell exit from the cluster μ_k could decline with cluster size k and in our additional analyses we tested two of such alternative models: a powerlaw model in which $\mu_k = k^\alpha \mu$ (defined for $k > 0$ with α and μ being constant) and an exponential model in which $\mu_k = k\mu e^{-\alpha k}$ (defined for $k > 0$ with α and μ being constant). When fitting these retention models to experimental data we did not derive the steady state solutions but instead used numerical solutions of the basic mathematical model (eqns. (1)–(2)).

An alternative mechanism for the formation of large clusters of CD8 T cells around the infection is an "attraction" model in which the rate of T cell entry into the cluster depends on cluster size. In this **density-dependent (DD) recruitment model** (Figure 1C) the entry rate into the cluster is given by $\lambda_k = \lambda_0 + \lambda_1 k$ while the total exit rate is density-dependent $\mu_k = k\mu$. Solving eqn. (3), the probability to observe k T cells around a parasite according to the DD recruitment model at the steady state is calculated numerically:

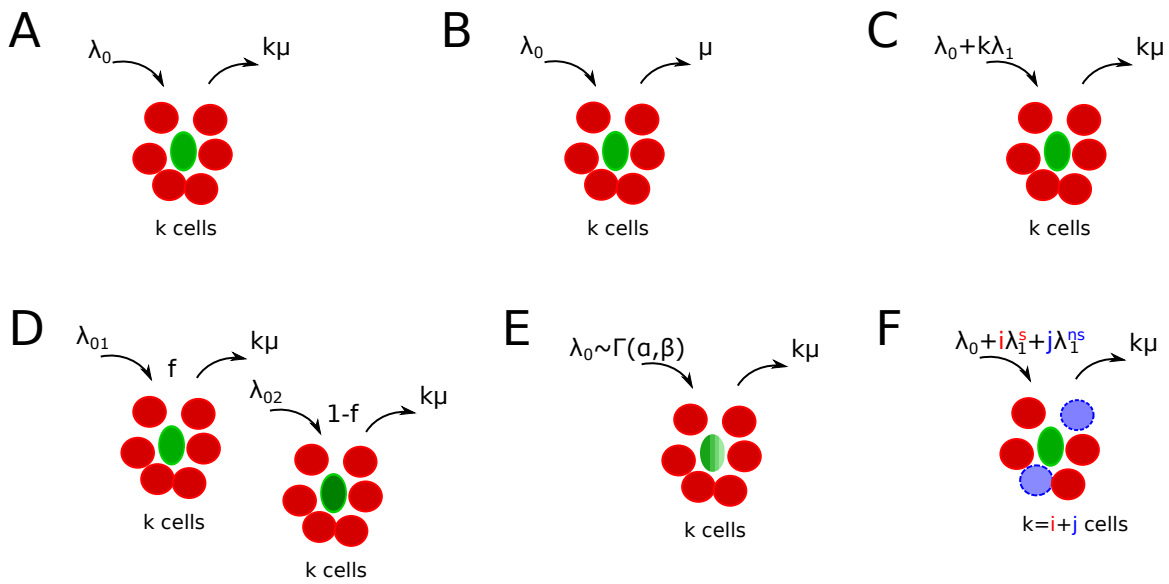

Figure 1: Schematics of alternative mathematical models of T cell cluster formation around Plasmodium yoelii (Py)-infected hepatocytes in mice. Py-specific T cells are labeled by red (discs), T cells of irrelevant specificity are colored by blue (dashed discs), and parasites are labeled by green (ovals). In the models the rate of T cell entry into a cluster is denoted as λ_k and rate of exit from the cluster is denoted as μ_k. Mathematical models include a random entry/exit (Poisson) model (A, eqn. (4), $\lambda_k = \lambda_0$ and $\mu_k = k\mu$), a density-independent (DI) exit model (B, eqn. (5), $\lambda_k = \lambda_0$ and $\mu_k = \mu$), a density-dependent (DD) recruitment model (C, eqn. (6), $\lambda_k = \lambda_0 + k\lambda_1$ and $\mu_k = k\mu$), a "two populations" model in which infected hepatocytes have either of two different "attractiveness" levels determined by λ_{01} and λ_{02} (D, eqn. (10), $\mu_k = k\mu$), a "gamma" model, in which the entry rate into clusters is distributed according to a gamma distribution with α and β being the rate and shape parameters (E, eqn. (11), $\mu_k = k\mu$), and finally a "co-clustering" model, in which clusters are formed by Plasmodium-specific T cells or T cells or irrelevant specificity (nonspecific T cells) (F, eqns. (12)–(15), $\lambda_k = \lambda_0 + i\lambda_1^s + j\lambda_1^{ns}$ and $\mu_k = k\mu$). For some of our analyses we characterized the model behavior using the ratio of entry to exit rates denoted as a relative entry rate $\theta_k = \lambda_k/\mu_k$.

$$P_k^* = P_0^* \frac{\prod_{i=0}^{k-1} \lambda_i}{\prod_{i=1}^{k} \mu_i} = P_0^* \frac{\prod_{i=0}^{k-1} \lambda_0 + i\lambda_1}{\mu^k k!} = P_0^* \frac{\prod_{i=0}^{k-1} \theta_0 + i\theta_1}{k!}, \tag{6}$$

where $\theta_0 = \lambda_0/\mu$ and $\theta_1 = \lambda_1/\mu$ and P_0^* is found by normalizing eqn. (6) assuming the maximal cluster size to be $k_{\max}$. In general, $\sum_{k=0}^{\infty} P_k^* \to \infty$ and therefore, the sum must be taken for a finite number of terms due to this reason [29].

To understand dynamics of cluster formation in the Poisson and DD recruitment models it is also useful and possible to derive the model describing the change in the average number of T cells around the parasite (average cluster size), $\langle k \rangle = \sum_{k=0}^{\infty} k P_k(t)$ using standard methods of physical chemistry [42]

$$\frac{\mathrm{d}\langle k \rangle}{\mathrm{d}t} = \lambda_0 + (\lambda_1 - \mu)\langle k \rangle, \tag{7}$$

which is a standard birth-death process with immigration which for $\langle k \rangle(0) = 0$ has the solution

$$\langle k \rangle(t) = \frac{\lambda_0}{\lambda_1 - \mu} \left(\mathrm{e}^{(\lambda_1 - \mu)t} - 1 \right). \tag{8}$$

In cases when $\lambda_1 > \mu$ the average cluster size grows indefinitely with time. When $\lambda_1 < \mu$, which is often found in our analyses (see Main text), average cluster size at the steady state is given by

$$\langle k \rangle^* = \frac{\lambda_0}{\mu - \lambda_1} = \frac{\theta_0}{1 - \theta_1}, \tag{9}$$

where θ_0 and θ_1 are defined after eqn. (6).

2.2.3 Sub-models assuming T cell-extrinsic clustering mechanisms (environment)

An alternative mechanism for the formation of T cell clusters around Plasmodium-infected hepatocytes is proposed in this paper, namely, that the formation of clusters is driven by the ability of different parasites to "attract" T cells. For example, some parasites while traveling from the blood to hepatocyte or while replicating in the hepatocyte may induce higher degree of inflammation than other parasites, thus, potentially increasing the chance of finding such "inflamed" sites by T cells. Indeed, sporozoites are able to induce inflammation in the liver [36].

We consider two versions of the "environment" model in which T cell recruitment to sites is determined by the variability in parasite's "attractiveness". In one such version, a **two population model**, we assume that there are parasites of two types found at frequencies f and $1 - f$, and these parasites differ in the rate at which T cells find them (Figure 1D). The formation of clusters around parasites of a given parasite type is given by random entry/exit model with rates λ_{01} and λ_{02} while the rate of exit of T cells from the cluster is $\mu_k = k\mu$. Then assuming a steady state the probability to observe clusters of size k is given by

$$P_k^* = f \frac{\theta_{01}^k}{k!} e^{-\theta_{01}} + (1 - f) \frac{\theta_{02}^k}{k!} e^{-\theta_{02}}, \tag{10}$$

where $\theta_{01} = \lambda_{01}/\mu$ and $\theta_{02} = \lambda_{02}/\mu$. Alternatively, the rate at which T cells find parasites could be given by a continuous function, and we tested a model in which entry rate into the cluster is given by a gamma distribution $g(\lambda_0; \alpha, \beta) = \frac{\beta^\alpha \lambda_0^{\alpha-1} e^{-\lambda_0 \beta}}{\Gamma(\alpha)}$, i.e., the probability for T cells to have an entry rate in the interval $(\lambda_0, \lambda_0 + \mathrm{d}\lambda_0)$ is $g(\lambda_0; \alpha, \beta)\mathrm{d}\lambda_0$. The probability to observe a cluster of size k given that clustering around a parasite "attracting" T cells at a rate λ follows a Poisson model is given by an integral

$$P_k^* = \int_0^\infty \frac{\lambda_0^k}{\mu^k k!} e^{-\lambda_0/\mu} \times \frac{\beta^\alpha \lambda_0^{\alpha-1} e^{-\lambda_0 \beta}}{\Gamma(\alpha)} d\lambda_0 = (\mu^{-1} + \beta)^{-(\alpha+k)} \frac{\beta^\alpha \Gamma(\alpha + k)}{\mu^k k! \Gamma(\alpha)}, \tag{11}$$

where α and β are the shape and rate parameters of the Gamma distribution, respectively, and $\Gamma(\alpha) = (\alpha - 1)!$.

2.2.4 A basic mathematical model for clustering of two cell types

In some of our experiments we tracked clustering of T cells of two specificities: one type of T cells was specific to Plasmodium sporozoites (PyTCR) and another type of T cells was specific to irrelevant antigen (OT1). To quantify the kinetics of clustering of Plasmodium-specific (PyTCR) and non-specific T cells (OT1) around Plasmodium-infected hepatocytes, we extended our basic model (eqns. (1)–(2)) to include two types of cells, t_1 and t_2, in the cluster to formulate a **co-clustering model** (Figure 1F). We define $P_{ij}(t)$ as the probability to observe i cells of type 1 and j cells of type 2 in a given cluster. Then the rate at which new T cells of type x, where $x = t_1, t_2$ enter the cluster with i T cells of type 1 and j T cells of type 2 is λ_{ij}^x. Similarly, μ_{ij}^x is the rate of exit of T cell of type x from a cluster with (i, j) T cells. The dynamics of the probability to observe a cluster with (i, j) T cells is given by equations

$$\begin{aligned}
\frac{\mathrm{d}P_{00}(t)}{\mathrm{d}t} &= -(\lambda_{00}^{t_1} + \lambda_{00}^{t_2})P_{00}(t) + \mu_{10}^{t_1} P_{10}(t) + \mu_{01}^{t_2} P_{01}(t), && (12)\\
\frac{\mathrm{d}P_{01}(t)}{\mathrm{d}t} &= -(\lambda_{01}^{t_1} + \lambda_{01}^{t_2} + \mu_{01}^{t_2})P_{01}(t) + \lambda_{00}^{t_2} P_{00}(t) + \mu_{11}^{t_1} P_{11}(t) + \mu_{02}^{t_2} P_{02}(t), && (13)\\
\frac{\mathrm{d}P_{10}(t)}{\mathrm{d}t} &= -(\lambda_{01}^{t_1} + \lambda_{01}^{t_2} + \mu_{10}^{t_1})P_{10}(t) + \lambda_{00}^{t_2} P_{00}(t) + \mu_{11}^{t_2} P_{11}(t) + \mu_{02}^{t_2} P_{02}(t), && (14)\\
\frac{\mathrm{d}P_{ij}(t)}{\mathrm{d}t} &= -(\lambda_{ij}^{t_1} + \lambda_{ij}^{t_2})P_{ij}(t) + \lambda_{(i-1)j}^{t_1} P_{(i-1)j}(t) + \lambda_{i(j-1)}^{t_2} P_{i(j-1)}(t)\\
&\quad + \mu_{(i+1)j}^{t_1} P_{(i+1)j}(t) + \mu_{i(j+1)}^{t_2} P_{i(j+1)}(t), \qquad i, j = 2..k_{\max} && (15)
\end{aligned}$$

The dynamics of the probability $P_{ij}(t)$ can be simulated by assuming different functional forms for the entry and exit rates (Table 1). For example, when the entry rate into the cluster is independent of the cluster size or cell type, $\lambda_{ij}^{t_1} = \lambda_{ij}^{t_2} = \lambda_0 = \text{const}$, and the exit rate is dependent on the number of T cells of a given specificity present near the parasite, $\mu_{ij}^{t_1} = i\mu$ and $\mu_{ij}^{t_2} = j\mu$ where $\mu = \text{const}$, clustering of T cells is independent and is described at the steady state by the Poisson distribution (results not shown). Another model is when the entry rate of T cells into the cluster is dependent only on the number of specific T cells (t_1) already in the cluster: $\lambda_{ij}^{t_1} = \lambda_{ij}^{t_2} = \lambda_0 + i\lambda_1$ with exit rates being similar to the random entry/exit model described above.

Model	$\boldsymbol{\lambda_{ij}^s}$	$\boldsymbol{\lambda_{ij}^{ns}}$	$\boldsymbol{\mu_{ij}^s}$	$\boldsymbol{\mu_{ij}^{ns}}$
Random entry/exit	λ_0	λ_0	μi	μj
Equal recruitment	$\lambda_0 + \lambda_1 \cdot (i+j)$	$\lambda_0 + \lambda_1 \cdot (i+j)$	μi	μj
Only specific T cells recruit	$\lambda_0 + \lambda_1 \cdot i$	$\lambda_0 + \lambda_1 \cdot i$	μi	μj
Only non-specific T cells recruit	$\lambda_0 + \lambda_1 \cdot j$	$\lambda_0 + \lambda_1 \cdot j$	μi	μj
Basic entry rates are type-specific and only specific T cells recruit	$\lambda_0^s + \lambda_1 \cdot i$	$\lambda_0^{ns} + \lambda_1 \cdot i$	μi	μj
Both T cell types recruit but with different rates	$\lambda_0 + \lambda_1^s \cdot i + \lambda_1^{ns} \cdot j$	$\lambda_0 + \lambda_1^s \cdot i + \lambda_1^{ns} \cdot j$	μi	μj
Both T cell types recruit but only their own type	$\lambda_0 + \lambda_1^s \cdot i$	$\lambda_0 + \lambda_1^{ns} \cdot j$	μi	μj

Table 1: Defining alternative models for co-clustering of Plasmodium-specific (s) and non-specific (ns) T cells around Plasmodium liver stages. For the general mathematical model describing co-clustering of two cell types (eqns. (12)–(15)) we define parameters determining the rate of T cell entry into the cluster (λ_{ij}^s and λ_{ij}^{ns}) and the rate of exit from the cluster (μ_{ij}^s and μ_{ij}^{ns}) where superscripts "s" and "ns" stand for Plasmodium-specific and non-specific T cells, respectively, ij denotes a cluster with i specific and j nonspecific T cells around a given liver stage. For example, λ_{ij}^s denotes the entry rate of specific T cells into a cluster with i specific and j non-specific T cells. Parameters λ_0 (the initial entry rate), μ (per capita exit rate), and λ_1 (increase in entry rate with cluster size) are found by fitting the numerical solution of the mathematical model (given in eqns. (12)–(15)) to the co-clustering data (dataset #3).

2.2.5 Stochastic simulations

We simulated cluster formation using the Gillespie algorithm as previously described [43]. In short, for every iteration we first determined randomly the time of the change in cluster size which is determined by the total rate at which clusters could increase or decrease in size (e.g., in the DD recruitment model this rate for a cluster of size k is $\lambda_0 + k\lambda_1 + k\mu$). The second step was to then choose at random which of two events (cluster size increase or decrease) occurs; this is determined by the relative value of the entry rate into the cluster (e.g., $\lambda_0 + k\lambda_1$) or exit from the cluster (e.g., $k\mu$).

2.2.6 Statistics

Our clustering data are given as the number of T cells found in the 40 μm radius from a given parasite following intravital microscopy imaging [29], i.e., the data are simply a column of integers representing T cell numbers per parasite. (In co-clustering experiments the data also represent the number of Plasmodium-specific and non-specific T cells found per parasite.) As the data shows, in many cases majority of parasites have no T cells associated with them [29, and see main text].

To estimate parameters of mathematical models we used a likelihood approach where the likelihood represents the product of probabilities to observe clusters of different sizes

$$L(parameters|data) = P(data|parameters) = \prod_{k=0}^{k_{\max}} (P_k)^{x(k)}, \tag{16}$$

where P_k is the mathematical model-predicted probability of observing a cluster of size k according to a set of parameter values, $x(k)$ is the number of clusters of size k in the data, and $k_{\max}$ is the maximal cluster size in the data. In this procedure, the probability $P_k(t)$ can be either given analytically as a steady state solution (e.g., eqn. (4)) or can be found by numerically solving the basic mathematical model predicting $P_k(t)$ at a particular time (e.g., eqns. (1)–(2)). When fitting numerical solutions of the model to experimental data in some cases we fixed the rate of exit of T cells from clusters μ to different values because we found that it is generally impossible to accurately estimate both entry and exit rates simultaneously (see Results section).

The models were fitted by calculating negative log-likelihood $\mathcal{L} = -\log L$ and using routine `FindMinimum` in `Mathematica` version 11. When alternative models were fitted to the same dataset, we compared quality of the model fits to the data by comparing Akaike weights w based on the corrected Akaike Information Criterion (AIC) [44]:

$$\mathrm{AIC} = -2\log L + 2p + \frac{2p(p+1)}{N-p-1}, \tag{17}$$

where p is the number of model parameters and N is the number of data points (parasites). AIC provides a score for each model based on its maximum likelihood value and the number of model parameters. The model with the lowest AIC score is considered to be best relative to the tested models. Weights of a given model can be treated as a likelihood of the model in the list of the tested models. As a rule of thumb, models with $w < 0.05$ can be considered to be inconsistent with experimental data in favor for models with higher weights. In addition, when comparing nested models, where one model is a special case of another, we used the likelihood ratio test.

3 Results

3.1 Pertussis toxin-treated PyTCR cells form clusters randomly

By using intravital microscopy two recent studies have observed that activated CD8 T cells, specific to Plasmodium antigens, often localize near Plasmodium liver stages [29, 30]. In our experiments we used GFP-expressing *Plasmodium yoelii* (Py) sporozoites and fluorescently-labeled activated Py-specific CD8 T cells (PyTCR cells) [29]. A peculiar feature of these observations was a variable number of T cells found near individual parasites: in many experiments multiple parasites had no T cells while some parasites had more than 20 T cells in a 40 μm radius [29]. We have developed three alternative mathematical models that assume that the formation of clusters of CD8 T cells around Plasmodium liver stages is T-cell-intrisitic and is mediated exclusively by T cells. These models include 1) the null model predicting formation of clusters by chance ("random entry/exit" or Poisson model, eqn. (4)), 2) a model in which T cells find parasites randomly but are retained near the parasite ("density-independent (DI) exit" model, eqn. (5)), and 3) a model in which T cells actively recruit other T cells to the infection site ("density-dependent (DD) recruitment" model, eqn. (6)). By fitting these models to experimental data on clustering we concluded that the DD recruitment model is best consistent with data on clustering of PyTCR cells [29]. By performing experiments in which either PyTCR cells or PyTCR cells, treated with pertussis toxin (PyTCR+PT cells), were transferred into mice, previously infected with Py, we found that PT-treated PyTCR cells form smaller clusters that untreated T cells (Figure 2). PT treatment inactivates G-protein-coupled receptors (GPCRs) in

the cells, and in particular, makes PT-treated cells to be unresponsive to chemokine gradients [45]. Fitting three basic mathematical models, predicting steady state distribution of T cells around the liver stages, to the data on clustering of PyTCR cells (see Materials and Methods for more detail on how models were fitted to data) confirmed our previous result that DD recruitment model provided the best fit of clustering data (Figure 2B). The random entry/exit model did not fit these data adequately, while the DI exit model provided a visually reasonable fit which was not supported by the Akaike weights or log-likelihood (result not shown). Interestingly, plotting the data and model fits on the log-scale revealed that even the best fit model did not accurately predict formation of a large cluster with 21 PyTCR cells (Figure 2B). Indeed, comparing the prediction of the DD recruitment model with data using a goodness-of-fit χ^2 test showed that the model described the clusters until size 8 well ($\chi^2_{18} = 7.2$, $p = 0.99$), but failed at describing all clusters including one with 21 cells ($\chi^2_{19} = 51.6$, $p < 0.001$). According to the prediction of the mathematical model, the average cluster size of T cells around the liver stages at the steady state is rather small, $\langle k \rangle^* \approx 1.5$ (see eqn. (9)), because most parasites were not found by T cells within 6 hours (Figure 2A).

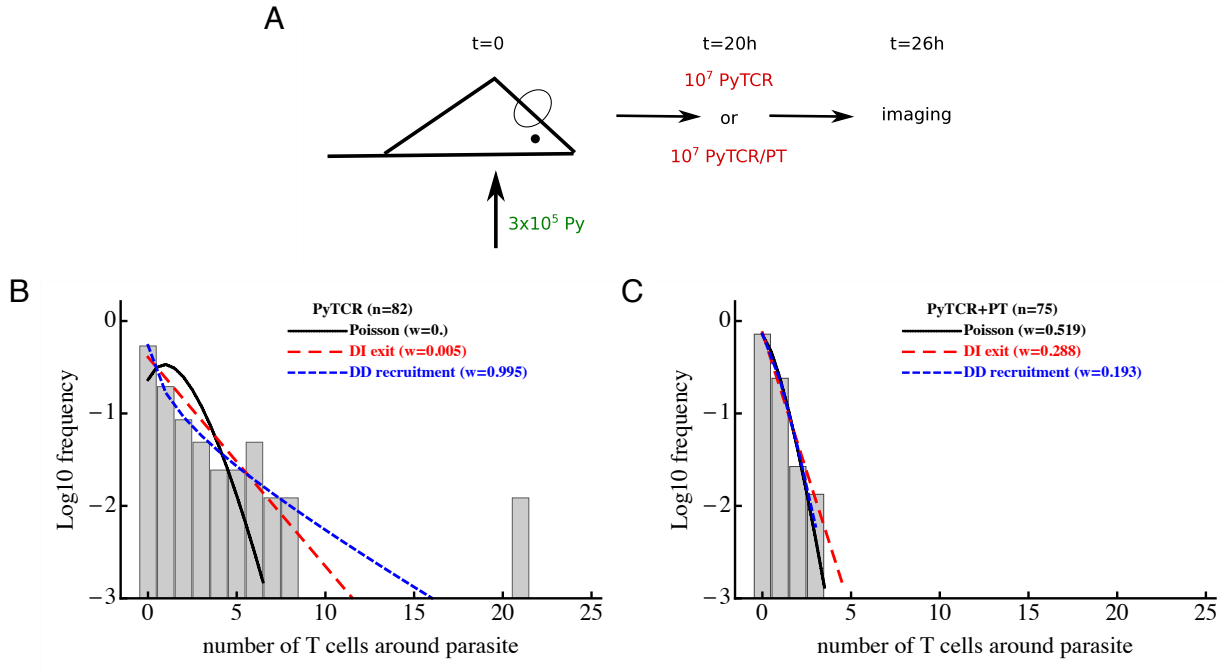

Figure 2: Pertussis toxin (PT)-treated T cells form clusters around Plasmodium yoelii (Py)-infected hepatocytes randomly. Panel A: mice were infected with 3×10^5 GFP-expressing Py sporozoites. Twenty hours later 10^7 Py-specific activated CD8 T cells (PyTCR) or 10^7 PyTCR T cells pretreated for one hour with 1 μg/mL PT were transferred into infected mice and imaged with intravital microscopy 6 hours later [29]. The number of T cells in 40 μm radius of randomly chosen $n = 82$ (PyTCR) or $n = 75$ (PyTCR+PT) parasites was recorded; the frequency of size of such clusters is shown in panels B-C by bars. Panels B and C: we fitted three mathematical models (Poisson (eqn. (4)), DI exit, (eqn. (5)), and DD recruitment (eqn. (6))) to these experimental data using likelihood approach (eqn. (16)) and calculated Akaike weights (w) (see Materials and Methods for more detail). Results suggest that clustering of PyTCR T cells is best described by density-dependent recruitment model while clustering of PT-treated T cells occurs mainly randomly. Parameter estimates of the best fit model and 95% CIs obtained by resampling the data with replacement 10^3 times in panel B are $\theta_0 = 0.31$ $(0.20 - 0.48)$ and $\theta_1 = 0.79$ $(0.59 - 0.88)$ (DD recruitment model) and in panel C are $\theta_0 = 0.33$ $(0.21 - 0.48)$ (Poisson model).

Importantly, however, all three models could accurately describe the formation of clusters by

PT-treated PyTCR cells (Figure 2C) suggesting that the formation of these clusters is most likely due to a random encounter between T cells and the Py-infected hepatocyte. Indeed, the estimated relative entry rate θ_0 was similar for two datasets (see legend of Figure 2) and was only slightly higher than the relative rate estimated in our previous work from other experiments involving co-transfer of PyTCR and OT1 cells [29, see Materials and Methods and below]. Thus, this result further supports the conclusion that GPCR-mediated signaling is important in the formation of CD8 T cell clusters around Plasmodium liver stages.

3.2 Clustering of endogenous CD8 T cells does not allow to discriminate between T-cell-intrinsic and T-cell-extrinsic models of cluster formation

Our previous analyses so far attempted to explain mechanisms behind the formation of clusters around Plasmodium liver stages from the T-cell-centric point of view; namely, we assumed that cluster formation is dependent on the presence of T cells (e.g., DD recruitment model, see eqn. (6)). However, it is possible that a very different alternative mechanism drives the formation of clusters, which is T-cell-extrinsic. In this case, variable clustering of T cells around the liver stages is driven by variability in the environment, for example, due to the level of "attractiveness" of individual parasites. This may arise because individual parasites may induce different degrees of inflammation as they migrate from the blood into the liver parenchyma, or some parasites may infect hepatocytes which are located in liver parts with a larger blood flow, increasing the chance of T cells to locate such parasites [35, 46].

To investigate whether a T-cell-extrinsic mechanism can be sufficient to explain the formation of clusters in our experiments we formulated two alternative mathematical models predicting the formation of clusters of different sizes: in the first model we assumed that there are two populations of parasites with different levels of attractiveness/rate of entry λ_{01} and λ_{02} ("2 population" model, Figure 1D and eqn. (10)), and in the second model we assumed that there is a distribution in the level of attractiveness of parasites given by a continuous Gamma distribution ("gamma" model, Figure 1E and eqn. (11)). To test whether models assuming T-cell-intrinsic or T-cell-extrinsic mechanisms of cluster formation perform better, we fitted the models to previously published data on clustering of endogenous CD8 T cells around Py liver stages in mice [29]. In these experiments, mice were left naive or were immunized with radiation-attenuated sporozoites (RAS) and then 10 days later infected with wild-type Py expressing GFP (Figure 3A). Clustering of CD8 T cells around GFP-expressing liver stages was visualized by injecting CD8-binding antibody. We fitted five mathematical models to these data using a likelihood approach (eqn. (16)). This analysis showed that all mathematical models could accurately describe the lack of formation of large ($k > 5$) clusters around Plasmodium liver stages in naive (unimmunized) mice, and the simplest (null) random entry/exit model was favored by the Akaike weights (Figure 3B). While all models provided similar likelihood of the model given the data, lower weights for 2 population and gamma models was due to a larger number of fitted parameters (3 in 2 population/gamma models vs. 1 in the Poisson model).

Consistent with our previous analysis [29], the random entry/exit (Poisson) model could not adequately describe the distribution of cluster sizes in RAS-immunized mice (Figure 3C). Interestingly, both 2 population and gamma models did not provide an improved fit of these clustering data as compared to DD recruitment or DI exit models which fitted the data with similar quality (Figure 3C). This was surprising given that the DD recruitment and DI exit models were not able to accu-

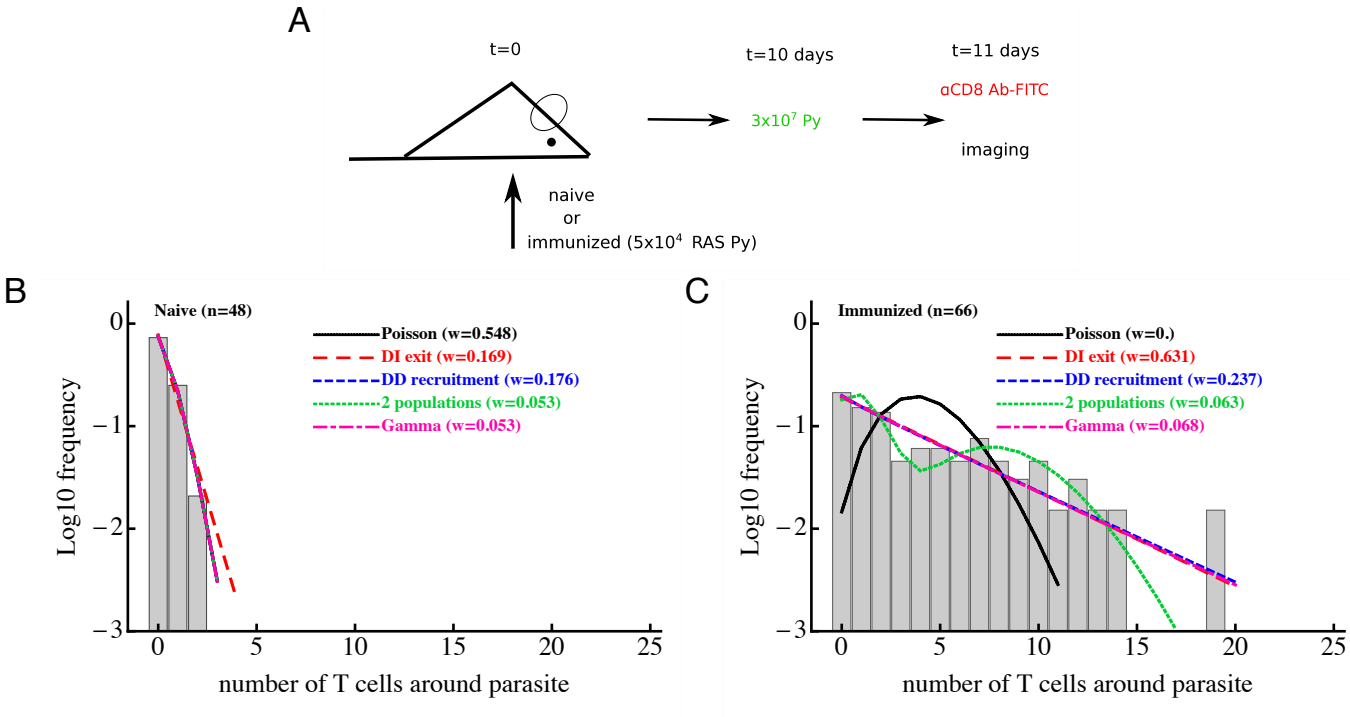

Figure 3: Models assuming time-invariant but spatially-variable environment are consistent with the data on clustering of CD8 T cells in mice immunized with radiation-attenuated sporozoites (RAS). Panel A: mice were immunized with 5×10^4 Py RAS or left unimmunized. Ten days later, mice were infected with 3×10^7 wild-type Py, expressing GFP. One day later CD8 T cells were labeled with 4 μg PE-conjugated anti-CD8 mAbs and clustering of CD8 T cells around Py-infected hepatocytes in the liver was imaged using intravital microscopy [29]. In total 48 (in naive mice, panel B) and 66 (in RAS-immunized mice, panel C) parasites were randomly chosen and the number of T cells in a 40 μm radius were counted. Five different mathematical models were fitted independently to the data on T cell clustering in naive and immunized mice, and the quality of the model fits was evaluated using Akaike weights (w). Clustering in naive mice is most consistent with the Poisson (random entry/exit) model, while in RAS-immunized mice models assuming constant environment ("2 populations" and "gamma" models) fit the data worse than other models, in part due to a larger number of parameters than in the DD recruitment or DI exit models. Parameter estimates of the best fit model and 95% CIs in panel B are $\theta_0 = 0.29$ $(0.17 - 0.42)$ (Poisson model) and in panel C are $\theta_0 = 0.81$ $(0.77 - 0.84)$ (DI exit model) or $\theta_0 = 0.80$ $(0.52 - 1.16)$ and $\theta_1 = 0.82$ $(0.73 - 0.88)$ (DD recruitment model). According to the DD recruitment model, the average cluster size in RAS-immunized mice at steady state (panel B) is $\langle k \rangle^* \approx 4.2$.

rately describe the two peaks in the cluster size distribution (at 0 and 7 T cells/parasite). A closer inspection revealed that the DD recruitment, DI exit, 2 population, and gamma models provided fits of nearly identical quality as based on the negative log-likelihood values ($\mathcal{L} \approx 168$), and lower weights were selected for models with more parameters. All models except the Poisson and 2 population models could accurately describe the data (based on goodness-of-fit χ^2 test); the 2 population model deviation was due to its inability to accurately predict the formation of one cluster with 19 cells (results not shown). Importantly, the 2 population and gamma models could fit other clustering data relatively well, e.g., data on clustering of PyTCR cells or PyTCR cells treated with PT (Figure 2, results not shown). Taken together, these results demonstrate that some clustering datasets do not allow to discriminate between T-cell-intrinsic and T-cell-extrinsic mechanisms of formation of CD8 T cell clusters around Plasmodium liver stages.

3.3 Environmental variability is not the main driver of cluster formation

To discriminate between T-cell-intrinsic and T-cell-extrinsic mechanisms of formation of CD8 T cell clusters around Plasmodium liver stages we turned to additional experimental data generated previously [29]. In these experiments, Py-specific T cells and T cells of irrelevant specificity (OT1) were transferred either separately or together into mice previously infected with Py-GFP, and the formation of clusters around Py liver stages was measured by intravital microscopy (Figure 4A). We have previously shown that the DD recruitment model describes best (based on Akaike weights) the data on the clustering of PyTCR cells when transferred alone or data on the clustering of PyTCR and OT1 cells when transferred together [29]. In contrast, the clustering of OT1 cells alone was best described by the random entry/exit model [29, Figure 4B-C]. Therefore, these data indicate that the clustering of T cells, which are not specific to Plasmodium depends on the presence of Py-specific T cells suggesting that variability in parasite's "attractiveness" alone cannot explain these data. We formally tested if the 2 population or gamma models can describe the clustering data of OT1 cells in the following way. We fitted the 2 population model to the data on clustering of OT1 cells alone or in the mixture with PyTCR cells simultaneously. We therefore fitted the models by allowing all three parameters of the model (θ_{01}, θ_{01}, and f, see eqn. (10)) to be different for the two datasets or by allowing only the fraction of parasites with different attractiveness level f to vary between two datasets while keeping other parameters the same between datasets (Figure 4B-C). In this way, we tested the hypothesis that the clustering of OT1 cells is driven exclusively by factors which are independent of Py-specific CD8 T cells. Because two fits are from nested models, comparing the quality of the fits revealed that the model assuming PyTCR-cell-independent environment fits the two datasets significantly worse ($\chi^2_1 = 12.4$, $p < 0.001$). Fitting the gamma model to the same two datasets assuming either identical or variable parameters between the two datasets also suggested that the model with constant parameters fits the data significantly worse (results not shown). Thus, these results strongly suggest that the T-cell-extrinsic models of cluster formation are not consistent with the data on different clustering patterns of OT1 cells in the absence or presence of Py-specific T cells.

It is important to note that the use a specific mathematical model (e.g., 2 population model) simply allows to formally test if distributions of cluster sizes of OT1 cells are different in two different conditions. This can be also done using a χ^2 test [47] which showed that these distributions are only marginally different ($\chi^2_8 = 16.1$, $p = 0.04$). Thus, the use of models allows to obtain much stronger statistical power at falsifying the T-cell-independent ("environment") hypothesis as the sufficient mechanism of cluster formation.

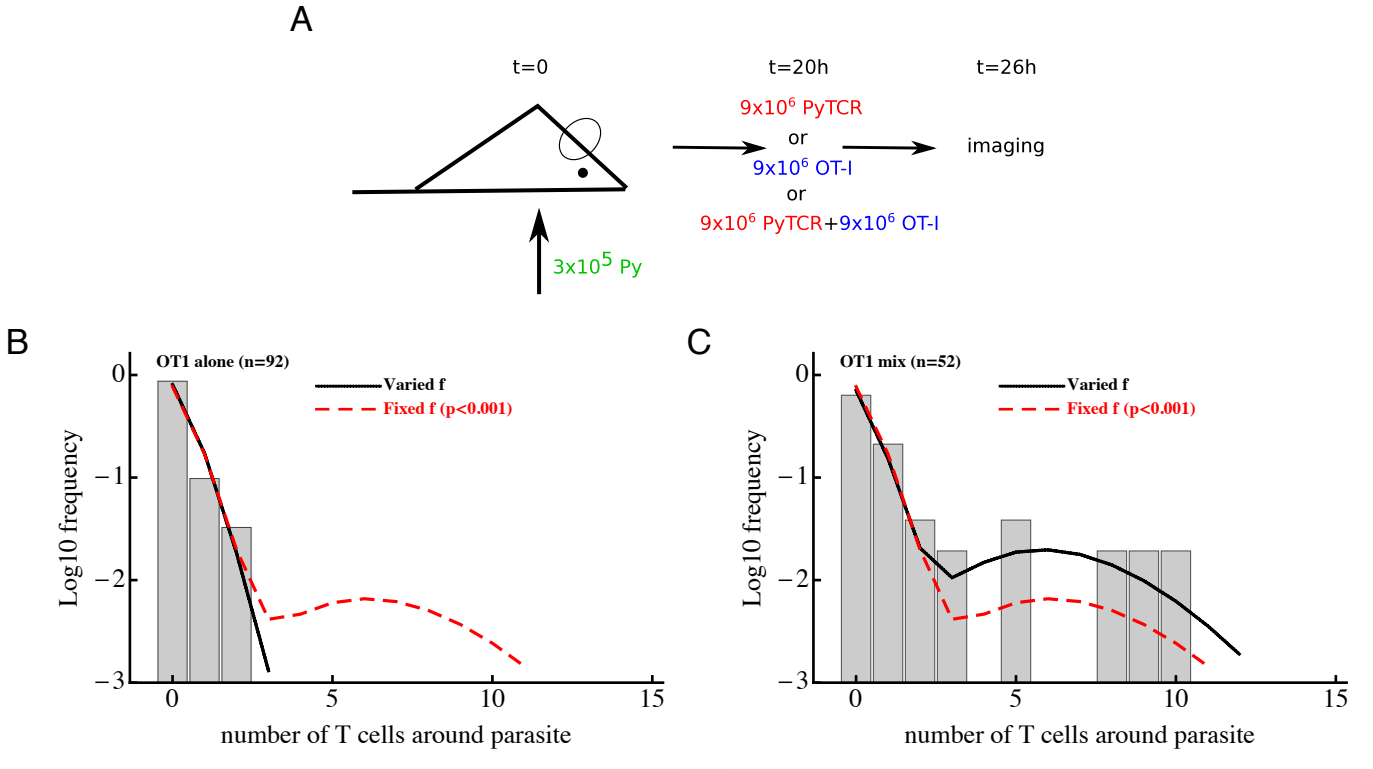

Figure 4: Models assuming time-invariant but spatially-variable environment are unable to accurately describe the clustering of T cells of irrelevant specificity in different conditions. Panel A: mice were infected with 3×10^5 GFP-expressing Py sporozoites. Twenty hours later 9×10^6 Py-specific activated CD8 T cells (PyTCR), 9×10^6 OT1 T cells (specific to chicken ovalbumin), or mixture of 9×10^6 PyTCR and 9×10^6 OT1 T cells were transferred into infected mice and livers of these mice were imaged using intravital microscopy 6 hours later. In total 92 (mice receiving only OT1 cells, panel B) and 52 (in mice receiving a mix of PyTCR and OT1 cells, panel C) parasites were randomly chosen and the number of T cells in a 40 μm radius were counted [29]. The "two population" mathematical model (eqn. (10)) was fitted to these two datasets simultaneously assuming two different entry rates θ_{01} and θ_{02} and either allowing the fraction of attracting parasites f to vary between the datasets (solid line) or to be fixed between the datasets (dashed line). Fixing the fraction f between the datasets significantly reduced the quality of the model fit of the data as compared to the model in which f could vary (likelihood ratio test, $\chi_1^2 = 12.4$, $p < 0.001$).

3.4 Several alternative retention models poorly describe data on clustering of PyTCR cells

While our experiments of clustering of OT1 T cells either alone or in presence of PyTCR T cells argue against T-cell-extrinsic clustering model, they do not allow to fully discriminate between alternative T-cell-intrinsic clustering models. Fitting the steady state prediction of the DI exit and DD recruitment model to clustering of PyTCR T cells (Figure 2) or clustering of OT1 T cells in the presence of PyTCR cells [29] favored the DD recruitment model (based on Akaike weights). However, it is possible that a specific form of the retention model, i.e., that the per capita exit rate is inversely proportional to the cluster size, was an accidentally poor choice. Therefore, we tested two alternative models of how exit rate from a cluster could depend on cluster size with $\mu_k = k^\alpha \mu$ or $\mu_k = k\mu e^{-\alpha k}$. We fitted the numerical solution of the basic mathematical model (eqns. (1)–(2)) to the clustering of PyTCR T cells (Figure 2A) using a likelihood approach. Both alternative retention models still described the data worse than the DD recruitment model ($w < 0.001$ and $w = 0.02$ for the two models, respectively, results not shown) suggesting limited support for the hypothesis that retention of T cells plays the major role in cluster formation. Therefore, in our following analyses we focus exclusively on the DD recruitment model.

3.5 No evidence that T cells of irrelevant specificity influence clustering

In our previous analysis we showed that the DD recruitment model-based fit of the data on the clustering of PyTCR and OT1 cells in the co-transfer experiments (Figure 4A) predicted similar relative recruitment rate parameters θ_0 and θ_1 (see Table S1 in Cockburn *et al.* [29]). However, the previous analysis treated clustering of PyTCR and OT1 cells in the co-transfer experiments independently, and here we extend this analysis by considering potential mechanisms behind co-clustering of these two cell populations. First, we found that there was no significant difference in the number of PyTCR or OT1 T cells clusters around Plasmodium liver stages with similar proportions of parasite having more PyTCR or OT1 cells (Figure 5A). To investigate whether the data on the co-clustering of T cells may provide evidence of OT1 T cells assisting in cluster formation we developed a mathematical model tracking the dynamics of co-clustering of two types of cells (see eqns. (12)–(15) in Material and Methods) and fitted that model to the co-clustering data (dataset #3) using a likelihood approach. As we show in the next section, our clustering data do not allow to identify both the rate of T cell entry into the cluster and exit rate from the cluster from measuring clusters at one time point. Therefore, in this analysis we fixed the per capita exit rate $\mu = 0.5$/h and estimated entry rates. Our overall results were robust to several other tested values of the exit rate such as $\mu = 0.1$/h or $\mu = 3$/h even though estimates of the entry rates were strongly dependent on the assumed exit rate (results not shown and see next section).

Using the DD recruitment model we tested several different mechanisms of how specific and non-specific T cells may participate in cluster formation (see Table 1). Despite the highly correlated numbers of the Py-specific and non-specific T cells around Plasmodium liver stages (Figure 5A), different roles of these two CD8 T cell types seem to be inherent in the data (Table 2). Specifically, the model in which PyTCR T cells attract all cells into the cluster was statistically better at describing these data as compared to any other model tested (based on Akaike weights); interestingly, an alternative model in which OT1 cells exclusively drive cluster formation could not fit the co-clustering data well (model "Only OT1 cells recruit" in Table 2).

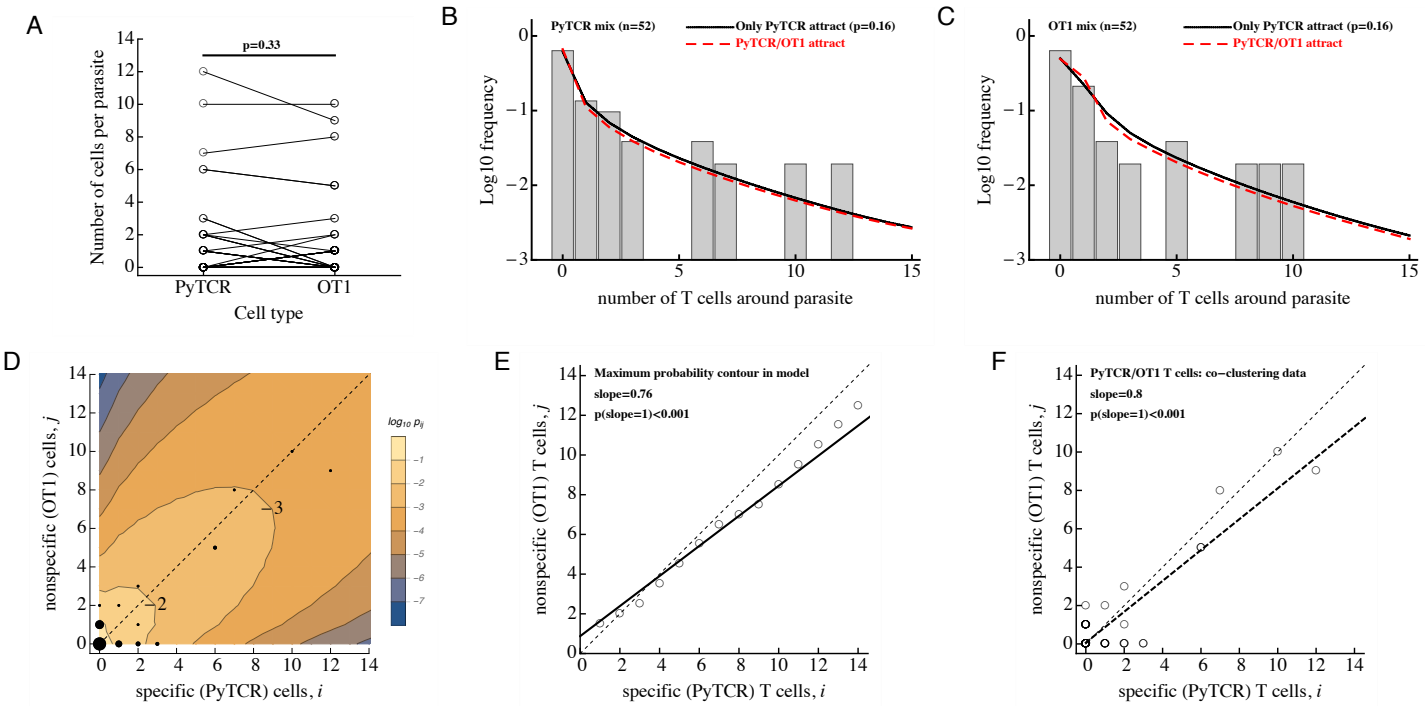

Figure 5: No evidence that activated CD8 T cells of irrelevant specificity play a significant role in cluster formation. Experiments were performed as described in Figure 4A and the number of T cells, specific to Py (PyTCR) and T cells of irrelevant specificity (OT1), found in a 40 μm radius of $n = 52$ randomly chosen parasites in the liver was counted using intravital imaging. Panel A: no difference in the number of Py-specific and non-specific T cells found around Py liver stages (Wilcoxon signed rank test). Panels B-C: we fitted a series of mathematical models assuming how Py-specific or nonspecific T cells mediate attraction to the infection site (co-clustering model, eqns. (12)–(15)), and fits of two models where either only PyTCR T cells attract (solid lines, B-C) or both PyTCR and OT1 T cells attract (dashed lines, B-C) as well as data (bars) are shown. A simpler model in which only PyTCR T cells mediate attraction describes the data as well as the more complex model (likelihood ratio test, $\chi^2_1 = 1.95$, $p = 0.16$). Bars in panels B-C are the observed frequencies of parasites with a variable number of PyTCR (panel B) or OT1 (panel C) T cells. In panel D we show the data (points) and predictions of the model in which only Py-specific T cells attract all activated cells to the infection site (contours); model prediction is the $\log_{10} P_{ij}(6)$ where P_{ij} is the probability to observe i Py-specific and j nonspecific T cells around the parasite at $t = 6$h after T cell transfer. The point size represent the number of parasites having a given number of PyTCR/OT1 cells in 40 μm radius, and thin dashed lines shows the line with slope = 1. In panel E points represent the prediction of the model on the number of PyTCR/OT1 T cells in clusters where P_{ij} reaches maximum for $i+j = \text{const}$ (see the contour plot in panel D). The solid line represents a regression line for the model predictions with slope = 0.76 which is significantly different from 1 (t-test, $p < 0.001$) and the dashed line shows the line with slope = 1. In panel F points represent co-clustering data of PyTCR and OT1 T cells around parasites (the same data are shown in panel D). Solid line represents a regression line with slope = 0.8 which is significantly different from 1 (t-test, $p < 0.001$). Results in panels D-F indicate model-predicted slight bias towards clustering of a larger number of PyTCR T cells which is not directly observed in the data (panel A).

In two separate models we tested whether OT1 cells "help" in the formation of clusters which is driven by Py-specific T cells. Perhaps unsurprisingly in both models ("PyTCR and OT1 cells recruit at different rates" and " PyTCR and OT1 cells recruit at different rates towards different cell types") we found no evidence that OT1 cells enhance cluster formation (Table 2 and Figure 5B-C).

Model	λ_0, 1/h	λ_1, 1/h	$\mathcal{L}$	AIC	w
Only PyTCR cells recruit	0.14	0.58	128.0	260.3	**0.40**
Only OT1 cells recruit	0.15	0.60	131.4	267.8	0.01
PyTCR and OT1 cells recruit at the same rate	0.12	0.32	130.7	265.6	0.03
PyTCR and OT1 cells recruit at different rates	0.17	PyTCR=0.67, OT1=-0.16	127.1	260.6	0.34
PyTCR and OT1 cells recruit at different rates towards different cell types	0.18	PyTCR:PyTCR= 0.73, PyTCR:OT1= 0.61, OT1:OT1= -0.15, OT1:PyTCR= -0.23	126.7	264.8	0.04

Table 2: Comparing alternative models which assume different contributions of Py-specific (PyTCR) and non-specific (OT1) T cells to cluster formation. We fit the basic mathematical model on co-clustering of Plasmodium-specific and non-speciific T cells (eqns. (12)–(15)) to the data on T cell clustering around Py liver stages assuming DD recruitment model and different mechanisms of how T cells contribute to cluster formation (see Table 1 for tested models). Here we list the estimated initial recruitment rate λ_0 and how recruitment rate changes with cluster size λ_1 (i.e., in the DD recruitment model the recruitment rate is $\lambda_k = \lambda_0 + k\lambda_1$), the negative log-likelihood $\mathcal{L}$, AIC, and Akaike weights w for the model fit. In these fits the total exit rate of T cells from the cluster of size k was fixed to $\mu_k = 0.5k$/h. In the column with estimates for λ_1 we list specifically the predicted change in the cluster "attractiveness" by a given type of T cell (specific or non-specific) and towards a given type of T cells. For instance, an estimate $\lambda_1 = 0.58$/h for the model in which only PyTCR cells recruit assumes that PyTCR cells recruit specific and nonspecific T cells at the same rate. In another model notation "PyTCR:OT1" denotes the recruitment rate induced by PyTCR cells for OT1 cells.

In contrast, parameter estimates suggested that OT1 cells may inhibit cluster formation because the estimated OT1-driven recruitment rates λ_1 were negative (Table 2); however, improvements of the fits of these two more complicated models were not supported by the likelihood ratio test ($p > 0.1$, see Figure 5B-C). Thus, our results suggest that non-specific T cells are "passive" participants in the clusters and do not significantly promote or impede the formation of clusters. A similar result was obtained recently using another Plasmodium experimental system [31].

Predictions of our best mathematical model in which only PyTCR cells recruit all activated T cells to the site of infection can be shown as the distribution of cluster sizes for each cell type (e.g., Figure 5B-C) as well as the probability to observe a cluster with i PyTCR and j OT1 cells (Figure 5D). Careful examination of this fit revealed that the model predicts a slight bias towards having more PyTCR cells per cluster than OT1 cells (Figure 5E). Linear regression analysis of the co-clustering data indeed suggests that there may be bias towards having more PyTCR cells than OT1 cells per cluster (Figure 5F); however, this result is not fully consistent with another analysis (e.g., Figure 5A), and the application of linear regression to data with integers may not fully appropriate. While the existence of such a bias is indeed in line with the analytical analysis of the steady state distribution of cluster sizes (see Supplement for mathematical proof), this bias is small (perhaps one extra PyTCR cell in clusters of a total size of 20), and biological relevance of such a bias for the killing of the parasite is most likely limited.

3.6 Clusters of Py-specific CD8 T cells around Py-infected hepatocytes are formed rapidly

Our analyses so far made an assumption that clusters around Plasmodium liver stages reach a steady state by 6-24 hours after T cell transfer. To understand potential limitations of this approach, we therefore performed several analyses using previously published experimental data.

Because our main mathematical model of T cell clustering (eqns. (1)–(2)) can be solved numerically and thus fitted to experimental data assuming a specific clustering mechanism (e.g., DD recruitment model), we investigated if the rates of T cell entry into the cluster (λ_0 and λ_1) and rates of exit from the cluster (μ) can be estimated from data in which PyTCR cell clusters around Py-infected hepatocytes were observed at 6 hours after T cell transfer. Interestingly, fitting the DD recruitment model (eqns. (1)–(2)) to data on the clustering of PyTCR cells transferred alone (Figure 2B or Figure 5A) revealed that model fits favored very high entry and exit rates, e.g., rates exceeding 20-30/h (results not shown). By fixing the exit rate from the cluster to multiple values we found that estimates of the absolute and relative values of the entry rate depended strongly on the exit rate values, and the relative entry rates (θ_0 and θ_1) approached constant values at high exit rates (Table 3). Importantly, all the fits of models with dramatically different exit rates were of nearly identical quality as based on negative log-likelihood suggesting that data on clustering of T cells at one time point are not sufficient to estimate entry and exit rates simultaneously. These results were confirmed for two independent datasets (experiments with PyTCR cells alone as shown in Figure 2A and Figure 5A) although exact values of parameter estimates such as λ_0 did slightly vary between two sets of experiments (see Figure 2 and estimates in Table S1 in [29]).

	μ=**0.06/h**	μ=**0.1/h**	μ=**0.3/h**	μ=**1/h**	μ=**3/h**
Estimated θ_0	1.45 (1.10-1.88)	0.92 (0.70-1.23)	0.41 (0.30-0.51)	0.24 (0.17-0.31)	0.20 (0.14-0.28)
Estimated θ_1	6.31 (4.37-7.71)	4.02 (2.77-4.91)	1.77 (1.29-2.09)	1.04 (0.84-1.17)	0.89 (0.77-0.98)

Table 3: Estimated relative entry rates in the DD recruitment model (θ_0 and θ_1) strongly depend on the value of assumed exit rate from the cluster μ. We fixed the value of the rate of exit of T cells from the cluster μ to different values (indicated in the top row) and fitted the DD recruitment model (with $\lambda_k = \lambda_0 + \lambda_1 k$ and $\mu_k = k\mu$, see eqns. (1)–(2)) to experimental data on clustering of PyTCR T cells ($n = 130$) when transferred alone (cluster formation was observed 6 hours after T cell transfer, see Figure 5A). The quality of model fits to data as judged by the negative log-likelihood were nearly identical between different fits ($\mathcal{L} \sim 194$); the fit of the model with $\mu = 3$/h is shown in our previous publication (see Figure S1A and Table S1 in Cockburn *et al.* [29]). To compare parameter estimates we show relative entry rates $\theta_0 = \lambda_0/\mu$ and $\theta_1 = \lambda_1/\mu$. Shown 95% confidence intervals for parameter estimates were obtained by bootstrapping the cluster data for individual parasites 1000 times. Interestingly, the ratio θ_1/θ_0 was relatively constant for different fits [48, and ms. in preparation].

To gain further insights into the kinetics of T cell cluster formation we analyzed additional data in which the same parasites ($n = 32$) were followed after T cell transfer over time and cluster sizes at different time points were recorded (Figure 6A and Figure S2). In these experiments imaging started between 4 and 8 hours after T cell transfer and followed for about 4 hours [29]. As expected there was a variable and statistically significant increase in the number of T cells found around individual Py-infected hepatocytes between T cell transfer and start of imaging (t_{start}, Figure 6B). In contrast, in the following $\sim 2-8$ hours there was a minor change in cluster sizes (t_{end}, Figure 6B). However,

because imaging of CD8 T cell cluster formation started at different time points after T cell transfer there may be biases associated with the simple analysis of the data which takes into account only start and end time points of the clusters (e.g., Figure 6B). To obtained more accurate insights we further analyzed these data using mathematical models of cluster formation.

To take full advantage of these "longitudinal" data in which T cell cluster formation was followed over time for individual parasites (Figure S2), we divided the data into individual "paths", i.e., the number of T cells found near the parasite at sequential time points. For example, a parasite which did not have any T cells nearby and for which measurements were done at 0, 4 and 6.2 hours after T cell transfer, the path is "$0 \to 0 \to 0$". For the parasite that was surrounded by most T cells in these experiments, the path is "$0 \to 8 \to 9$" for times 0, 4, and 6.2 h post T cell transfer (Figure 6B). A mathematical model of cluster formation can then be used to calculate the likelihood of a particular path by assuming that individual "sub-paths" along the path are independent (and thus by multiplying likelihoods of the model for individual sub-paths). For example, the probability to observe the path "$0 \to 8 \to 9$" at times $(0, 4, 6.2)$ h is simply the product of the probability to observe 9 T cells in the cluster at 6.2 hours given that at 4 hours there were 8 T cells in the cluster and the probability to observe 8 T cells in the cluster at 4 hours given that at 0 hours there were 0 T cells in the cluster:

$$p_{\text{path}} = P_9(6.2|k = 8, t = 4) \times P_8(4|k = 0, t = 0), \tag{18}$$

where the probability $P_k(t|i, t_0)$ was calculated using the basic model (see eqns. (1)–(2)) with initial conditions $P_i(t_0) = \theta_{ij}$ and θ_{ij} is Kronicker delta ($\theta_{ij} = 1$ if $i = j$ and $\theta_{ij} = 0$, otherwise). Fitting the DD recruitment model to these "longitudinal" data subdivided into "paths" resulted in the following entry/exit rates $\lambda_0 = 0.14/\text{h}$, $\lambda_1 = 0.16/\text{h}$, $\mu = 0.09/\text{h}$. Additional analysis by fixing exit rate μ to different values and then comparing quality of the model fit using likelihood ratio test revealed that estimate of the parameters are relatively robust (i.e., fixing the exit rate to much lower or much higher values resulted in fits lower quality as judged by likelihood ratio test). Furthermore, by resampling the paths with replacement we found relatively small confidence intervals for the estimated parameters suggesting that measurement of T cell clusters longitudinally allows for a relatively accurate estimates of all three parameters of the DD recruitment model determining the kinetics of cluster formation (results not shown).

Parameter estimates of the model fitted to "longitudinal" (paths) data suggest that rates of entry into the cluster and exit from the cluster are relatively small, and this appears to contradict the formation of relatively large clusters already in 4 hours after T cell transfer (Figure 6C). Indeed, model fits did not accurately predict formation of clusters with > 5 T cells (results not shown). In addition, while the estimate of $\theta_0 = \lambda_0/\mu$ was reasonable, the estimate of relative recruitment rate $\theta_1 = \lambda_1/\mu$ was too low when compared with model estimates for clustering of T cells at 6 hours after transfer (e.g., Table 3 for $\mu = 0.1/\text{h}$).

The major caveat of this analysis is the assumption that the parameters determining T cell clustering are constant over time. Our data indicate that formation of clusters may be slowing down over time (Figure 6B). Therefore, we fitted the DD recruitment model to the longitudinal/path data assuming that parameters determining kinetics of cluster formation depend on the time since T cell transfer. Given how the data were collected (Figure S2) for our analysis we made the simplest assumption that the rates are constant in two time intervals: $(0 - 4)$ h and $(4 - 12)$ h but may be different between the time intervals. Assuming that in the DD recruitment model recruitment rates

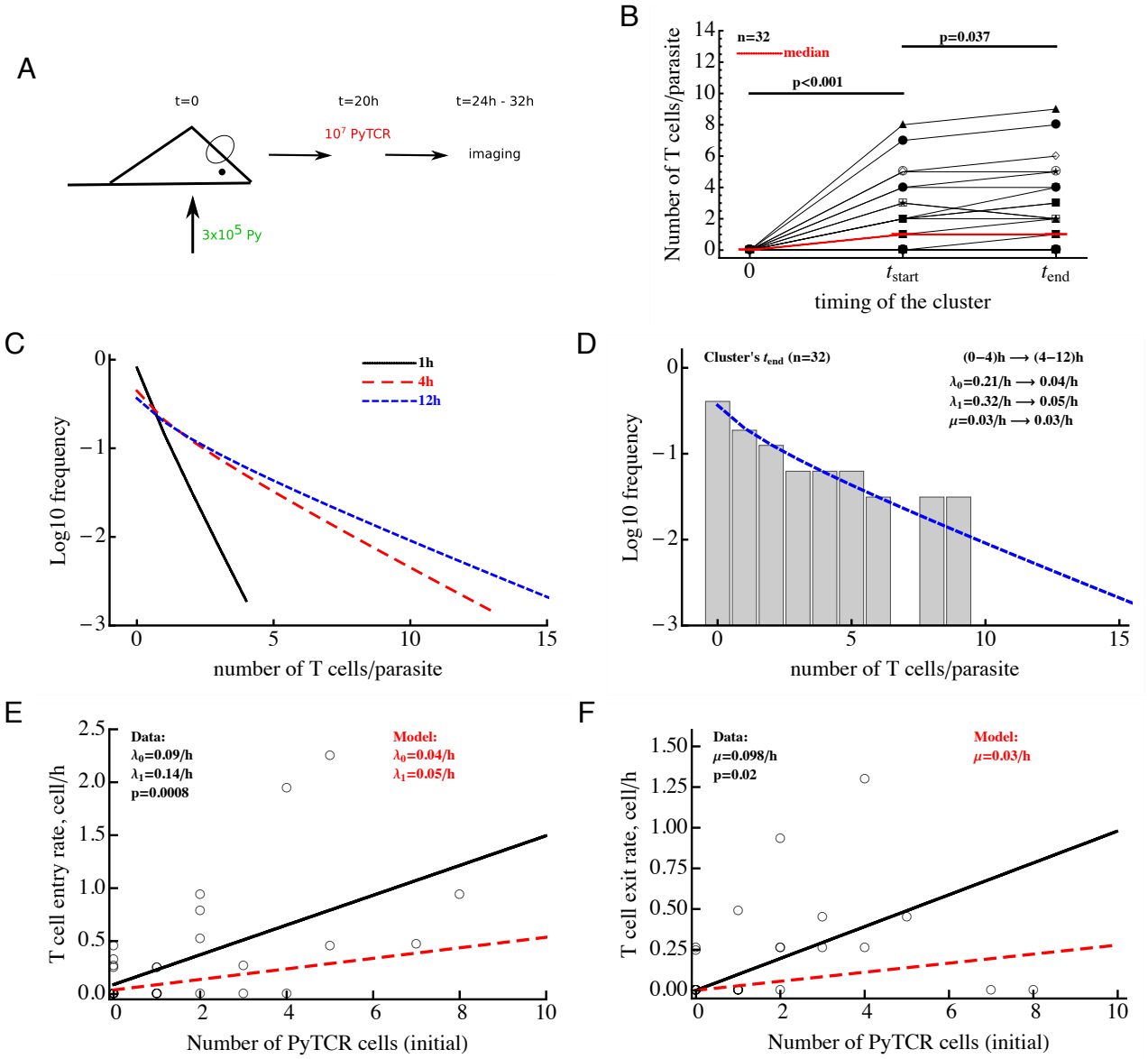

Figure 6: Clusters of T cells around the parasite are largely formed by 4 hours post T cell transfer. Panel A: mice were infected with 3×10^5 GFP-expressing Py sporozoites. Twenty hours later 10^7 Py-specific activated CD8 T cells (PyTCR) were transferred into infected mice and livers of these mice were imaged using intravital microscopy between 4 to 12 hours later. In total 32 parasites were randomly chosen and number of T cells in 40 μm radius of the same parasites were counted at both times [29]. Panel B: significant increase in the median size of the cluster around Py-infected hepatocytes is observed in the first time period and there was a moderate increase in the median cluster size in the following 4-8 hours (Wilcoxon sum rank test). Thick red line shows change in the median number of T cells per parasite. In these experiments, 44% and 38% of all parasites did not have a single CD8 T cell nearby for first and last measurement of T cell clusters, respectively. Panel C: we plot the distribution of cluster sizes as predicted by the best fit model at different times after T cell cluster 4. The best fit model was a model assuming DD recruitment (eqns. (1)–(2)) with entry rates into the cluster being dependent on the time period (0-4h and 4-12h) but with the same exit rate during 12 hour period. Estimated parameters and their 95% confidence intervals for 0-4 h time interval are $\lambda_0 = 0.21$ $(0.11 - 0.34)$/h and $\lambda_1 = 0.32$ $(0.11 - 0.49)$/h; for 4-12h time interval are $\lambda_0 = 0.04$ $(0.0 - 0.10)$/h and $\lambda_1 = 0.05$ $(0.02 - 0.08)$/h with the exit rate $\mu = 0.030$ $(0.0 - 0.086)$/h. Panel D: we show the observed distribution of cluster sizes at the last measurement for each parasite and predictions of the DD recruitment model for 12 hours after T cell transfer. Panels E&F: correlation between the T cell entry rate into the cluster (panel E) or exit rate from the cluster (panel F) as the function of the initial number of PyTCR T cells in the cluster. Points are experimentally measured values from Cockburn *et al.* [29], solid lines show the regression lines with estimated intercept $\lambda_0 = 0.09$/h and slope $\lambda_1 = 0.14$/h (panel E) or slope $\mu = 0.098$/h (panel F); both slopes are significantly different from zero (t-test). Dashed lines in panels E&F show prediction of the mathematical model for the recruitment and exit rates estimated by fitting DD recruitment model to the clustering data.

λ_0 and λ_1 are time-dependent and the exit rate μ is time-independent, the model fitted the data significantly better than the DD recruitment model with constant parameters (likelihood ratio test, $\chi^2_2 = 30.0$, $p < 10^{-6}$). Parameter estimates suggest a six fold reduction in both λ_0 and λ_1 4 hours after T cell transfer (see legend of Figure 6 for actual parameter estimates). A similar decline in both λ_0 and λ_1 at 4 hours after T cell transfer was confirmed by fitting the model in which both rates declined by the same amount α; such a model fitted the data with a similar quality as the model that allowed for different decline in the two rates with time since T cell transfer (likelihood ratio test, $\chi^2_1 = 0.02$, $p = 0.89$). Because the distribution of cluster sizes was measured experimentally at different time points it was not possible to visualize the model fits of the data. However, because model predictions suggested little change in cluster size distributions between 4 and 12 hours after T cell transfer (Figure 6C), the model predicted well the distribution of cluster sizes for each of the parasite at the end of imaging (Figure 6D, $\chi^2_8 = 1.14$, $p = 1$). Interestingly, this analysis indicated an extremely slow rate of T cell exit from clusters at 4-12 hours after T cell transfer suggesting that nearly every cell that enters the cluster after 4 hours post T cell transfer remains in the cluster which is an indirect support for the "retention" model.

An alternative DD recruitment model is one in which recruitment rates into the cluster remain constant over time but exit rates from the cluster change with time. This model did slightly improve the model fit of the data as compared to the model with constant parameters (likelihood change of 3.32, $\chi^2_2 = 6.6$, $p = 0.01$) and predicted that constant recruitment rates are $\lambda_0 = 0.14$/h and $\lambda_1 = 0.19$/h, and exit rate for the first 4 hours is $\mu = 0$/h and for after 4 hours is $\mu = 0.19$/h. This model suggests an alternative interpretation of the cluster formation dynamics — namely that T cells are recruited into the cluster and retained during the first four hours after T cell transfer — but after the initial time additional recruited T cells have a high chance of leaving. Because the quality of this model fit of the data was significantly worse than that of the model with time-dependent recruitment rates (ΔAIC = 21 or Akaike weight $w < 0.001$ for time-dependent exit rate model), our data appear to be more consistent with the time-dependent recruitment and constant exit. This suggests that the best explanation of the longitudinal clustering data is that the formation of clusters is driven by the DD recruitment model in which the rate of T cell recruitment into the cluster declines over time. Parameter estimates also suggest that the formation of clusters around Py-infected hepatocytes occurs mainly during the first 4 hours after T cell transfer.

The dynamics of change in the number of T cells near the parasite between 4 and 12 hours were followed by time-lapse intravital microscopy which allowed to calculate the number of T cells entering the cluster and leaving the cluster in this time period [29, Figure 6E-F]. Analysis showed that both entry and exit rates were strongly dependent on the cluster size k even though there was large variability in the number of T cells entering and exiting individual clusters. Interestingly, the slopes of the dependence of recruitment and exit rates was 2-3 fold higher for experimentally measured rates as compared to the parameters found by fitting DD recruitment model to longitudinal data (Figure 6E-F). One potential explanation of this difference is that perhaps not all cells that come near the parasite (i.e., within 40 μm distance) recognize the infection and leave, thus, increasing the overall observed T cell surveillance rate. In contrast, the model only accounts for T cells which actually formed clusters and thus most likely have recognized the parasite.

4 Discussion

Studies from two independent groups showed that activated Plasmodium-specific CD8 T cells form clusters around Plasmodium-infected hepatocytes and that such clusters are correlated with elimination of the Plasmodium liver stages [29–31]. Application of mathematical models to data on distribution of the number of Py-specific CD8 T cells around randomly chosen parasites suggested that formation of the clusters is not a random process; the model in which activated T cells of different specificities are attracted at a rate proportional to the number of Py-specific T cells already present near the parasite, described the data with best quality [29]. More recent work also suggested that formation of CD8 T cell clusters around Plasmodium-infected hepatocytes depends on $CD11c^+$ cells and that activated CD8 T cells, specific to irrelevant antigens, do not appear to play a significant role in protection against Plasmodium challenge [31]. We analyzed experimental data with the use of mathematical modeling to provide further insights into potential mechanisms of the formation of clusters around Py-infected hepatocytes.

First, we found that several independent experimental datasets are fully consistent with the model in which variability in the number of activated Py-specific CD8 T cells located near the parasite-infected hepatocytes is driven by variability in the "environment" around the infected hepatocytes, providing indirect support for the T cell-extrinsic mechanism of cluster formation (e.g., Figure 3). These results suggested that data on clustering of Py-specific T cells alone may be insufficient to discriminate between T cell-intrinsic and T cell-extrinsic mechanisms of cluster formation [49]. A key experiment, rejecting the "variability of the environment" hypothesis as the sufficient mechanism explaining distribution of cluster sizes is one involving either transfer of only OT1 T cells (which are not specific to Py antigens) or OT1 T cells together with PyTCR T cells – only in the latter case, OT1 T cells form co-clusters with Py-specific T cells (Figure 4 and see [29, 31]). The mathematical model assuming fixed yet variable (between individual parasites) environment was not able to accurately explain such data (Figure 4). The result, however, does not mean that inflammation is irrelevant for parasite's replication in the liver. In fact, recent work suggested that sporozoite infection of the liver does lead to inflammation [36].

Second, while OT1 T cells of irrelevant specificity are found in clusters together with Py-specific CD8 T cells, we found no evidence that OT1 improve cluster formation (Figure 5). If anything, OT1 T cells may in fact reduce the rate of recruitment of other T cells into the cluster as indicated by the negative values for the recruitment rate λ_1 (Table 2); however, this value was not significantly different from zero. Mathematical modeling also suggested that there may be a slight bias in the clusters to have more Py-specific T cells than T cells of irrelevant specificity per cluster but the biological relevance of such a small bias remains unclear. The limited role of T cells of irrelevant specificity in the formation of T cell clusters in Py-infected mice is consistent with the observation that transfers of large numbers of activated CD8 T cells with irrelevant specificity into Plasmodium-infected mice did not impact efficiency at which parasites were killed by Plasmodium-specific T cells [31]. Interestingly, and perhaps surprisingly, this result contradicts a recent observation of suppression of development of T-cell-driven type 1 diabetes by islet-non specific CD8 T cells [50].

Third, by following longitudinal changes in the number of CD8 T cells around individual parasites over time we found that T cell clusters are formed rather rapidly, at least within the first 4 hours after T cell transfer (Figure 6B), and mathematical modeling predicted recruitment of T cells to the parasite and retaining of the T cells in the cluster in that time period. Interestingly, the rates of entry into and exit from the clusters declined after the 4 hours six fold further supporting the conclusion

that clusters are formed rapidly and few cells enter and exit the cluster after 4 hours since T cell transfer (Figure 6C-D). Stochastic simulations of the formation of clusters assuming DD recruitment model with different entry/exit rates also suggested that between 4 and 8 hours post T cell transfer, entry and exit rates cannot be large (Figure S3). This is because when these rates are large, changes in the cluster size in the 4-8 hour time period are highly variable with some clusters growing in size exponentially while other clusters nearly disappearing (e.g., Figure S3C&F). This, however, was not observed in experimental data (shown by dotted histogram in Figure S3D-F and see Figure 6B and Figure S2). Rapid recruitment of CD8 T cells to the liver stages in the first 4 hours after T cell transfer may be the result of the specific experimental set-up as it is expected that immediately after intravenous injection, large numbers of T cells would be passing through the liver increasing chances of T cells finding the infection site [51], and that the number of liver-resident CD8 T cells tends to reach a steady state at 2-3 hours after T cell transfer [52, James O'Connor and Ian Cockburn (ms. in preparation)].

Interestingly, there was some discrepancy between the estimated rates of T cell entry into the cluster and exit from the cluster measured experimentally and predicted by the model (Figure 6E-F) most likely indicating that not all T cells that were observed to come in close proximity with the parasite recognize it. We also found a strong correlation between experimentally measured rates of T cell entry into and exit from the clusters (Figure S4) which may indicate that in addition to T cell-intrinsic mechanisms of clustering, some parasites may be more "attractive" to T cells. Indeed, none of our tested models could well explain the formation of extremely large T cell clusters around Py-infected hepatocytes (e.g., with 15 or more T cells, see Figure 2B) which could indicate the need for future models to include both DD recruitment and variability in parasite's attractiveness.

In this paper we analyzed a number of different datasets that involve different cell types, different experimental set-ups, and different mice. We found it encouraging that some of these datasets were in a way "consistent". Specifically, we observed similar clustering of CD8 T cells in naive mice (Figure 3A), PT-treated PyTCR T cells (Figure 2B), or activated OT1 T cells of irrelevant specificity (Figure 4A) and the random entry/exit model described these data with nearly identical parameters (likelihood ratio test, $\chi^2_2 = 5.38$, $p = 0.07$). The DD model could describe the distribution of cluster sizes of PyTCR T cells in three different experiments (Figure 2A and data in [29]) with identical parameters ($\chi^2_4 = 3.31$, $p = 0.51$). However, the clustering of CD8 T cells following immunization with radiation-attenuated sporozoites (Figure 3B) did not match well the clustering of the mixture of PyTCR and OT1 T cells ($\chi^2_2 = 12.21$, $p = 0.002$) perhaps highlighting potential differences between active and passive immunizations (the latter involving transfer of pre-activated CD8 T cells).

In multiple analyses we found that a DI exit (retention) model did not describe well the clustering data. However, a poorer fit of the model (as compared to other tested models) does not necessarily falsify a model [49], and additional experiments will be needed to formally rule out this model. Fitting the DI exit model to the "longitudinal" data on change in cluster size around individual parasites (e.g., Figure S2) revealed that this model could not accurately describe the data assuming a constant entry and time-dependent exit rates based on likelihood of the model ($\mathcal{L} = 100.2$ vs. $\mathcal{L} = 83.0$ of the DD recruitment model with time-dependent recruitment and constant exit rates, results not shown). In addition, if the rate of T cell exit from the clusters found in the DD recruitment model is constant over the course of the first 12 hours since T cell transfer, it would suggest that T cells mostly enter the clusters and rarely exit them (given $\mu = 0.028$/h corresponding to the residence time of T cells in the cluster of about $1/\mu \approx 36$ h), providing some indirect support for the retention model.

Conversely, our result that the DD recruitment model describes most of the data with best quality

does not prove that this model is the true mechanism of the formation of large clusters around Py-infected hepatocytes. Future experiments will have to test the major prediction of the model — that clusters of a large size attract more T cells per unit of time. Such experiments may involve measurement of T cell movements in the liver using intravital microscopy and estimating bias in T cell shift towards the parasite. Indeed, our recent work suggested that there is a bias in PyTCR cell movements towards Py-infected cells [34] but more analyses are needed to evaluate whether such a bias depends on the number of T cells already present at the parasite and whether a small bias is sufficient to explain the formation of larger clusters (with $k \sim 5 - 10$ of CD8 T cells per parasite) within few hours after T cell transfer. Detecting a bias in T cell movement towards the infection site may be complicated as our current analysis predicts that "attraction" seems to be present only during the first 4 hours after T cell transfer (Figure 6C-D).

Our analysis has several potential limitations. The biggest issue is that by using numerical solutions of the DD recruitment model we showed that the distribution of cluster sizes at a single time point does not allow to accurately estimate the rates of T cell entry into the clusters and T cell exit from the clusters, and thus, most of our analysis were restricted to estimating relative recruitment rates. Ongoing analysis based on analytical solutions in Bailey [48] has also demonstrated this point using analytical derivations (ms. in preparation). While the estimated values of the recruitment rates λ_0 and λ_1 in the DD recruitment model directly depend on the assumed exit rate μ (see Table 3) we showed that the likelihood of the model fit to data assuming a steady state or dynamics for clusters at a given exit rate μ were nearly identical strongly suggesting that our results on best fit models obtained assuming steady states are robust. However, the actual values of the entry and exit rates cannot be found with certainty as these depend on the actual value of the assumed exit rate (Table 3).

Another complexity in the analysis comes from our finding that rates of T cell entry into the cluster are time-dependent (Figure 6). To investigate whether this impacts our selection of best fit models assuming steady state solutions we did the following. We fitted the DD recruitment model to the clustering data at one time point by assuming that early recruitment rates λ_0 and λ_1 are unknown and that late entry rates are fixed to values found from the analysis of longitudinal data (Figure 6) and that the exit rate μ is constant. Under these minimal assumptions the model fit was of nearly identical quality as the model fit of the data assuming a steady state (results not shown). Therefore, even for time-dependent parameters our results determining which models are not consistent with clustering data remain valid.

An important experimental limitation of our data is the way of how experiments were performed whereby pre-activated T cells were transferred into mice that had been already infected with Plasmodium sporozoites (e.g., Figure 2A). This sequence of events does not fully match the physiological situation in which activated or memory CD8 T cells are already present at the time of sporozoite challenge. In fact, the rapid predicted decline in the rates of T cell recruitment into clusters with time suggests that it may be an artifact of the experimental system. Whether change in the experimental protocol will lead to support of the same mathematical models of cluster formation remains to be determined (and is the focus of our ongoing experiments and analyses).

Mathematical methodologies used in this work provided deeper understanding of how CD8 T cells form clusters around Plasmodium-infected hepatocytes. While formation of such clusters was a novel observation in malaria infection of the liver, clusters of immune cells have been observed in multiple systems including herpes simplex virus (HSV) [53] and Mycrobacterium tuberculosis (Mtb) [54]. In fact, formation of granulomas in Mtb-infected animals and humans is a classical example of

T cell clustering around the infection site. Interestingly, both Mtb-specific CD4 T cells and CD4 T cells of irrelevant specificities were found in granulomas of Mtb-infected monkeys [55] which could be explained by the DD recruitment model extended in this work. Movement of neutrophils towards an injury site may also depend on the number of neutrophils that have already reached the site [56]. It may be useful to combine mathematical modeling tools for deeper understanding of the mechanisms of formation of clusters of immune cells in these and other systems.

While our work provides some clarification regarding mechanisms of CD8 T cell cluster formation around Plasmodium-infected hepatocytes, many questions remain. In particular, while clusters appear to be important for the death of the parasite [29, 30], whether clusters of a larger size kill the parasites faster remains unknown. Classical work involving killing of chromium-labeled target cells by cytotoxic T lymphocytes (CTLs) suggested a faster killing of targets bound by multiple CTLs [57], and *in vivo*, death of peptide-pulsed targets is directly proportional to the number of peptide-specific CTLs [33]. Recent work also suggested that the probability of death of virus-infected cells in skin *in vivo* was higher when the infected cell was contacted by several antigen-specific CD8 T cells [32]. Whether the same relationship holds for T cells killing Plasmodium parasites in the liver remains to be determined. However, several studies have shown that probability of clearance of i.v. injected Plasmodium sporozoites does depend on the number of vaccine-induced CD8 T cells in the liver [22, 23, 58]. However, these previous studies are numerically inconsistent suggesting that either 3×10^4 [22] or 10^6 [23] memory CD8 T cells in the liver are needed for sterilizing protection. Further work is required to accurately quantify the number of T cells needed for protection.

Our results suggest that activated CD8 T cells of irrelevant specificities do not play a major role in cluster formation, and elegant experiments demonstrated that large numbers of non-specific T cells do not impair the ability of Plasmodium-specific T cells to eliminate the parasites [31]. However, the latter result was found by using only two different ratios of Plasmodium-specific and non-specific T cells and 3 mice per group, so it remains to be determined if competition between such cells for the access to infected cells occurs at higher ratios, e.g., as has been observed in another system [50]. In natural settings we do expect that Plasmodium-specific T cells will be likely outnumbered by memory T cells specific to other infections, and therefore, deeper understanding of such competition may be of relevance to malaria vaccines, inducing liver-resident memory CD8 T cells for protection [59].

Accumulation of large numbers of CD8 T cells around Plasmodium-infected cells raises an intriguing possibility that parasites may in fact attempt to attract T cells. While this may be detrimental to an individual parasite, as a population this may give an advantage if attracting many T cells to one site prevents T cells from effectively locating parasites at other sites. To cause blood-stage infection, there is a need for only one liver stage to mature and release differentiated parasites (merozoites) into the blood stream. Indeed, it should be noted that in most of our experiments, many of surveyed parasites did not have a T cell nearby at 6-8 hours after T cell transfer. Future studies may be needed to investigate whether such a strategy is indeed evolutionarily advantageous.

It remains unclear how relevant our results are for T cell-mediated protection of humans against malaria. Because of the need of imaging i.v. injected sporozoites in the liver, large numbers of parasites must be used. This is in contrast with very few sporozoites that humans are likely to be exposed to when bit by infectious mosquitoes.

Taken together, here we illustrated the power of combining the use of detailed quantitative experimental data with mathematical modeling, and limitations that come from inability to make solid conclusions from extensive yet limited experimental data. The field of immunology will likely benefit from closer collaborations between experimentalists and modellers where experimentalists

being involved in data analyses and modeling, and modellers are cooperating with experimentalists in designing experiments to test and potentially falsify alternative mathematical models.

5 Acknowledgements

This work was supported by the NIH grant (R01 GM118553) to VVG. This manuscript has been released as a preprint at BioArxiv [60].

6 Author contributions

RKK, IAC, and VVG designed the study. RKK and VVG run most of analyses and wrote first drafts of the paper. HR confirmed some of the numerical results and provided derivation of the bias in co-clustering of antigen-specific and non-specific T cells. IAC performed additional data analyses and provided feedback on data interpretation and modeling. The final version of the paper was written by VVG but all authors contributed to the writing of the final version of the paper via comments and suggestions.

References

1. Murray, C. J. L., Ortblad, K. F., Guinovart, C., Lim, S. S., Wolock, T. M., Roberts, D. A., Dansereau, E. A., Graetz, N., Barber, R. M. *et al.* 2014 Global, regional, and national incidence and mortality for HIV, tuberculosis, and malaria during 1990-2013: a systematic analysis for the Global Burden of Disease Study 2013. *Lancet*, **384**(9947), 1005–1070. doi:10.1016/S0140-6736(14)60844-8.

2. WHO 2017 World malaria report. Tech. rep., WHO, Geneva: World Health Organisation. Http://www.who.int/tb/publications/factsheets/en/.

3. Cowman, A. F., Healer, J., Marapana, D. & Marsh, K. 2016 Malaria: Biology and disease. *Cell*, **167**, 610–624. doi:10.1016/j.cell.2016.07.055.

4. Zuzarte-Luis, V., Mota, M. M. & Vigario, A. M. 2014 Malaria infections: what and how can mice teach us. *J Immunol Methods*, **410**, 113–122. doi:10.1016/j.jim.2014.05.001.

5. Beier, J. C., Davis, J. R., Vaughan, J. A., Noden, B. H. & Beier, M. S. 1991 Quantitation of Plasmodium falciparum sporozoites transmitted in vitro by experimentally infected Anopheles gambiae and Anopheles stephensi. *Am J Trop Med Hyg*, **44**(5), 564–570.

6. Medica, D. L. & Sinnis, P. 2005 Quantitative dynamics of Plasmodium yoelii sporozoite transmission by infected anopheline mosquitoes. *Infect Immun*, **73**(7), 4363–4369. doi: 10.1128/IAI.73.7.4363-4369.2005.

7. Amino, R., Thiberge, S., Martin, B., Celli, S., Shorte, S., Frischknecht, F. & Ménard, R. 2006 Quantitative imaging of Plasmodium transmission from mosquito to mammal. *Nat Med*, **12**(2), 220–224. doi:10.1038/nm1350.

8. yue Xu, W., xiang Wang, X., Qi, J., hua Duan, J. & sheng Huang, F. 2010 Plasmodium yoelii: influence of immune modulators on the development of the liver stage. *Exp Parasitol*, **126**(2), 254–258. doi:10.1016/j.exppara.2010.05.005.

9. Tavares, J., Costa, D. M., Teixeira, A. R., Cordeiro-da Silva, A. & Amino, R. 2017 In vivo imaging of pathogen homing to the host tissues. *Methods*, **127**, 37–44. doi:10.1016/j.ymeth.2017.05.008.

10. Murphy, J. R., Baqar, S., Davis, J. R., Herrington, D. A. & Clyde, D. F. 1989 Evidence for a 6.5-day minimum exoerythrocytic cycle for Plasmodium falciparum in humans and confirmation that immunization with a synthetic peptide representative of a region of the circumsporozoite protein retards infection. *J Clin Microbiol*, **27**(7), 1434–1437.

11. Hermsen, C. C., Telgt, D. S., Linders, E. H., van de Locht, L. A., Eling, W. M., Mensink, E. J. & Sauerwein, R. W. 2001 Detection of Plasmodium falciparum malaria parasites in vivo by real-time quantitative PCR. *Mol Biochem Parasitol*, **118**(2), 247–251.

12. Sturm, A., Amino, R., van de Sand, C., Regen, T., Retzlaff, S., Rennenberg, A., Krueger, A., Pollok, J.-M., Menard, R. *et al.* 2006 Manipulation of host hepatocytes by the malaria parasite for delivery into liver sinusoids. *Science*, **313**(5791), 1287–1290. doi:10.1126/science.1129720.

13. Miller, L. H., Ackerman, H. C., zhuan Su, X. & Wellems, T. E. 2013 Malaria biology and disease pathogenesis: insights for new treatments. *Nat Med*, **19**(2), 156–167. doi:10.1038/nm.3073.

14. Weiss, W. R., Sedegah, M., Beaudoin, R. L., Miller, L. H. & Good, M. F. 1988 CD8+ T cells (cytotoxic/suppressors) are required for protection in mice immunized with malaria sporozoites. *Proc Natl Acad Sci U S A*, **85**(2), 573–576.

15. Rodrigues, M., Nussenzweig, R. S. & Zavala, F. 1993 The relative contribution of antibodies, CD4+ and CD8+ T cells to sporozoite-induced protection against malaria. *Immunology*, **80**(1), 1–5.

16. Romero, P., Maryanski, J. L., Corradin, G., Nussenzweig, R. S., Nussenzweig, V. & Zavala, F. 1989 Cloned cytotoxic T cells recognize an epitope in the circumsporozoite protein and protect against malaria. *Nature*, **341**(6240), 323–326. doi:10.1038/341323a0.

17. Rodrigues, M. M., Cordey, A. S., Arreaza, G., Corradin, G., Romero, P., Maryanski, J. L., Nussenzweig, R. S. & Zavala, F. 1991 CD8+ cytolytic T cell clones derived against the Plasmodium yoelii circumsporozoite protein protect against malaria. *Int Immunol*, **3**(6), 579–585.

18. Schmidt, N. W., Butler, N. S., Badovinac, V. P. & Harty, J. T. 2010 Extreme CD8 T cell requirements for anti-malarial liver-stage immunity following immunization with radiation attenuated sporozoites. *PLoS Pathog*, **6**(7), e1000 998. doi:10.1371/journal.ppat.1000998.

19. Butler, N. S., Schmidt, N. W., Vaughan, A. M., Aly, A. S., Kappe, S. H. I. & Harty, J. T. 2011 Superior antimalarial immunity after vaccination with late liver stage-arresting genetically attenuated parasites. *Cell Host Microbe*, **9**(6), 451–462. doi:10.1016/j.chom.2011.05.008.

20. Krzych, U., Dalai, S., Zarling, S. & Pichugin, A. 2012 Memory CD8 T cells specific for plasmodia liver-stage antigens maintain protracted protection against malaria. *Front Immunol*, **3**, 370. doi:10.3389/fimmu.2012.00370.

21. Krzych, U., Zarling, S. & Pichugin, A. 2014 Memory T cells maintain protracted protection against malaria. *Immunol Lett*, **161**(2), 189–195. doi:10.1016/j.imlet.2014.03.011.

22. Olsen, T. M., Stone, B. C., Chuenchob, V. & Murphy, S. C. 2018 Prime-and-trap malaria vaccination to generate protective CD8+ liver-resident memory T cells. *J Immunol*, **201**, 1984–1993. doi:10.4049/jimmunol.1800740.

23. Gola, A., Silman, D., Walters, A. A., Sridhar, S., Uderhardt, S., Salman, A. M., Halbroth, B. R., Bellamy, D., Bowyer, G. *et al.* 2018 Prime and target immunization protects against liver-stage malaria in mice. *Science translational medicine*, **10**. doi:10.1126/scitranslmed.aap9128.

24. Schofield, L., Villaquiran, J., Ferreira, A., Schellekens, H., Nussenzweig, R. & Nussenzweig, V. 1987 Gamma interferon, CD8+ T cells and antibodies required for immunity to malaria sporozoites. *Nature*, **330**(6149), 664–666. doi:10.1038/330664a0.

25. Vanderberg, J., Mueller, A.-K., Heiss, K., Goetz, K., Matuschewski, K., Deckert, M. & Schlüter, D. 2007 Assessment of antibody protection against malaria sporozoites must be done by mosquito injection of sporozoites. *Am J Pathol*, **171**, 1405–6; author reply 1406. doi: 10.2353/ajpath.2007.070661.

26. Klaunig, J. E., Goldblatt, P. J., Hinton, D. E., Lipsky, M. M., Chacko, J. & Trump, B. F. 1981 Mouse liver cell culture. I. Hepatocyte isolation. *In Vitro*, **17**(10), 913–925.

27. Jones, L. D., Nielsen, M. K. & Britton, R. A. 1992 Genetic variation in liver mass, body mass, and liver:body mass in mice. *J Anim Sci*, **70**(10), 2999–3006.

28. Sohlenius-Sternbeck, A.-K. 2006 Determination of the hepatocellularity number for human, dog, rabbit, rat and mouse livers from protein concentration measurements. *Toxicol in Vitro*, **20**(8), 1582–1586. doi:10.1016/j.tiv.2006.06.003.

29. Cockburn, I. A., Amino, R., Kelemen, R. K., Kuo, S. C., Tse, S.-W., Radtke, A., MacDaniel, L., Ganusov, V. V., Zavala, F. *et al.* 2013 In vivo imaging of CD8+ T cell-mediated elimination of malaria liver stages. *Proc Natl Acad Sci U S A*, **110**(22), 9090–9095. doi: 10.1073/pnas.1303858110.

30. Kimura, K., Kimura, D., Matsushima, Y., Miyakoda, M., Honma, K., Yuda, M. & Yui, K. 2013 CD8+ T cells specific for a malaria cytoplasmic antigen form clusters around infected hepatocytes and are protective at the liver stage of infection. *Infection and immunity*, **81**, 3825–3834. doi: 10.1128/IAI.00570-13.

31. Akbari, M., Kimura, K., Bayarsaikhan, G., Kimura, D., Miyakoda, M., Juriasingani, S., Yuda, M., Amino, R. & Yui, K. 2018 Non-specific CD8+ T cells and dendritic cells/macrophages participate in CD8+ T cell-mediated cluster formation against malaria liver-stage infection. *Infec Immun.* doi:10.1128/IAI.00717-17.

32. Halle, S., Keyser, K. A., Stahl, F. R., Busche, A., Marquardt, A., Zheng, X., Galla, M., Heissmeyer, V., Heller, K. *et al.* 2016 In vivo killing capacity of cytotoxic T cells is limited and involves dynamic interactions and T cell cooperativity. *Immunity*, **44**, 233–245. doi: 10.1016/j.immuni.2016.01.010.

33. Ganusov, V. V., Barber, D. L. & De Boer, R. J. 2011 Killing of targets by CD8 T cells in the mouse spleen follows the law of mass action. *PLoS One*, **6**(1), e15 959.

34. Kelemen, R., He, G., Woo, H., Lane, T., Rempe, C., Wang, J., Cockburn, I. A., Ganusov, V. V. & Berry, M. W. 2014 Classification of T cell movement tracks allows for prediction of cell function. *Int. J. Comput. Biol Drug Desig*, **7**, 113–129.

35. Elias, H. & Sherrick, J. C. 1969 *Morphology of the liver*. London (& New York): Academic Press.

36. Inácio, P., Zuzarte-Luís, V., Ruivo, M. T. G., Falkard, B., Nagaraj, N., Rooijers, K., Mann, M., Mair, G., Fidock, D. A. *et al.* 2015 Parasite-induced ER stress response in hepatocytes facilitates Plasmodium liver stage infection. *EMBO reports*, **16**, 955–964. doi:10.15252/embr.201439979.

37. McNamara, H. A., Cai, Y., Wagle, M. V., Sontani, Y., Roots, C. M., Miosge, L. A., O'Connor, J. H., Sutton, H. J., Ganusov, V. V. *et al.* 2017 Up-regulation of LFA-1 allows liver-resident memory T cells to patrol and remain in the hepatic sinusoids. *Science Immunology*, pp. 1–10.

38. Severgnini, M., Sherman, J., Sehgal, A., Jayaprakash, N. K., Aubin, J., Wang, G., Zhang, L., Peng, C. G., Yucius, K. *et al.* 2012 A rapid two-step method for isolation of functional primary mouse hepatocytes: cell characterization and asialoglycoprotein receptor based assay development. *Cytotechnology*, **64**, 187–195. doi:10.1007/s10616-011-9407-0.

39. Morales-Navarrete, H., Segovia-Miranda, F., Klukowski, P., Meyer, K., Nonaka, H., Marsico, G., Chernykh, M., Kalaidzidis, A., Zerial, M. *et al.* 2015 A versatile pipeline for the multi-scale digital reconstruction and quantitative analysis of 3d tissue architecture. *eLife*, **4**. doi: 10.7554/eLife.11214.

40. Vanderberg, J. P., Nussenzweig, R. S. & Most, H. 1968 Further studies on the Plasmodium berghei-Anopheles stephensi–rodent system of mammalian malaria. *J Parasitol*, **54**, 1009–1016.

41. Vanderberg, J. P. 1977 Plasmodium berghei: quantitation of sporozoites injected by mosquitoes feeding on a rodent host. *Experimental parasitology*, **42**, 169–181.

42. Kampen, N. G. v. 1981 *Stochastic processes in physics and chemistry*. North-Holland.

43. Gillespie, D. 1977 Exact stochastic simulation of coupled chemical-reactions. *J Phys Chem*, **81**(25), 2340–2361.

44. Burnham, K. P. & Anderson, D. R. 2002 *Model selection and multimodel inference: a practical information-theoretic approach*. Springer-Verlag, New York.

45. Spangrude, G. J., Sacchi, F., Hill, H. R., Van Epps, D. E. & Daynes, R. A. 1985 Inhibition of lymphocyte and neutrophil chemotaxis by pertussis toxin. *J Immunol*, **135**, 4135–4143.

46. Bromley, S. K., Mempel, T. R. & Luster, A. D. 2008 Orchestrating the orchestrators: chemokines in control of T cell traffic. *Nat Immunol*, **9**(9), 970–980. doi:10.1038/ni.f.213.

47. Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. 1996 *Numerical recipes in C*, vol. 2. Cambridge university press Cambridge.

48. Bailey, N. T. J. 1964 *The elements of stochastic processes with applications to the natural sciences*. New York: Wiley.

49. Ganusov, V. V. 2016 Strong Inference in Mathematical Modeling: A Method for Robust Science in the Twenty-First Century. *Front Microbiol*, **7**, 1131. doi:10.3389/fmicb.2016.01131.

50. Christoffersson, G., Chodaczek, G., Ratliff, S. S., Coppieters, K. & von Herrath, M. G. 2018 Suppression of diabetes by accumulation of non-islet-specific CD8 effector T cells in pancreatic islets. *Science immunology*, **3**. doi:10.1126/sciimmunol.aam6533.

51. Ganusov, V. V. & Auerbach, J. 2014 Mathematical modeling reveals kinetics of lymphocyte recirculation in the whole organism. *PLoS Comp Biol*, **10**, e1003 586.

52. Guidotti, L. G., Inverso, D., Sironi, L., Di Lucia, P., Fioravanti, J., Ganzer, L., Fiocchi, A., Vacca, M., Aiolfi, R. *et al.* 2015 Immunosurveillance of the liver by intravascular effector CD8(+) T cells. *Cell*, **161**, 486–500. doi:10.1016/j.cell.2015.03.005.

53. Tang, V. A. & Rosenthal, K. L. 2010 Intravaginal infection with herpes simplex virus type-2 (HSV-2) generates a functional effector memory T cell population that persists in the murine genital tract. *Journal of reproductive immunology*, **87**, 39–44. doi:10.1016/j.jri.2010.06.155.

54. Ramakrishnan, L. 2012 Revisiting the role of the granuloma in tuberculosis. *Nat Rev Immunol*, **12**(5), 352–366. doi:10.1038/nri3211.

55. Gideon, H. P., Phuah, J., Myers, A. J., Bryson, B. D., Rodgers, M. A., Coleman, M. T., Maiello, P., Rutledge, T., Marino, S. *et al.* 2015 Variability in tuberculosis granuloma t cell responses exists, but a balance of pro- and anti-inflammatory cytokines is associated with sterilization. *PLoS pathogens*, **11**, e1004 603. doi:10.1371/journal.ppat.1004603.

56. Lammermann, T., Afonso, P. V., Angermann, B. R., Wang, J. M., Kastenmüller, W., Parent, C. A. & Germain, R. N. 2013 Neutrophil swarms require LTB4 and integrins at sites of cell death in vivo. *Nature*, **498**, 371–375. doi:10.1038/nature12175.

57. Perelson, A., Macken, C., Grimm, E., Roos, L. & Bonavida, B. 1984 Mechanism of cell-mediated cytotoxicity at the single cell level. VIII. Kinetics of lysis of target cells bound by more than one cytotoxic T lymphocyte. *J Immunol*, **132**(5), 2190–8.

58. Spencer, A. J., Longley, R. J., Gola, A., Ulaszewska, M., Lambe, T. & Hill, A. V. S. 2017 The threshold of protection from liver-stage malaria relies on a fine balance between the number of infected hepatocytes and effector CD8+ T cells present in the liver. *J Immunol*, **198**, 2006–2016. doi:10.4049/jimmunol.1601209.

59. Cockburn, I. A. & Seder, R. A. 2018 Malaria prevention: from immunological concepts to effective vaccines and protective antibodies. *Nature immunology*, **19**, 1199–1211. doi:10.1038/s41590-018-0228-6.

60. Kelemen, R. K., Rajakaruna, H., Cockburn, I. A. & Ganusov, V. V. 2018 Clustering of activated CD8 T cells around malaria-infected hepatocytes is rapid and is driven by antigen-specific T cells. Bioaxiv: doi:10.1101/508796.

Clustering of activated CD8 T cells around malaria-infected hepatocytes is rapid and is driven by antigen-specific T cells

Reka K. Kelemen, Harshana Rajakaruna, Ian A. Cockburn, and Vitaly V. Ganusov

Supplemental Information

Bias in clustering of Plasmodium-specific CD8 T cells and T cells of irrelevant specificity

In fitting mathematical models to data on co-clustering of PyTCR (Py-specific) and OT1 (OVA-specific) CD8 T cells we found slight bias in the number of PyTCR cells in a given cluster. Here we provide mathematical justification of this observation.

Defining the problem. The linear system of ODEs given in eqns. (1)–(2) can be written as follows:

$$\frac{d\vec{P}_{i,j}}{dt} = A(\lambda_{i,j}, \mu_{i,j})\vec{P}_{i,j}(t), \tag{S.1}$$

where $A(\lambda_{i,j}, \mu_{i,j})$ is the probability transition matrix of time-independent entry and exit rate parameters $\lambda_{i,j}$ and $\mu_{i,j}$, respectively, and the subscript (i, j) denotes the number of PyTCR and OT1 cells of a given cluster, respectively, specific to the probabilities and rate parameters. Here, we consider the PyTCR density-dependent recruitment co-cluster model, where we assume that entry rates $\lambda_{i,j}$ of cell combinations (i, j) are time-constant, and both PyTCR and OT1 cells are attracted to a cluster depending on the density of PyTCR cells (type i) in the cluster, which is called the DD recruitment model. Thus, we can write $\lambda_{i,j} = \lambda_0 + i\lambda_1$; where λ_0 and λ_1 are parameters. The exit rate $\mu_{i,j}$ from a cluster are per-capita with respect to the particular cell density, i.e., $\mu_{i,j} = i\mu$ or $\mu_{i,j} = j\mu$, where $\mu > 0$ is a constant.

Defining $\theta_0 = \lambda_0/\mu$, $\theta_1 = \lambda_1/\mu$, we can write

$$\frac{d\vec{P}_{i,j}}{dt} = \mu A(\theta_0, \theta_1)\vec{P}_{i,j}(t). \tag{S.2}$$

Consider the solutions $P^n_{i,j}$ of the above system of equations written in a matrix of size $n \times n \in \Re$, with the largest possible co-cluster size $k_{max} = n$, i.e., $\max(i + j) = n$, where i (i.e., PyTCR cells in a cluster) denotes the rows, and j (i.e., OT1 cells in a cluster) denotes the columns of the matrix. The previous results of fitting DD recruitment model to the data showed that probabilities $P_{i,j}$ of entries (i, j) for smaller j of the upper right triangle of the matrix are likely to be larger than their respective mirror entries (j, i) of the lower left triangle of the matrix, and in contrast, $P_{i,j}$ of entries (i, j) for larger j of the upper right triangle of the matrix are likely to be larger than their respective mirror entries (j, i) of the lower left triangle of the respective matrix. That is, the probabilities for smaller clusters to have a larger number of OT1 cells are greater, whereas the probabilities for larger clusters to have a larger number of PyTCR cells are greater. Here, we show the formation of this systematic bias in the DD recruitment model for $P_{i,j}$ solutions for both the steady state and time-evolving conditions.

Explanation of bias for the system in steady state. Consider the system at steady state, i.e., where the rate of evolution of probabilities of all possible combinations of PyTCR and OT1 cells over time in the ODEs are zero. Thus, we get $A(\theta_0, \theta_1)P_{i,j}(t) = 0$ for $\mu > 0$. Note below that superscript n of $P^n_{i,j}$ indicates the maximum cluster size of the matrix.

I) Case $n = 1$ [i.e., $\max(i + j) = 1$].

$i =$ PyTCR, $j =$ OT1	$j = 0$	$j = 1$
$i = 0$	$P^1_{0,0}$	$P^1_{0,1}$
$i = 1$	$P^1_{1,0}$	

Table S1: Probability table for $n = 1$.

The steady state equations of the $n = 2$ matrix system are

$$P^1_{0,0} = (P^1_{1,0} + P^1_{0,1})/2\theta_0, \tag{S.3}$$

$$P^1_{1,0} = \frac{\theta_0}{[(2\theta_0 + 1) + 2\theta_1]} P^1_{0,0}, \tag{S.4}$$

$$P^1_{0,1} = \frac{\theta_0}{(2\theta_0 + 1)} P^1_{0,0}. \tag{S.5}$$

Since $\theta_0, \theta_1 > 0$, we get $P^1_{1,0} < P^1_{0,1}$. Here, $\sum P^1_{i,j} = 1$ for $i = 0, 1$ and $j = 0, 1$, s.t. $(i + j) \leq 1$. And, the ratio $2\theta_1/\theta_0$ determines how large the $P^1_{1,0} < P^1_{0,1}$ relationship is. Also, note that for any system of equations of size $\max(i + j) = n$ at steady state, the relationship in eq. (S.3) satisfies for probabilities $P^n_{0,0}, P^n_{0,1}, P^n_{1,0}$. That is,

$$P^n_{0,0} = (P^n_{1,0} + P^n_{0,1})/2\theta_0. \tag{S.6}$$

Corollary I: Due to similarity of coeffcients between the linear equations (S.3) and (S.6) of a given system, regardless of the matrix size n, s.t., $\max(i + j) = n$, with fixed rate parameters, $P^n_{0,1}$ and $P^n_{1,0}$ given by eq. ((S.6)) can be expressed as $\alpha_n P^1_{0,1}$ and $\alpha_n P^1_{1,0}$, respectively, for any given $\alpha_n P^1_{0,0}$ of the eq. (S.3), where $\alpha_n \in \Re$. Here, $\alpha_n = 1/\sum P^n_{i,j}\ \forall(i, j)$ entries.

Thus, the condition $P^n_{1,0} < P^n_{0,1}$ holds true for every solution of a given system of size n.

II) Case $n = 2$ [i.e., $\max(i + j) = 2$].

$i =$ PyTCR, $j =$ OT1	$j = 0$	$j = 1$	$j = 2$
$i = 0$	$P^2_{0,0}$	$P^2_{0,1}$	$P^2_{0,2}$
$i = 1$	$P^2_{1,0}$	$P^2_{1,1}$	
$i = 2$	$P^2_{2,0}$		

Table S2: Probability table for $n = 2$.

Here, we get,

$$P^2_{2,0} = \frac{(\theta_0+\theta_1)}{(2\theta_0+2)+2(2\theta_1)} P^2_{1,0} \tag{S.7}$$

$$P^2_{0,2} = \frac{\theta_0}{(2\theta_0+2)} P^2_{0,1} \tag{S.8}$$

Since we can rewrite $P^2_{1,0} = \alpha_2 P^1_{1,0}$, $P^2_{0,1} = \alpha_2 P^1_{0,1}$ and $P^2_{0,0} = \alpha_2 P^1_{0,0}$, s.t.. $\alpha_2 = 1/\sum P^2_{i,j} \forall (i,j)$ entries, s.t., $\max(i+j) = 2$, from Corollary I, we get

$$P^2_{2,0} = \alpha_2 \frac{(\theta_0+\theta_1)}{[(2\theta_0+2)+2(2\theta_1)]} \frac{\theta_0}{[(2\theta_0+1)+2\theta_1)]} P^1_{0,0}, \tag{S.9}$$

$$P^2_{0,2} = \alpha_2 \frac{\theta_0}{(2\theta_0+2)} \frac{\theta_0}{(2\theta_0+1)} P^1_{0,0}. \tag{S.10}$$

Corollary II: If $(x < y)$ and $(\alpha < \beta)$ then $(x+\alpha)/(y+\beta) > (x/y)$ only if $(x/y < \alpha/\beta)$

It follows from Corollary II that in eqns. (S.9) and (S.10), $\frac{\theta_0+\theta_1}{(2\theta_0+2)+2(2\theta_1)} > \frac{\theta_0}{(2\theta_0+2)}$, and also, $\frac{\theta_0}{(2\theta_0+1)+(2\theta_1)} < \frac{\theta_0}{(2\theta_0+1)}$. Thus, whether $P^2_{2,0} \neq P^2_{0,2}$ is determined by the condition $\frac{\theta_0+\theta_1}{(2\theta_0+2)+2(2\theta_1)} \frac{(2\theta_0+2)}{\theta_0} \neq \frac{\theta_0}{(2\theta_0+1)} \frac{(2\theta_0+1)+2\theta_1}{\theta_0}$, which simplifies to $\frac{1}{4\theta_0} - \theta_0 - 0.25 \neq \theta_1$. Here $\theta_0 < 1$ in general.

Corollary III: It follows from Corollary I, in general, that for a given system of matrix size $n = h+1$, i.e., $\max(i+j) = h+1$, the $P^{h+1}_{k,l}$'s for $\max(i+j) = h-1$ are given as functions of entries $P^h_{k,l}$ s.t., $\max(k+l) = h$, multiplied by the factor α_{h+1}, where $\alpha_{h+1} = 1/\sum P^{h+1}_{i,j} \forall (i,j)$.

$i =$ PyTCR, $j =$ OT1	$j=0$	$j=1$	$j=2$	$\dots$	$j=h-1$	$j=h$
$i=0$	$P^h_{0,0}$	$P^h_{0,1}$	$P^h_{0,2}$	..	$P^h_{0,h-1}$	$P^h_{0,h}$
$i=1$	$P^h_{1,0}$	$P^h_{1,1}$	$P^h_{1,2}$	..	$P^h_{0,h-1}$	
$\vdots$	$\vdots$	$\vdots$	$\vdots$	..		
$i=h-1$	$P^h_{h-1,0}$	$P^h_{h-1,1}$				
$i=h$	$P^h_{h,0}$					

Table S3: Probability table for $n = h$.

Thus, for the general case for matrix $\max(i+j) = h$, comparing $P^h_{h,0}$ vs. $P^h_{0,h}$, we get

$$P^h_{h,0} = \alpha_h \frac{(\theta_0+(h-1)\theta_1)}{[(2\theta_0+h)+h(2\theta_1)]} \cdots \frac{(\theta_0+\theta_1)}{[(2\theta_0+2)+2\theta_1)]} \frac{\theta_0}{[(2\theta_0+1)+2\theta_1)]} P^1_{0,0}, \tag{S.11}$$

$$P^h_{0,h} = \alpha_h \frac{\theta_0}{(2\theta_0+h)} \cdots \frac{\theta_0}{(2\theta_0+2)} \frac{\theta_0}{(2\theta_0+1)} P^1_{0,0}, \tag{S.12}$$

for constant α_h. Note that equations (S.11) and (S.12) are functions of $P^1_{0,0}$.

And, thus the $P^h_{h,0} \neq P^h_{0,h}$ holds true depending on the condition if

$$\prod_{m=1}^{h} \big(1 + (m-1)\frac{\theta_1}{\theta_0}\big) \neq \prod_{m=1}^{h} \big(1 + \frac{\theta_1}{\theta_0/m + 0.5}\big) \tag{S.13}$$

as every multiplicative terms, factors after $h > 2$ has a potential to yield $\frac{\theta_0+(h-1)\theta_1}{(2\theta_0+h)+h(2\theta_1)} > \frac{\theta_0}{(2\theta_0+h)}$ for large h, following the Corollary II, for $\frac{(h-1)\theta_1}{h(2\theta_1)} > \frac{\theta_0}{(2\theta_0+h)}$, i.e., $\frac{1}{2} - \frac{1}{2h} > \frac{1}{(2+h/\theta_0)}$ for $\frac{1}{2} - \frac{1}{2h}$ goes from 0 to 1/2, while $\frac{1}{(2+h/\theta_0)}$ goes from 1/2 to zero for h tending from 1 to infinite. In other words, when h gets larger ($h \gg 1$), L.H.S of condition in equation (S.13) turns greater than R.H.S of the condition.

It follows that for small clusters with only one type of cells, the probabilities for the number of OT1 cells to be greater than that of PyTCR cells is greater, i.e., $P^h_{h,0} < P^h_{0,h}$, whereas for large clusters the probabilities for the number of PyTCR cells to be greater than that of OT1 cells is greater, i.e., $P^h_{h,0} > P^h_{0,h}$.

For the case of diagonal entries of matrix size $\max(i+j) = 3$, we can write

$$P^3_{1,1} = \alpha_3 \frac{(\theta_0 + \theta_1)P^2_{1,0} + \theta_0 P^2_{0,1}}{(2\theta_0 + 1) + (2\theta_1)}. \tag{S.14}$$

Here, the proportional contribution of $P^2_{1,0}$ on $P^3_{1,1}$ is greater than that by $P^2_{0,1}$ on $P^3_{1,1}$, whereas, we also get that the proportion of exits from $P^3_{1,1}$ towards $P^2_{1,0}$ is less than that of $P^2_{0,1}$ as $\frac{1}{2\theta_0+1+\theta_1}P^3_{1,1} < \frac{1}{2\theta_0+1}P^3_{1,1}$ from the respective functions of $P^2_{1,0}$ and$P^2_{0,1}$ from steady state equations. Therefore, there is a net differential (flow) of probabilities from PyTCR cells towards the cells with more OT1 cells of the matrix size $n = 3$.

Furthermore, for the general case of the above, i.e., diagonal entries of matrix size $\max(i+j) = 2h+1$ (odd number), we can show that

$$P^{2h+1}_{h,h} = \alpha_{2h+1} \frac{(\theta_0 + h\theta_1)P^{2h}_{h,h-1} + (\theta_0 + (h-1)\theta_1)P^{2h}_{h-1,h}}{(2\theta_0 + 2h) + h(2\theta_1)}. \tag{S.15}$$

Here, when $h \to \infty$, the proportional contribution from the entries of $P^{2h}_{h,h-1}$ and $P^{2h}_{h-1,h}$ of matrix $n = 2h$ converges to a single value (constant). The convergence is also true for the proportionate exits from $P^{2h+1}_{h,h}$ into the two entries $P^{2h}_{h,h-1}$ and $P^{2h}_{h-1,h}$. Thus, it follows that when h is larger, the flow of probability from PyTCR cells towards the cells with more OT1 cells converges to zero.

Any probabilities of mirror entries on either side of (h,h) entry of large h, we note that

$$P^{h+1}_{h+1,h} = \alpha_{h+1} \frac{(\theta_0 + (h+1)\theta_1)P^h_{h+1,h-1} + (\theta_0 + h\theta_1)P^h_{h,h}}{(2\theta_0 + 2h + 1) + (h+1)(2\theta_1)}, \tag{S.16}$$

$$P^{h+1}_{h,h+1} = \alpha_{h+1} \frac{(\theta_0 + (h-1)\theta_1)P^h_{h-1,h+1} + (\theta_0 + h\theta_1)P^h_{h,h}}{(2\theta_0 + 2h + 1) + h(2\theta_1)}. \tag{S.17}$$

As before, as h increases, the contribution of $P^h_{h+1,h-1}$ towards $P^{h+1}_{h+1,h}$ converges to that of $P^h_{h-1,h+1}$ towards $P^{h+1}_{h,h+1}$ as per the Corollary II. Thus, for large h, $P^{h+1}_{h+1,h} > P^{h+1}_{h,h+1}$ depending on $P^h_{h+1,h-1} > P^h_{h-1,h+1}$, for contributions of $P^h_{h,h}$ towards $P^{h+1}_{h+1,h}$ and $P^{h+1}_{h,h+1}$ converge.

Furthermore, the special case of any two mirror entries $(h, 1)$ and $(1, h)$ of matrix size $\max(i+j) = h+1$

$$P^{h+1}_{h,1} = \alpha_{h+1}\frac{(\theta_0 + h\theta_1)P^h_{h,0} + (\theta_0 + (h-1)\theta_1)P^h_{h-1,1}}{(2\theta_0 + h + 1) + h(2\theta_1)}, \tag{S.18}$$

$$P^{h+1}_{1,h} = \alpha_{h+1}\frac{\theta_0 P^h_{0,h} + (\theta_0 + \theta_1)P^h_{1,h-1}}{(2\theta_0 + h + 1) + (2\theta_1)}. \tag{S.19}$$

By Corollary II, here the coeffcients of respective entries in function $P^{h+1}_{h,1}$ are greater than that in the function of $P^{h+1}_{1,h}$ for larger h. Thus, it follows that $P^{h+1}_{h,1} > P^{h+1}_{1,h}$ as $P^h_{h,0} > P^h_{0,h}$ and $P^h_{h-1,1} > P^h_{1,h-1}$ as h gets larger.

For any two mirror entries (i, j) and (j, i) for the case $\max(i + j) = 3$ matrix, we can write

$$P^3_{2,1} = \alpha_3\frac{(\theta_0 + 2\theta_1)P^2_{2,0} + (\theta_0 + \theta_1)P^2_{1,1}}{(2\theta_0 + 3) + 2(2\theta_1)}, \tag{S.20}$$

$$P^3_{1,2} = \alpha_3\frac{(\theta_0)P^2_{0,2} + (\theta_0 + \theta_1)P^2_{1,1}}{(2\theta_0 + 3) + (2\theta_1)}. \tag{S.21}$$

By comparing coefficients of equations (S.20) and (S.21), we note that the contribution of $P^2_{1,1}$ towards $P^3_{1,2}$ is greater than that towards $P^3_{2,1}$. From the same, the contribution of $P^2_{2,0}$ towards $P^3_{2,1}$ is greater than that of $P^2_{0,2}$ towards $P^3_{1,2}$ as per the Corollary II. Thus, if $P^2_{2,0} > P^2_{0,2}$, then $P^3_{2,1} > P^3_{1,2}$, depending on the magnitude of the contribution by $P^2_{1,1}$ towards each, and vice-versa.

For the general case of the above special case, for any (k, l) and (l, k) entries, s.t. $(l + k) = h$, and $l < k$ we get

$$P^h_{k,l} = \alpha_h\frac{(\theta_0 + k\theta_1)P^{h-1}_{k,l-1} + (\theta_0 + (k-1)\theta_1)P^{h-1}_{k-1,l}}{(2\theta_0 + h) + k(2\theta_1)}, \tag{S.22}$$

$$P^h_{l,k} = \alpha_h\frac{(\theta_0 + l\theta_1)P^{h-1}_{l,k-1} + (\theta_0 + (l-1)\theta_1)P^{h-1}_{l-1,k}}{(2\theta_0 + h) + l(2\theta_1)}. \tag{S.23}$$

For $k > l$, using Corollary II, we get that the coefficients of respective entries in $P^h_{k,l}$ are greater than those of mirror entries in $P^h_{l,k}$. Thus, it follows that $P^h_{k,l} > P^h_{l,k}$ if $P^{h-1}_{k-1,l} > P^{h-1}_{l-1,k}$ and $P^{h-1}_{k-1,l} > P^{h-1}_{l-1,k}$ as h gets larger ($h \gg 1$), also because the flow of probability through $P^h_{h,h}$ tends to zero for large h as we have shown above.

Thus, once the probability entries in the lower left triangle, i.e., where PyTCR cell numbers are higher, turn greater than those mirror entries in the upper right triangle, i.e., where OT1 cell numbers

are higher, depending on the condition in equation (S.13), and those follow similarly, the pattern of bias remains the same for large h regardless of how large the matrix is.

Explanation of the bias for the system of evolving probabilities over time. The solution to the system at time $t_2 = t_1 + \delta t$, evolved from time t_1 yield $\vec{P}_{i,j}(t_1 + \delta t) = e^{\mu A(\theta_0,\theta_1)\delta t}\vec{P}_{i,j}(t_1)$ where e denotes the matrix exponent with elements $\in \Re$, can be written as $e^{A\delta t} = I + \frac{A\delta t}{1!} + \frac{A^2\delta t^2}{2!} + \ldots$, where I is the identity matrix. Assuming higher orders (> 1) of the series expansion are negligible compared to its first order solution, we can write the solution to the system, evolved from time t_1 to t_2 as $\vec{P}_{i,j}(t_2) = (I + \mu A(\theta_0,\theta_1)\delta t)P_{i,j}(t_1)$, for elements in A and t are < 1. This yields $\vec{P}_{i,j}(t_1+\delta t) - \vec{P}_{i,j}(t_1) = \mu A(\theta_0,\theta_1)\delta t P_{i,j}(t_1)$. Note that the steady state solution was obtained by setting this difference to 0. We can write the above as $\frac{\delta \vec{P}_{i,j}(t_1)}{\mu\delta t} = A(\theta_0,\theta_1)\vec{P}_{i,j}(t_1)$.

Thus, the solutions of $\frac{\delta \vec{P}_{i,j}(t_1)}{\mu\delta t}$ of the above to each cluster combination (i,j) as a function of $A(\theta_0,\theta_1)\vec{P}_{i,j}(t_1)$, can be written as, for example from equation (S.3), for the case $n = 1$, $\frac{\delta \vec{P}^1_{0,0}(t_1)}{\mu\delta t} = -P^1_{0,0}(t_1) + \frac{P^1_{1,0}(t_1)+P^1_{0,1}(t_1)}{2\theta_0}$, and similarly from equation 2,$\frac{\delta \vec{P}^1_{1,0}(t_1)}{\mu\delta t} = -P^1_{1,0}(t_1) + \frac{\theta_0 P^1_{0,0}(t_1)}{(2\theta_0+1)+2\theta_1}$, and so on $\forall(i,j)$ combinations given for the case $n = 1$ and so forth for all $n = 2,3,4,\ldots$. These can be also written as $P^1_{0,0}(t_1) + \frac{\delta \vec{P}^1_{0,0}(t_1)}{\mu\delta t} = \frac{P^1_{1,0}(t_1)+P^1_{0,1}(t_1)}{2\theta_0}$, and $P^1_{1,0}(t_1) + \frac{\delta \vec{P}^1_{1,0}(t_1)}{\mu\delta t} = \frac{\theta_0}{(2\theta_0+1)+2\theta_1}P^1_{0,0}(t_1)$ and so forth for all $n = 2,3,4,\ldots$. Therefore, the patterns in bias, explained under the steady state solutions,$P^n_{i,j}$ in the main text are qualitatively similar for solutions $\left(P^n_{i,j}(t_1) + \frac{\delta P^n_{i,j}(t_1)}{\mu\delta t}\right)$ for all combinations of (i,j), and also for matrices A of any size n. This is because $\left(P^n_{i,j}(t_1) + \frac{\delta P^n_{i,j}(t_1)}{\mu\delta t}\right)$ are functionally equivalent to $P^n_{i,j}$ solutions under the steady sate condition. Thus, the bias discussed in the main text is similar for any $P^n_{i,j}(t_m)$ computed discretely in the form of $P^n_{i,j}(t_2) = P^n_{i,j}(t_1) + \delta\vec{P}^n_{i,j}(t_1)$ step-wise, starting from $t = 0$ to any t, choosing δt, s.t., $\mu\delta t = 1$.

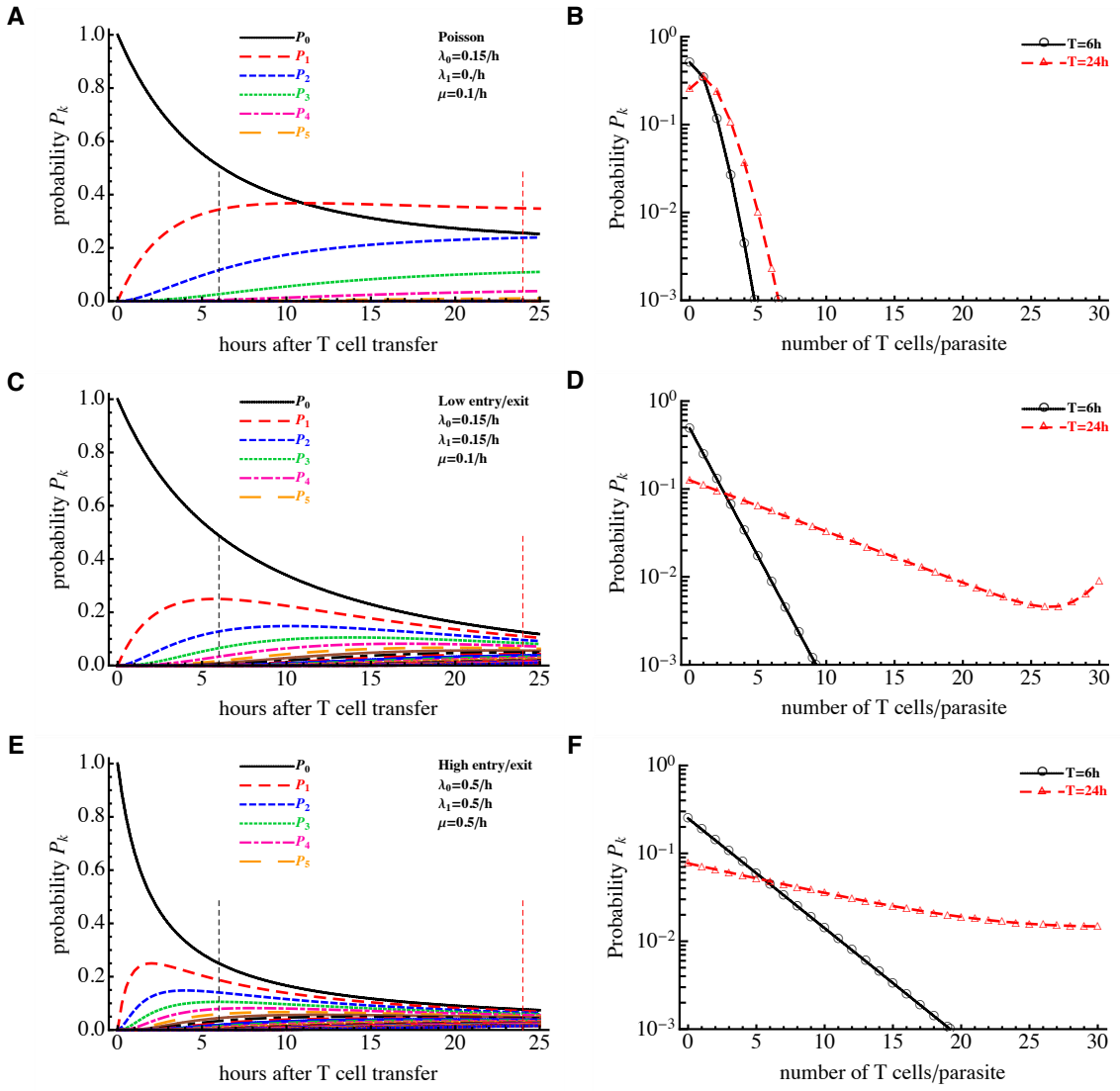

Figure S1: Dynamics of cluster formation as predicted by the random entry/exit (Poisson) and density-dependent (DD) recruitment models. We simulated the dynamics of cluster formation using eqns. (1)–(2) in the main text and plot the probability to observe k T cells around a parasite, $P_k(t)$, as predicted by the Poisson model (panels A-B, $\lambda_0 = 0.15$/h, $\mu = 0.1$/h), DD recruitment model with low entry into/exit from the cluster rats (panels C-D, $\lambda_0 = 0.15$/h, $\lambda_1 = 0.15$/h, $\mu = 0.1$/h), or the DD recruitment model with high entry into/exit from the cluster rates (panels E-F, $\lambda_0 = 0.5$/h, $\lambda_1 = 0.5$/h, $\mu = 0.5$/h). We show the dynamics of the probability P_k over time (panels A, C, E) or the distribution of cluster sizes at 6 or 24 hours after start of cluster formation (panels B, D, F).

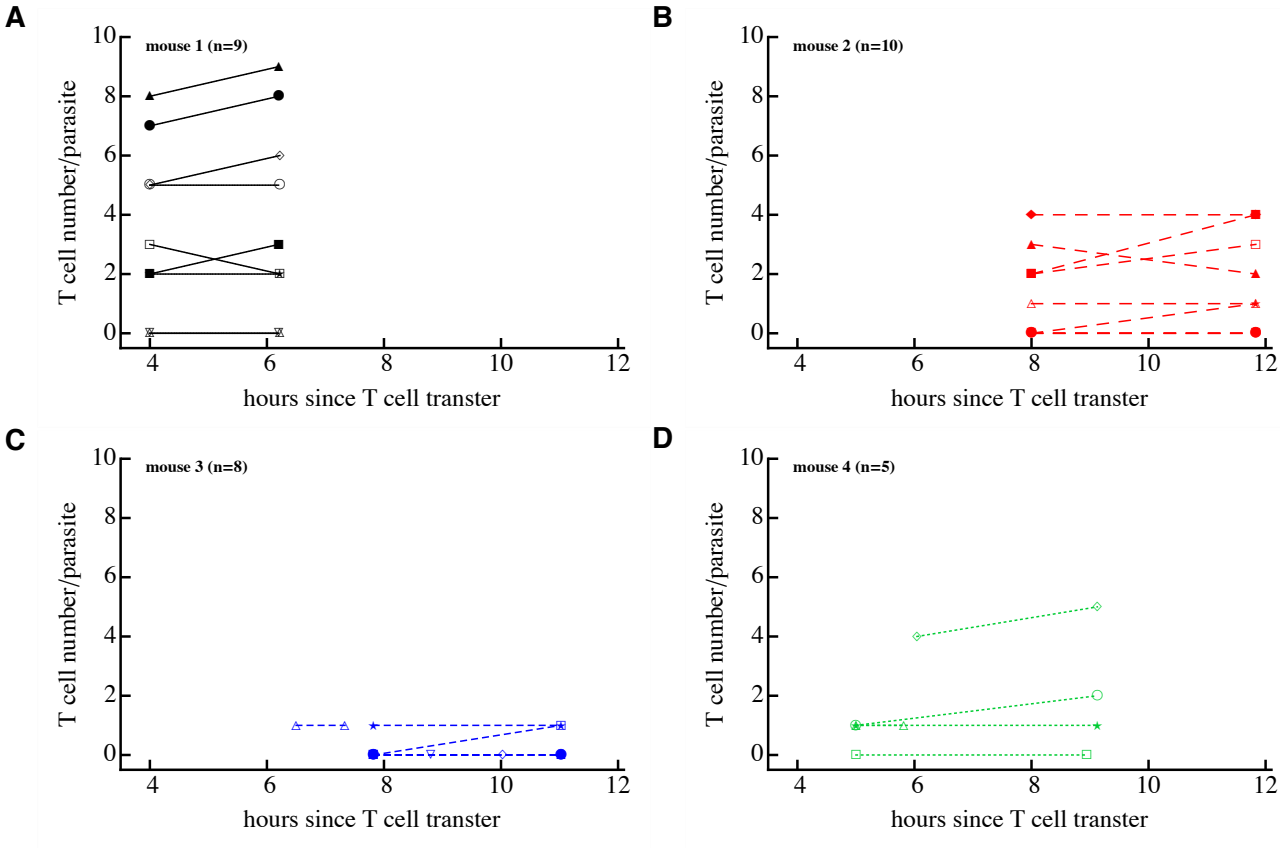

Figure S2: Moderate change in the T cell cluster size over time. We performed imaging experiments as described in Figure 6A and counted the number of T cells found around individual parasites at start and end of intravital imaging done after T cell transfer. Individual panels show change in T cell cluster size around $n = 32$ parasites in four individual mice. Imaging of T cell clusters started at different times in individual mice and followed for different lengths of time. Note that as we observed before, 12 parasites had no T cells near them at both observations. Overall, there was a statistically significant but small change in the cluster size of the imaging period (as summarized in Figure 6B).

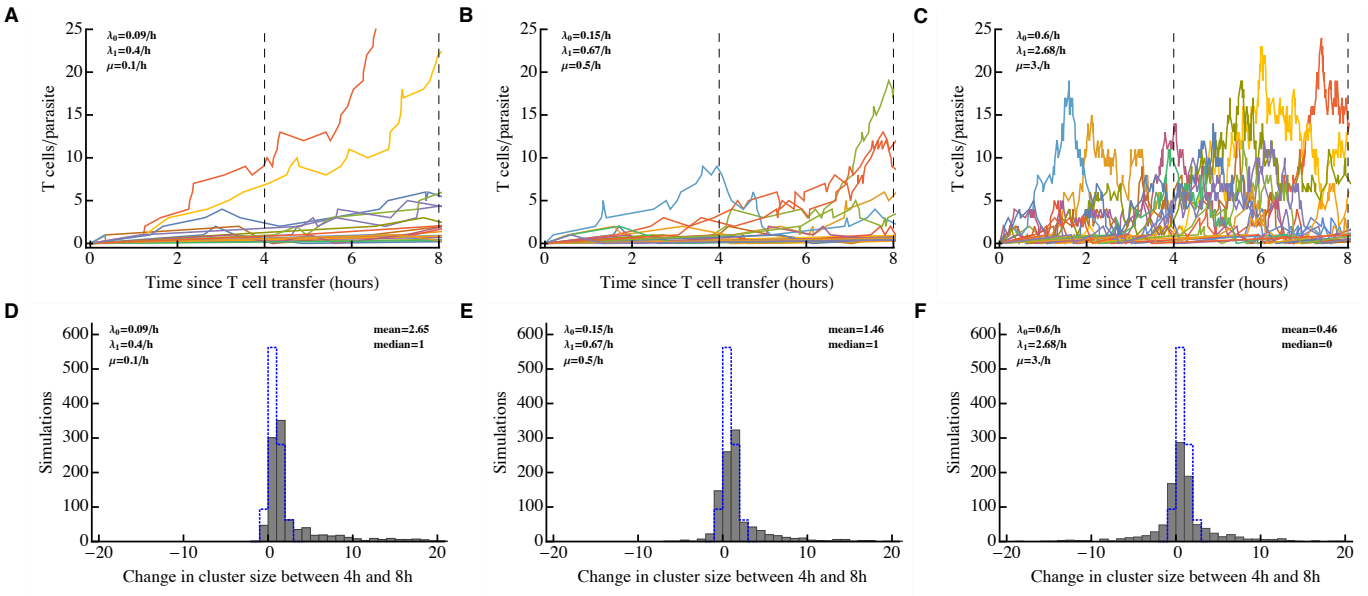

Figure S3: Stochastic simulations of cluster formation suggest an upper limit on the rate of T cell exit from the clusters. We ran Gillespie simulations of the cluster formation assuming different constant (time-independent) values for the entry rates into the cluster (λ_0 and λ_1) and exit rates from the cluster (μ) found by fitting the DD recruitment model to experimental data in Figure 6C-D. Three values of the exit rate were fixed: $\mu = 0.1/\text{h}$ (panels A&D), $\mu = 0.5/\text{h}$ (panels D&E), and $\mu = 3/\text{h}$ (panels C&F) and remaining parameters were estimated by fitting the model (eqns. (1)–(2)) to data (Figure 2B). These parameters are shown on individual panels. We simulated changes in cluster size for $n = 10^3$ parasites. Panels A-C show sample trajectories of cluster sizes of 20 of such simulations, and panels D-F show the change in the size of the cluster between 4 and 8 hours after start of simulation for all simulations (solid bars) or changes in cluster sizes as was observed in experimental data (dotted bars, see also Figure 6B). These simulations indicate that at high exit rates ($\sim \mu = 1 - 3/\text{h}$) and at high entry rates there are large fluctuations in the cluster sizes between 4 and 8 hours (panels C&F) which is not observed in experimental data. Thus, in the 4-8 hour time period exit and entry rates cannot be extremely large for the DD recruitment model to be consistent with experimental data. Furthermore, simulations with smaller rates (panels A&D) also indicate increase in the average cluster size over time (since $\lambda_1 > \mu$) which is also not consistent with the change in cluster size at 4-8 hours post T cell transfer.

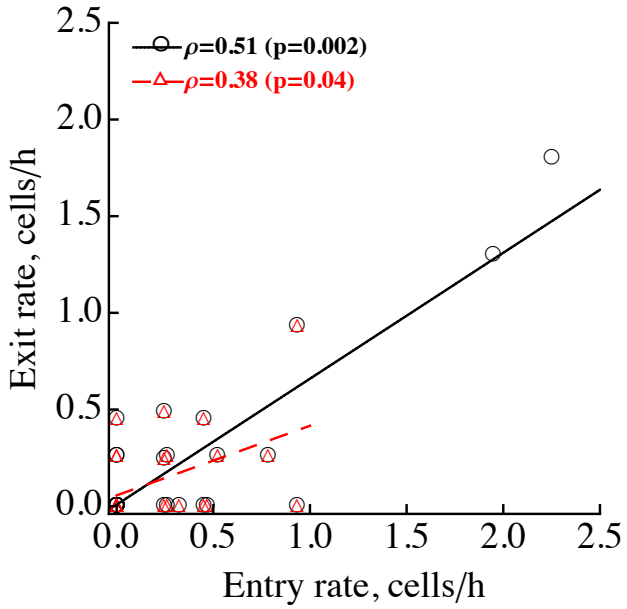

Figure S4: Experimentally measured rate of T cell exit from the cluster correlates with the rate of T cell entry into the cluster. We plotted the correlation between the experimentally measured number of T cells coming with a 40 μm radius of a given parasite per unit of time (entry rate, see Figure 6E) and the number of T cells leaving a given cluster per unit of time (exit rate, see Figure 6F) for $n = 32$ parasites. P-values were calculated using Spearman Rank correlation test (with correlation coefficient ρ indicated), and lines indicate trends of the correlation found using a linear regression. The statistical significance of the correlation is shown for all data (circles) or for data that excluded two potential outliers (triangles).