Mathematical modeling of T cell clustering following malaria infection in mice

A Thesis Presented for the

Master of Science

Degree

The University of Tennessee, Knoxville

Reka Katalin Kelemen

May 2014

© by Reka Katalin Kelemen, 2014
All Rights Reserved.

I would like to dedicate this work to all the mice, whose lives and bodies were sacrificed in this study to gain insight into the mammalian immune response. They have not been forgotten, and here I show my efforts to create as much knowledge as I can from the data that was generated using them.

Abstract

Malaria is the result of the immune system's unsuccessful clearance of hepatocytes (liver cells) infected by the eukaryotic pathogen of the *Plasmodium* genus. It has been shown that CD8 T cells are required and sufficient for protective immunity against malaria in mice [29, 36], but the mechanisms by which they find and eliminate infected hepatocytes are not known yet. Recently we reported the formation of CD8 T cell clusters consisting of up to 25 cells around infected cells [8]. Our mathematical modeling and data analysis revealed that malaria-specific T cells likely recruit each other and also non-malaria-specific T cells to infected hepatocytes in a process that is G-protein-coupled receptor (GPCR) dependent. Evidence exists for variable T-cell-extrinsic factors in explaining the large variation of cluster sizes around infected cells, for which we propose further experiments to discriminate between T-cell-dependent and -independent mechanisms in cluster formation. Our results suggest that CD8 T cells specific to an ongoing infection tend to cluster around infected cells and their presence also enhances irrelevant CD8 T cells' propensity to clustering. However, analysis of T cell movement tracks shows that irrelevant T cells have a significantly lower average frequency of movements towards the parasite than antigen-specific T cells. Since clusters of CD8 T cells might be more efficient at killing infected cells than a CD8 T cell alone, understanding the requirements for large T cell clusters might have important implications in the fields of immunology and medicine.

Table of Contents

1 Mathematical modeling of CD8 T cell cluster size distributions supports the hypothesis of recruitment by antigen-specific T cells 1

1.1 Introduction 2

1.2 Materials and methods 3

1.2.1 Data 3

1.2.2 Mathematical models 5

1.2.3 Model fitting and selection 8

1.3 Results 9

1.4 Discussion 13

2 Classification of T cell movement tracks allows for prediction of cell function 16

2.1 Introduction 16

2.2 Implementation 18

2.2.1 Dataset 1: PyTCR and OT-1 cells in infected mice 19

2.2.2 Dataset 2: PyTCR cells in infected mice 19

2.2.3 Dataset 3: Parasite specific T cells in infected and uninfected mice 20

2.2.4 Movement Attributes 20

2.2.5 Spatial Analysis 22

2.2.6 Classifier: Support Vector Machine 23

2.2.7 Hierarchical clustering 23
2.3 Results and Discussion 23
2.3.1 T cell movement with respect to distance from parasite 23
2.3.2 Spatial Analysis 25
2.3.3 Classifier: Support Vector Machine 26
2.3.4 Hierarchical Clustering 31
2.4 Conclusions and Future Work 33

3 Exploring the possible effects of T-cell-extrinsic and T-cell-intrinsic factors in driving the clustering of T cells at sites of infection 36
3.1 Introduction 36
3.2 Materials and methods 38
3.2.1 Data 38
3.2.2 Mathematical models 39
3.3 Results 43
3.3.1 Entry and exit rates observed 4-8 hours PTCT suggest that cluster formation is near steady state at the time of imaging . 43
3.3.2 If cluster sizes are not at equilibrium, DD recruitment model still fits PyTCR data best 47
3.3.3 Tests of alternative hypotheses that assume variance in T cell entry rates due to T-cell-extrinsic or T-cell-intrinsic factors . . 49
3.4 Discussion 61

Bibliography 64

Vita 71

List of Tables

1.1 Parameter estimates (and their 95% CIs) for the three models (random entry/exit, density-dependent (DD) recruitment, density-independent (DI) exit) fitted to each dataset. Here θ is the ratio of the entrance to exit rate, and θ_1 denotes the density-dependent additional entrance rate, λ_1, divided by the exit rate, μ. The prediction of the model with DD recruitment was normalized assuming that $k_{max} = 30$.The lowest AIC score indicates the best fitting model for each dataset. AIC values were used to calculate Akaike weights shown in Figure 1.3. Dash (-), not applicable. 12

2.1 Summary of support vector machine results. Accuracies vary depending on input attributes, folds used for cross validation, and whether whole cells or overlapping *chunks* of 6 timepoints were used. Cross-validation folds are given, meaning the dataset was divided into 3 or 10 equal parts (folds), trained on *n-1* folds, and tested on 1 fold. Accuracy denotes the percent of the testing dataset that was correctly classified by the support vector machine. All distance attributes include path length, average jump, mininum jump, maximum jump, net x displacement vector, net y displacement vector, net z displacement vector, and total displacement. All angle attributes include mean angle, minimum angle, maximum angle, sum of angles, mean angle to parasite, minimum angle to parasite, maximum angle to parasite, and sum of angles to parasite). 29

2.2 Summary of support vector machine results: OT-1 and PyTCR data. The overall accuracy using 10-fold RBF and linear cross-validation was 67.7% and 72%, respectively. The number of correctly or incorrectly classified testing *chunks* of cell timepoints are given in the true and false categories for each cell type; the SVM appears to correctly classify at least 72% of the testing dataset. (**T**=true, **F**=false). 30

2.3 Summary of support vector machine results: infected (**inf**) versus uninfected (**uni**). The overall 10-fold RBF cross-validation accuracy was 69.8%. The number of correctly or incorrectly classified testing *chunks* of cell timepoints are given in the true and false categories for each cell type; the SVM appears to correctly classify at least 97% of the testing dataset in these iterations. (**T**=true, **F**=false). 31

3.1 DD recruitment model still fits best when using various fixed exit rates for fitting the "PyTCR alone" dataset using numerical solutions at 6 hours PTCT. We fixed the per capita exit rates, μ_0 values, to those shown in the top row, and used numerical solutions in a MLE approach to find best fit and estimate entry/exit rate parameters, of which the ones corresponding to the DD recruitment model are shown. AIC scores were then calculated for each model's numerical fit with exit rates fixed to the various values shown. 47

3.2 Estimated entry/exit rate ratio parameters of the DD recruitment model decrease as the exit rate used for model fitting is increased. At steady state, which is approached closely in 6 hours PTCT if μ_0=3/hour, the estimated rates are minimal. θ_1/θ_0 ≈4.3 consistently in all cases. 49

3.3 Gamma distributed entry rates explain PyTCR, OT-1 and PTx data well, but yield highly different means and variances, which suggest different effect of T-cell-extrinsic factors on different cell types. When we constrained either the mean or variance of entry rates consistently over datasets, AIC values showed no support for those models. Likelihood ratio test was also performed and found a significant increase in fit quality if added T-cell-type-specific means and variances of entry rates. The exit rate was fixed to 3/hour when estimating the entry rate parameters, and the table shows the ratio of estimated entry rates to the exit rate. A discrete model with two subsets of entry rates into clusters did not explain these three datasets well, as shown by the high AIC values. The value f stands for the fraction of infected cells with entry and exit rate ratio, θ_{0f_1}, thus $1-f$ is the fraction of infected cells with rates ratio θ_{0f_2}. 54

3.4 Estimated mean and variance of entry rates by OT-1 cells into clusters increase by nearly 10-fold when co-transferred with PyTCR cells, when assuming a Gamma model, where entry rates only vary because of T-cell-extrinsic reasons. In the density-dependent recruitment model, however, OT-1 cells have similar estimated basic entry rates in both experiments, and only the T-cell-density-dependent recruitment rates are changed. During fitting of the Gamma model, we used exit rates of 3/hour. 57

3.5 This table lists the basic entry/exit rate values $\theta_0 = \lambda_0/\mu_0$, which are shared by both cell types. Likelihood ratio found no significant increase in fit quality when estimating cell-type-specific entry rates. Additional entry rates, due to density-dependent recruitment, divided by exit rates are indicated by $\theta_1 = \lambda_1/\mu_1$. For instance, "PyTCR" in the column of θ_1 means that the recruitment rate/exit rate is exerted by PyTCR cells and extends onto both cell types, while "PyTCR:OT-1" denotes the recruitment rate/exit rate by PyTCR cells that is sensed by OT-1 cells. The mixed cluster model that defines the recruitment rates into clusters to be only dependent on the number of PyTCR cell density predicts the observed mixed clusters the best, as shown by AIC scores. The addition of an OT-1 dependent recruitment gives a fit with a similar AIC value, but likelihood ratio test of these nested models shows that this more complex model does not significantly increase the quality of the fit ($p > 0.25$). Interestingly, the estimated recruitment rate is also about 4.3 times the estimated basic entry rate. 60

List of Figures

1.1 Panels a), b) and c) give a cartoon representation of total entry,λ_k, and exit rates, μ_k, in and out of a cluster of size k, and their dependences on k (values shown on arrows). Random entry/exit, density-independent exit and density-dependent recruitment are shown in a), b) and c) respectively. Panel d) shows the time points of the imaging of endogenous CD8 T cells during malria-infection. These T cells were generated by immunization with radiation attenuted sporozoites (RAS) 10 days before infection. Panel e) demonstrates the experiments involving T cell transfer. Either OT-1 or PyTCR cells were transferred 20 hours infection or they were transferred together. PyTCR cells were either treated with pertussis toxin (PTx) or not, and the PTx experimental results will be discussed in Chapter 3. . . 4

1.2 The endogenous CD8 T cell cluster size distribution rejects the T-cell-density-independent, random model of clustering. "w" stands for AIC weights, which give the likelihood of a model relative to other models ($w < 0.01$ suggests an unlikely model). 10

1.3 Modeling of clustering formation by specific and nonspecific T cells. (A) Frequency distribution of (i) OT-1 CD8+ T cells and (ii) PyTCR CD8+ T cells around infected hepatocytes when transferred to different animals compared with the distribution that would be expected for three different models for T-cell clustering: random entry and exit (red line), density-independent exit (gray line), and density-dependent recruitment (green line). (B) Frequency distribution of (i) OT-1 CD8+ T cells and (ii) PyTCR CD8+ T cells around infected hepatocytes when transferred to the same animal compared with the distribution that would be expected for three different models for T-cell clustering: random entry and exit (red line), density-independent exit (gray line), and density-dependent recruitment (green line). Akaike weights (w) for every model fitted to the experimental data are shown on individual panels. 11

1.4 PKH-26 labeled effector PyTCR cells (1×10^7) were transferred to mice 20 h after infection with 3×10^5 PyGFP and imaged 48 h later; data are pooled from 32 movies in four independent experiments. Correlation of (i) entry rate and (ii) per capita exit rate of PyTCR with the total number of PyTCR cells around each infected hepatocyte. Infected hepatocytes not associated with any T cell were excluded from the analysis. 13

2.1 Image of malaria-specific PyTCR cells labeled with red fluorescent membrane dye (PKH-26) congregating around a green fluorescent protein expressing malaria parasite inside an infected liver cell. Dashed circle denotes a hypothetical outline of an infected hepatocyte (radius 40 μm). Image courtesy of Ian Cockburn. 19

2.2 Movement attributes of T cells. The distance between two consecutive positions of a T cell was considered a movement, and the sum of all movements for a T cell was its total path length. The distance and the vector between the start and end points of a T cell are called its displacement and displacement vector. 21

2.3 Direction trends of PyTCR (malaria-specific) and OT-1 (non-specific for malaria) cells (Dataset 1) based on distance from infection site. The direction of T cell movement at each time point was grouped by distance from parasite: up to 39 μm (near a parasite-infected cell, see Figure 2.1) or farther than 40 μm (away from a parasite-infected cell). The probability of moving towards the parasite is calculated by dividing the number of angles less than 90^0 relative to the direction of the infection site by total number of data points at that distance interval. Both cell types head towards the site of infection at a probability greater than 50%, suggesting a general tendency for T cells to move towards the parasite. However, at distances greater than 40 μm, malaria-specific (PyTCR) cells will travel towards the parasite more often than nonspecific (OT-1) cells. The number of data points analyzed at each bin is indicated by n. Distance bins with less than 3 data points are not included. NS = not significant. $*P < 0.05$ and $**P < 0.002$ (unpaired two-tailed t-test, Kruskall-Wallis rank sums, binomial test). 25

2.4 Colors represent density based clusters of T cell positions at all timepoints from a parasite-infected mouse liver (Dataset 2). Blue circles are cell positions that did not cluster. These colors correspond to categories of the *zone* attribute used in the SVM, which highlight potential areas of intense traffic denoted by many dots and a single color that may include one or more individual cells. While this dataset was too sparse to visualize complete *highways*, this method could be used to predict junctions of sinusoids or the presence of recruiting cells in the liver with a larger dataset. 26

2.5 Each of these images contains malaria-specific CD8+ T cell paths that belong to different clusters based hierarchical clustering of different movement patterns (Section 3.4). We see that T cells tend to move either in small jumps and remain localized (top panel) or in large jumps and traverse large distances (bottom panel). Colors represent different cells. 32

2.6 Hierarchical clustering of parasite-specific T cell movement patterns in an infected mouse (red bars) and an uninfected mouse (blue bars) based on mean angle, sum of angles, max. angle, min. angle, net y displacement, path length, max jump, total displacement, net x displacement, avg. jump, and min. jump. Clustering patterns are visible in the attribute similarities, but do not appear to correlate with whether a CD8+ T cell is from an infected or uninfected mouse. . . . 34

3.1 Cartoon representation of three models introduced in this chapter. a) shows the Gamma distributed entry rate model, where α and β are the rate and shape parameters for the distribution that the T cell entry rates for each infected cell are drawn from. b) represents the two entry rate model, where a fraction f of the infected hepatocyte population is associated with one entry rate, and $1 - f$ fraction is associated with another entry rate. c) demonstrates a case under the mixed cluster modeling framework, where the recruitment rate is dependent on the density of only cell type 1. 41

3.2 Cluster size changes measured 4-8 hours PTCT are small, but significantly depend on the mean number of cells in the cluster.(A and B) Mean entry and per capita exit rates as a function of the mean cells in the clusters between 4 and 8 hours PTCT. A regression line is shown in (A), and the estimated basic entry rate and recruitment rate are shown. F-test yielded a significant dependence on cluster size. (B) shows the mean per capita exit rate, which was found to be independent of cluster size (F-test p=0.5). (C) 32 observed clusters show cluster size change that is not significantly different from an expected mean change of 0.25 cells. (D) When simulating cluster size changes 4-8 hours PTCT using our DD recruitment model with $\mu_0 = 3$/hour, we have found a mean change of 0.25 with the displayed distribution. 44

3.3 Probabilities of cluster sizes were estimated with our DD recruitment model using entry and exit rates observed 4-8 hours PTCT or using our steady state parameters $\theta_0 = 0.2$, $\theta_1 = 0.88$, and percent differences from the observed frequencies were calculated. (A) shows that using parameters observed 4-8 hours PTCT the distribution at 6 hours PTCT has too low probabilities for large clusters. (B) demonstrates the percent differences from observed if we use the estimated steady state distribution. (C and D) We sampled 1000 times a 130 clusters from our obtained distributions 6 hours PTCT and repeated the process 1000 times. Shown are the histograms of the probabilities of seeing clusters larger than 25 cells. The DD recruitment model cannot give rise to any clusters of those sizes with the parameters observed 4-8 hours PTCT, while at steady state the average probability of large clusters is 0.4. . 46

3.4 Percent differences from steady state probabilities as the per capita exit rate, μ_0, is fixed to the values indicated. The DD recruitment model was used to obtain the cluster size distributions at 6 hours PTCT. As apparent in the figure, a per capita exit rate of $\mu_0 = 3/hour$ results in cluster size frequencies close to the estimated steady state. 48

3.5 Fitting our three basic models, which are described in section 1.2.2 to the PTx data we find not enough support from the DI exit and DD recruitment models to reject the Random entry and exit model. "w" stands for AIC weights, which give the likelihood of a model relative to other models ($w < 0.01$ suggests unlikely model). 50

3.6 Gamma distributed entry rate models fit PyTCR, OT-1 and PTx datasets well, but with high discrepancies between estimated distributions of entry rates of the cell types. When mean entry rate is shared among three cell types the comparison of fits and AIC weights of these models to the respective best models shows no statistical support for the hypothesis that variance in entry rate alone can explain small and large clusters. Exit rates were fixed to 3/hour. 53

3.7 Endogenous CD8 T cells do not show strong evidence for T-cell-extrinsic reasons to observed cluster distribution, based on our Gamma distributed entry rate model and model with two subsets of infected cells with different T cell entry rates associated with them. "w" stands for AIC weights, which give the likelihood of a model relative to other models ($w < 0.01$ suggests unlikely model). 56

3.8 (A) Estimated probability distribution of mixed cluster sizes according to the best fitting model, where only PyTCR cells recruit, with data overlaid using points whose diameters are proportional to the number of clusters observed. Legend shows the natural logarithm of the probability values. (B and C) Estimated and observed probabilities of PyTCR and OT-1 cell numbers around infected cells when transferred together. 59

Chapter 1

Mathematical modeling of CD8 T cell cluster size distributions supports the hypothesis of recruitment by antigen-specific T cells

This chapter contains parts of a previous publication by the author, in which she contributed to data analysis and writing the paper.

I. A. Cockburn, R. Amino, R. K. Kelemen, S. C. Kuo, S. W. Tse, A. Radtke, L. Mac-Daniel, V. V. Ganusov, F. Zavala, and R. Menard. In vivo imaging of CD8+ T cell-mediated elimination of malaria liver stages. Proc. Natl. Acad. Sci. U.S.A., 110(22):90909095, May 2013

1.1 Introduction

CD8+ T cells are specialized cells of the adaptive immune system that specifically kill pathogen-infected cells and tumors presenting cognate antigen on MHC class I molecules. In malaria infection, antigen-specific CD8 T cells are capable of eliminating pre-erythrocytic liver-stage parasites that develop in hepatocytes [25]. The liver stages develop after sporozoites, the parasite form injected by the mosquito, migrate to the liver and infect hepatocytes. In natural infections only a few dozen liver stages exist in the host, which mature to form merozoites that establish blood stage infection. The maturation time of the liver stages is generally short: from 48 h with rodent-infecting *Plasmodium* species to 7 d with *Plasmodium falciparum*; therefore, to be protective, liver stage-specific CD8 T cells must rapidly find and eliminate the rare infected hepatocytes. We were the first group to successfully image the killing of an infected cell by a CD8 T cell in vivo and record the number of immune cells involved in each killing process. In contrast with previous belief, supported by in vitro experiments, that CD8 T cells eliminate infected cells mostly alone, this in vivo study of the immune response to liver-stage malaria reported that 64% of the observed infected cells were closely associated with more than one T cell. The imaged T cell clusters were in some cases as big as 25 cells for a single infected cell, which raises questions about the mechanism that would allow for the formation of such large clusters. The function and benefit of T cell cluster formation around infected cells, as well as the driving forces of clustering are unknown and their understanding requires analysis of the in vivo imaging data. We use datasets of observed T cell cluster size frequencies, which were obtained for five experimental setups, to test several hypotheses about the process of CD8 T cell cluster formation during malaria infection. The data allow us to analyze the behavior of fluorescently labeled malaria-specific T cells (PyTCR cells) and non-malaria-specific T cells (OT-1 cells) either transferred into mice separately or together. We use quantitative methods and mathematical models to compare alternative hypotheses about the driving forces and

possible requirements behind CD8 T cell cluster formation during liver-stage malaria in mice.

1.2 Materials and methods

1.2.1 Data

Endogenous CD8 T cells

To visualize the interaction between activated CD8+ T cells and *Plasmodium*-infected hepatocytes in vivo, our experimental collaborators undertook imaging in the mouse liver using spinning-disk confocal microscopy [33]. GFP-expressing *Plasmodium yoelii* sporozoites (PyGFP) were used to infect naïve mice or mice immunized 10 d previously with *P. yoelii* radiation-attenuated sporozoites (RAS). To visualize CD8+ T cells, Phycoerythrin (PE)-conjugated -CD8 antibodies were injected into the mice 24 h after infection. The mice were then immediately anesthetized and subjected to surgery to expose the liver for imaging. At each imaged infected hepatocyte the number of clustering CD8 T cells within a 40 μm-diameter circle were recorded for a total of 66 infected cells. Hepatocytes are approximately 40 μm long, so this choice of diameter for defining a cluster is aiming to include T cells that are in direct contact with the infected cell.

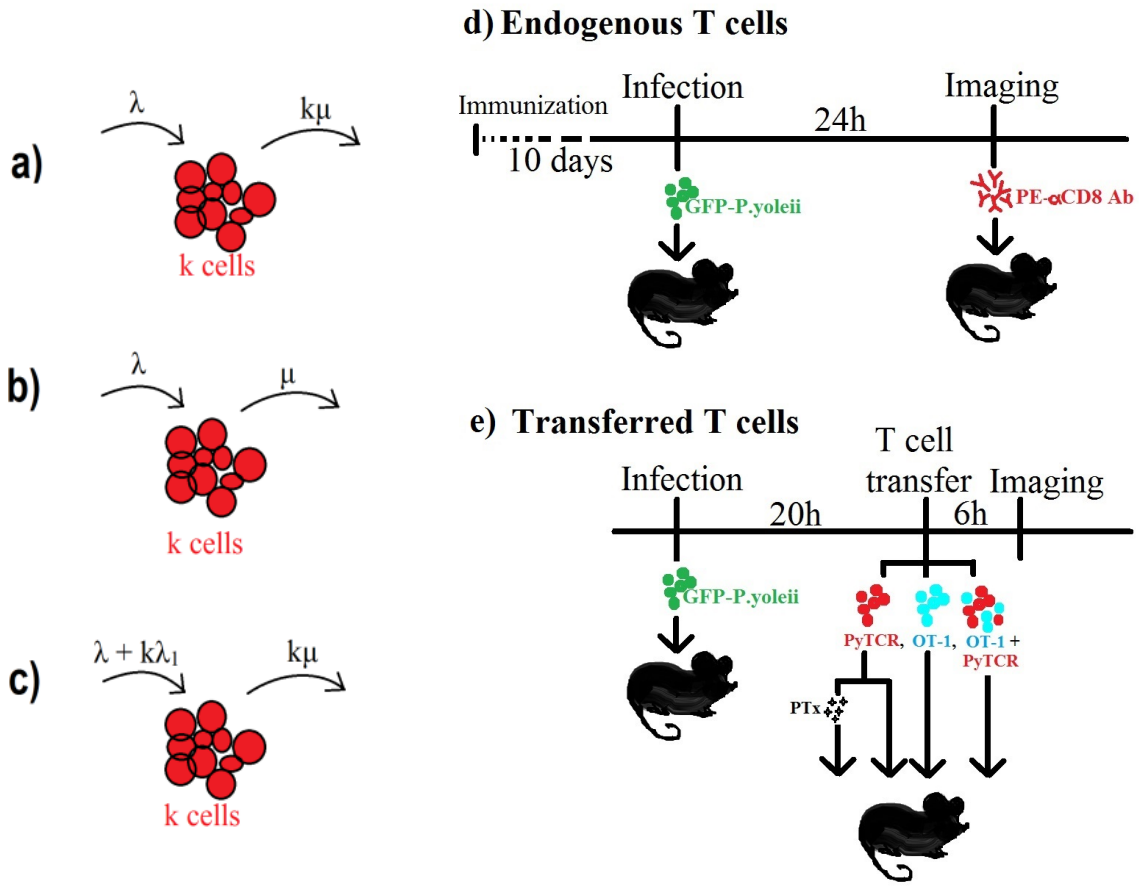

Figure 1.1: Panels a), b) and c) give a cartoon representation of total entry,λ_k, and exit rates, μ_k, in and out of a cluster of size k, and their dependences on k (values shown on arrows). Random entry/exit, density-independent exit and density-dependent recruitment are shown in a), b) and c) respectively. Panel d) shows the time points of the imaging of endogenous CD8 T cells during malria-infection. These T cells were generated by immunization with radiation attenuted sporozoites (RAS) 10 days before infection. Panel e) demonstrates the experiments involving T cell transfer. Either OT-1 or PyTCR cells were transferred 20 hours infection or they were transferred together. PyTCR cells were either treated with pertussis toxin (PTx) or not, and the PTx experimental results will be discussed in Chapter 3.

Malaria-specific and non-malaria-specific CD8 T cells

Using α-CD8 antibodies we could not discriminate whether the CD8+ T cells clustering around infected hepatocytes were parasite-specific. Moreover this technique is not compatible with time-lapse imaging because antibodies may alter the behavior

of labeled T cells. To overcome these limitations, we then examined the behavior of in-vivo-activated T cell receptor (TCR) transgenic CD8+ T cells specific for the CS280-288 (SYVPSAEQI) epitope of the circumsporozoite protein of *P. yoelii* (PyTCR cells) in the context of H^2-K^d [26]. As a control the behavior of similarly activated OT-I cells that recognize the irrelevant SIINFEKL epitope from chicken ovalbumin in the context of H^2-K^b were also examined. To ensure histocompatibility, both TCR transgenic donors and the wild-type recipient mice were BALB/c $\times$ C57BL/6 F1 (CB6) hybrid mice. The in vivo activated CD8+ T cells were labeled and transferred, either separately or together, into mice that had been inoculated 20 h previously with PyGFP sporozoites. Cells were transferred 20 h after infection so as to prevent parasites from being eliminated before they become visible. Six hours after transfer, mice were anesthetized and subjected to surgery to expose the liver, and interactions between infected hepatocytes and transferred CD8+ T cells were imaged by either static or time-lapse imaging. For a total of 130 infected hepatocytes the numbers of PyTCR cells were recorded in a 40 μm-diameter vicinity. The numbers of OT-1 cells were recorded around 92 infected hepatocytes. Equal amounts of PyTCR and OT-1 cells were transferred into mice together in a third experiment and the numbers of each cell type were recorded in the mixed clusters around 52 infected hepatocytes.

1.2.2 Mathematical models

To describe the kinetics of formation of clusters around a *P. yoelii*-infected hepatocyte we used a standard "birth-death" model. In this model we defined a probability to observe k cells in a cluster at time t as $P_k(t)$. Cells are recruited into the cluster at a rate λ_k and exit from the cluster at a rate μ_k. In general, both rates may depend on the number of cells in the cluster, and we are assuming an infinite labeled T cell population, as the entry rate does not depend on the availability of free-moving T cells. Due to the design of the T-cell-transfer experiments, there is expected to be a minimum of approximately 30 T cells available for every infected cell. The dynamics

of the probability to observe k cells in a cluster of lymphocytes around an infected cell at time t is given by the following system of differential equations

$$\begin{aligned}
\frac{dP_0}{dt} &= -\lambda_0 P_0 + \mu_1 P_1, \\
\frac{dP_1}{dt} &= -(\lambda_1 + \mu_1) P_1 + \mu_2 P_2 + \lambda_0 P_0, \\
\frac{dP_k}{dt} &= -(\lambda_k + \mu_k) P_k + \mu_{k+1} P_{k+1} + \lambda_{k-1} P_{k-1}, k > 1,
\end{aligned} \tag{1.1}$$

Because in our experimental data most clusters are largely formed before the time of imaging and few cells exit or enter the clusters in the observed time period, we focus on the analysis of the steady state distribution of cluster sizes. We relax this assumption in Chapter 3. The steady state distribution of cluster sizes can be found by letting each of the above equations equal 0. After some algebra we find

$$P_k = P_0 \frac{\prod_{i=0}^{k-1} \lambda_i}{\prod_{i=1}^{k} \mu_i} \tag{1.2}$$

and $P_{(k=0)} = P_0$ where P_0 is found by normalizing Eqn. 1.2 so the sum is equal to 1. The following models describe specific solutions depending on how the entrance rate into a cluster of size k, λ_k, and exit rate from a cluster of size k, μ_k, depend on the cluster size, k.

Density-independent recruitment, density-dependent exit model (random entry and exit)

Our first model assumes a constant, cluster size-independent entrance rate of cells into a cluster and constant per capita exit rate, $\lambda_k = \lambda$ and $\mu_k = k\mu$. In this scenario we assume that there is a constant influx of cells into cluster and the total rate of exit of any cell from a cluster is proportional to the cluster size. Then the probability of

forming a cluster of size k is

$$P_k = P_0 \frac{\prod_{i=0}^{k-1} \lambda_i}{\prod_{i=1}^{k} \mu_i} = P_0 \frac{\lambda^k}{\mu^k k!} = \frac{\theta^k}{k!} e^{-\theta}, \tag{1.3}$$

where $\theta = \frac{\lambda}{\mu}$.

Density-dependent recruitment model (cells are attracted to a cluster)

The density-dependent entry model modifies the entrance rate to be a combination of constant flux and density-dependent rate, $\lambda_k = \lambda + k\lambda_1$. The assumption is that cells in a larger size cluster will attract more cells than cells in a smaller size cluster. The per capita exit rate is constant, and thus the exit rate is $\mu_k = k\mu$. Cluster of size k has the following probability of forming

$$P_k = P_0 \frac{\prod_{i=0}^{k-1} \lambda_i}{\prod_{i=1}^{k} \mu_i} = P_0 \frac{\prod_{i=0}^{k-1} \lambda + i\lambda_1}{\mu^k k!} = P_0 \frac{\prod_{i=0}^{k-1} \theta + i\theta_1}{k!}, \tag{1.4}$$

where $\theta = \frac{\lambda}{\mu}$ and $\theta_1 = \frac{\lambda_1}{\mu}$ and P_0 is found by normalizing Eqn. 1.4 by letting the maximal cluster size to be k_{max}. In general, $\sum_{k=0}^{\infty} P_k \to \infty$ and therefore, the sum must be limited due to this reason.

Density-independent exit model (cells keep each other from leaving cluster)

Our third model leaves the entrance rate of cells into a cluster constant but considers the per capita exit rate to be density-dependent, so that the overall exit rate from the cluster is independent of the cluster size, i.e., $\lambda_k = \lambda$ and $\mu_k = \mu$. In this model there is a simple relationship between the number of lymphocytes around an infected cell and the per capita exit rate, as it is equal to the exit rate divided by the cluster size, k. Hence the total exit rate from a cluster of size k is $\mu_k = k\frac{\mu}{k} = \mu$. This model allows for bigger clusters to increase in size more rapidly as they will lose fewer cells

per capita. In this case the formation of a cluster of size k has the probability

$$P_k = P_0 \frac{\prod_{i=0}^{k-1} \lambda_i}{\prod_{i=1}^{k} \mu_i} = P_0 \frac{\lambda^k}{\mu^k} = (1-\theta)\theta^k, \tag{1.5}$$

where $\theta = \frac{\lambda}{\mu}$.

1.2.3 Model fitting and selection

The likelihood of the parameters λ and μ in a certain model, given the experimental data, is measured as the total probability of the observed data, given the model with parameters λ and μ. Therefore the likelihood function using the cluster size frequency data can be defined as:

$$L(\lambda, \mu | data) = P(data | \lambda, \mu) = \prod_{k=0}^{k_{max}} P_k^{x(k)} \tag{1.6}$$

where P_k is the probability of seeing a cluster of size k, given the parameters λ and μ (Eqns. 1.3-1.5), $x(k)$ is the number of clusters of size k, and k_{max} is the maximal cluster size in the data. By maximizing the likelihood with respect to the parameters, best-fit parameters can be found. It should be noted, however, that since we are assuming a steady state distribution of cluster sizes we are unable to estimate individual rate parameters such as λ and μ but only their ratio.

The model fitting, parameter estimation, and the statistical comparison of the models were done using R software (version 2.15.1). Confidence intervals for the parameters of the models were computed using the confint tool in R, which uses the likelihood profile method [18]. The relative goodness-of-fit of the three models were calculated using the Akaike information criterion (AIC), which evaluates each model by its level of variance from the data and the number of its parameters [15, 19]. The AIC score is minimal for the best tested model, which is then used to compute the AIC differences of other models relative to the best model. An AIC score difference of less than 2, between 4 and 7 and above 10 indicates substantial, significantly less

and essentially no evidence, respectively, for a model [19]. In this study, we calculate Akaike weights for every model using standard techniques [19], which are interpreted as the relative weight of a given model among all tested models. A low weight value for a particular model (e.g., < 0.01) indicates that this model is inferior at describing experimental data compared with other tested models.

1.3 Results

When we examined the frequency distribution of CD8+ T cells around infected hepatocytes, we were struck by the large variability in the number of CD8+ T cells in clusters (Figure 1.2). We compared this experimental distribution with three mathematical models of cluster formation (see Section 1.2.2). If clusters formed as a result of random interactions between T cells and an infected hepatocyte we would expect T cells to enter clusters at a constant rate and leave clusters at a rate proportional to the number of T cells in the cluster. Steady-state distribution of the number of CD8+ T cells surrounding a given parasite in this case corresponds to a Poisson distribution (Figure 1.2, red line) that does not fit the data well. In contrast two other models in which T cells were either preferentially attracted to larger clusters (density-dependent recruitment; Figure 1.2, green line) or preferentially retained in larger clusters (density-independent exit; Figure 1.2, gray line) fitted the data significantly better, suggesting that clusters may form by a nonrandom directed process.

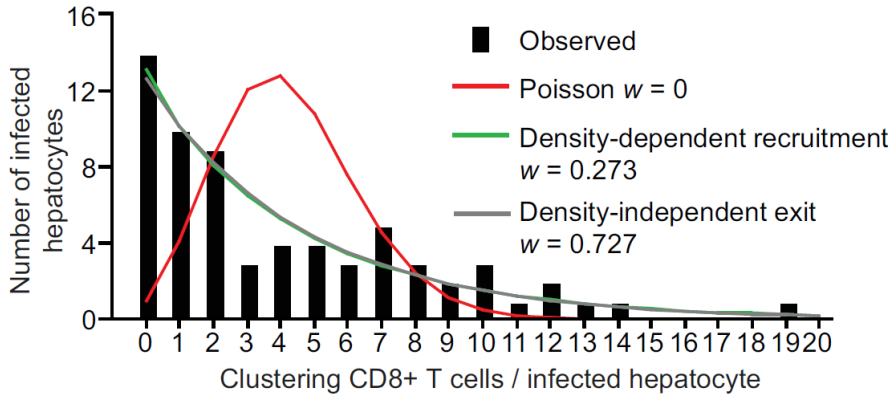

Figure 1.2: The endogenous CD8 T cell cluster size distribution rejects the T-cell-density-independent, random model of clustering. "w" stands for AIC weights, which give the likelihood of a model relative to other models ($w < 0.01$ suggests an unlikely model).

When PyTCR cells were transferred alone to infected mice, they strikingly recapitulated the phenotype seen in the endogenous response, with large clusters of up to 25 antigen-specific cells seen around some infected hepatocytes (Figure 1.3 A, i). Comparison of the frequency distribution of PyTCR cells around infected hepatocytes with the mathematical models (Figure 1.2) suggested that clusters were likely formed by the density-dependent recruitment of T cells rather than by chance or by density-independent exit of T cells (Figure 1.3 A, i). In contrast, when irrelevant OT-I cells were transferred to infected mice, 80 of 92 infected hepatocytes (87%) were not surrounded by any OT-I cells, whereas no infected hepatocyte was surrounded by more than 2 OT-I cells and. Moreover, the frequency distribution of OT-I cells around infected hepatocytes is consistent with the model in which clusters are formed by chance (Figure 1.3 A, ii). Together these data show that antigen-specific CD8+ T cells are necessary and sufficient for the formation of CD8+ T-cell clusters. Since we used the steady state solutions of our mathematical models to fit the data we were only able to estimate the ratio of per capita entry and recruitment rates to the per capita exit rates. The per capita entry rate to exit rate ratios were approximately 0.2

consistently for PyTCR and OT-1 cells whether they were transferred separately or together (Table 1.1).

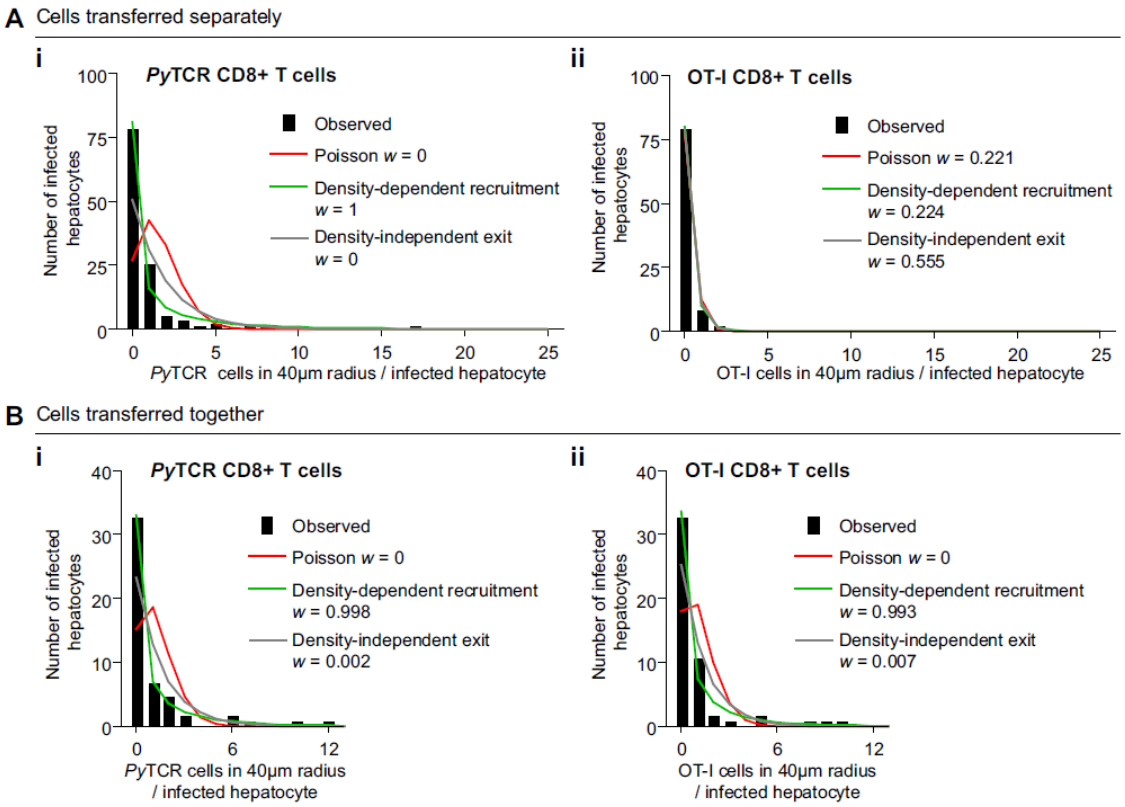

Figure 1.3: Modeling of clustering formation by specific and nonspecific T cells. (A) Frequency distribution of (i) OT-1 CD8+ T cells and (ii) PyTCR CD8+ T cells around infected hepatocytes when transferred to different animals compared with the distribution that would be expected for three different models for T-cell clustering: random entry and exit (red line), density-independent exit (gray line), and density-dependent recruitment (green line). (B) Frequency distribution of (i) OT-1 CD8+ T cells and (ii) PyTCR CD8+ T cells around infected hepatocytes when transferred to the same animal compared with the distribution that would be expected for three different models for T-cell clustering: random entry and exit (red line), density-independent exit (gray line), and density-dependent recruitment (green line). Akaike weights (w) for every model fitted to the experimental data are shown on individual panels.

Table 1.1: Parameter estimates (and their 95% CIs) for the three models (random entry/exit, density-dependent (DD) recruitment, density-independent (DI) exit) fitted to each dataset. Here θ is the ratio of the entrance to exit rate, and θ_1 denotes the density-dependent additional entrance rate, λ_1, divided by the exit rate, μ. The prediction of the model with DD recruitment was normalized assuming that $k_{max} = 30$.The lowest AIC score indicates the best fitting model for each dataset. AIC values were used to calculate Akaike weights shown in Figure 1.3. Dash (-), not applicable.

Condition	Model	θ	θ_1	AIC
Endogenous	Random entry/exit	4.2 (3.75–4.74)	—	461
	DD recruitment	0.7 (0.47–1.1)	0.8 (0.74–0.94)	337
	DI exit	0.8 (0.77–0.87)	—	338
OT1 cells alone	Random entry/exit	0.16 (0.10–0.26)	—	91
	DD recruitment	0.12 (0.06–0.22)	0.24 (0.01–0.77)	90
	DI exit	0.14 (0.083–0.21)	—	89
*Py*TCR cells alone	Random entry/exit	1.6 (1.4–1.8)	—	772
	DD recruitment	0.20 (0.14–0.28)	0.88 (0.81–0.94)	391
	DI exit	0.61 (0.56–0.66)	—	448
OT1 cells mixed	Random entry/exit	1.1 (0.80–1.36)	—	208
	DD recruitment	0.22 (0.12–0.38)	0.79 (0.62–0.92)	140
	DI exit	0.51 (0.42–0.61)	—	150
*Py*TCR cells mixed	Random entry/exit	1.2 (0.95–1.6)	—	231
	DD recruitment	0.21 (0.12–0.37)	0.83 (0.68–0.94)	149
	DI exit	0.55 (0.46–0.64)	—	162

Measurement of PyTCR entry and exit rates into 32 clusters between 4-8 hours PTCT supports the hypothesis that larger clusters have higher entry rates, but per capita exit rates are not influenced by the cluster size. Entry and per capita exit rates are plotted as a function of total T cells seen in each cluster in Figure 1.4.

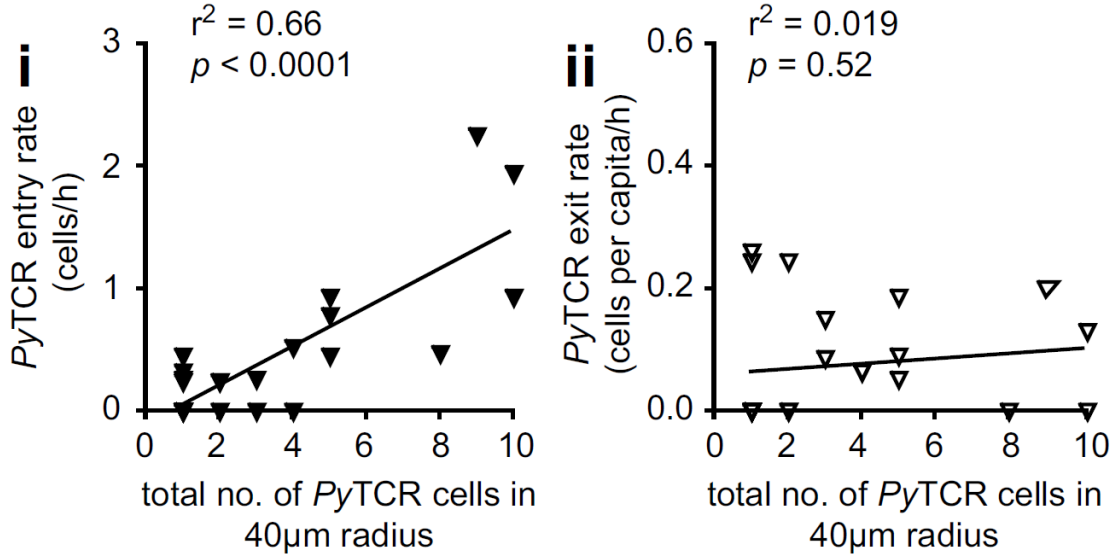

Figure 1.4: PKH-26 labeled effector PyTCR cells (1×10^7) were transferred to mice 20 h after infection with 3×10^5 PyGFP and imaged 48 h later; data are pooled from 32 movies in four independent experiments. Correlation of (i) entry rate and (ii) per capita exit rate of PyTCR with the total number of PyTCR cells around each infected hepatocyte. Infected hepatocytes not associated with any T cell were excluded from the analysis.

1.4 Discussion

This work presents the analysis of T cell cluster formation, which was frequently observed around *Plasmodium*-infected hepatocytes in mice. Our experimental data, supported by our mathematical modeling analysis, enables us to propose a model of how these clusters form. In this model an antigen-specific CD8+ T cell interacting with an infected hepatocyte alters the local microenvironment surrounding the site of infection, probably via the local secretion of cytokines and chemokines, resulting in the recruitment of additional effector CD8+ T cells of diverse specificities to form a cluster. This model is consistent with our finding that once an antigen-specific CD8+ T cell found an infected hepatocyte, both antigen-specific and nonspecific activated T cells are attracted to the cluster. Importantly, clustering seems to

be a physiological phenomenon and not an artefact of the use of TCR transgenic cells. We were able to observe clusters of endogenous T cells after immunization with RAS, which is in agreement with a previous postmortem histology study [17]. These clusters would be expected to contain both *Plasmodium*-specific cells and potentially other effector and memory CD8+ T cells from concurrent or previous immune responses. Cluster formation during *Plasmodium* infection may, however, require the presence of large numbers of pre-existing specific T cells [27]. Such a large number may only be induced by vaccination with subunit vaccines or immunogens such as RAS, which have been shown to provide depots of persisting antigen that may help induce and maintain CD8+ T-cell responses of broad specificity [9]. The relative roles of antigen and microenvironment in altering effector T-cell behavior have been examined in a variety of infections and tumor models. Similarly to our findings with *Plasmodium*, nonspecific CD8+ T cells were recruited to tumors only in the presence of tumor-specific CD8+ T cells [3]. However, in this model the recruited nonspecific cells had very distinct motility characteristics from the specific cells [22]. Similarly, in infection models of *Toxoplasma*, *Leishmania*, *Listeria*, and *Mycobacterium* [7, 10, 12, 34], non-specific T cells have been observed infiltrating sites of infection and in some cases showing altered motility relative to cells outside the foci of infection [10, 34]. In our system the T-cell behavior we observed is indistinguishable between specific and non-specific cells at the site of infection. One reason for this may be kinetic: we are looking at T-cell behavior in an acute infection very shortly after infection, whereas most imaging studies look at more persistent challenges days to weeks after infection or cell transfer. The long-term behavior of specific and non-specific T cells is a matter of interest that merits further studies. A number of key questions remain: we do not know how the first antigen-specific CD8+ T cell finds the infected hepatocyte, or whether this is a random or directed process. A recent study has reported that *Toxoplasma*-specific effector CD8+ T cells adopt a generalized Lévy random walk, which optimizes their ability to find rare infected cells in the brain [16]. It will be interesting to determine whether a

similar process is happening with *Plasmodium*-specific cells in the liver. Moreover, we do not know which GPCRs are required for further T-cell recruitment and cluster formation; interestingly, it has been demonstrated that both specific and non-specific T cells are recruited to the site of influenza virus infection in the lungs via CCR5 [21], whereas in the lymph node, CXCL9-CXCR3 interactions have been shown to be critical for the clustering of antigen-specific effector CD8+ T cells around DCs in viral infections [20, 32]. Although a single CD8+ T cell is probably sufficient to kill a parasite-infected cell, clustering likely increases the probability of parasite killing. Clustering of CD8+ T cells may aid the elimination of pathogens in a number of ways. CD8+ T cells are known to vary greatly in their ability to degranulate and produce effector cytokines in response to antigen [24]. Therefore, CD8+ T cells with limited effector capacity may initially contact infected cells. These cells may, however, be able to act as sentinels and recruit other CD8+ T cells to the infected cell. A requirement for CD8+ T-cell clustering for efficient parasite killing is compatible with previous work that showed that a large number of *Plasmodium*-specific CD8+ T cells are needed to ensure sterile protection against parasite challenge [25, 29]. In conclusion, this intravital dynamic imaging of *Plasmodium* elimination reveals that parasite destruction frequently involves the recruitment of multiple CD8+ T cells to a single infected cell and suggests that it can be achieved by distinct mechanisms.

Chapter 2

Classification of T cell movement tracks allows for prediction of cell function

This chapter contains a previous publication by the author.
R. K. Kelemen, G. F. He, H. L. Woo, T. Lane, C. Rempe, J. Wang, I. A. Cockburn, R. Amino, V. V. Ganusov and M. W. Berry. Classification of T cell movement tracks allows for prediction of cell function. Int. J. Computational Biology and Drug Design. (in press)

2.1 Introduction

CD8+ T cells are part of the immune system and are responsible for seeking and killing damaged, cancerous or infected cells in mammals. In infectious diseases such as malaria, which takes about one million lives annually and affects 300 times more, CD8+ T cells have been proven to play crucial roles in the host's ability to clear the parasite during an adaptive immune response [36]. It has been long known that CD8+ T cells are essential for immunity against malaria, and recently [29] has shown that

reaching a certain large, but definable threshold of CD8+ T cells specific for malaria antigen is sufficient for clearing the infection in 98% of the cases in mice [28, 29].

Despite their importance in the immune response against many infections, the mechanisms by which parasite-specific CD8+ T cells survey the site of infection and what factors influence T cell behavior and movement patterns remains poorly understood. Current hypotheses about CD8+ T cell movement are based on in vitro or in vivo immunological experiments, such as the detection of increasing numbers of T cells as the concentration of chemoattractant molecules increases in a certain area. The results of such experiments suggest that T cells are recruited to the site of infection by a concentration gradient of chemokines, which are expressed by the infected cell and other immune cells. Challenges remain in experimentally detecting the radius of such a chemokine concentration gradient and the effect they have on T cell movement.

With the latest advancements in visualization techniques, high resolution videos produced by spinning disk confocal and two-photon microscopy give biologists and immunologists a whole new perspective on T cells and their movement in the body [13]. Numerous studies have employed in vivo or in vitro visualization techniques using fluorescently labeled immune cells, and analyzed their behavior and movement trajectories to either visually or statistically test for phenomena such as Levy flights [16].

While qualitative analyses are important tools in gaining understanding of the immune response to infections, quantitative approaches to the same questions can give us more unbiased results and in many cases highlight underlying patterns otherwise undetectable. Here, we introduce the use of data mining and spatial and hierarchical clustering in analyzing T cell movement coordinate data obtained from in vivo imaging of mouse livers. We test several hypotheses regarding T cell movement, such as that T cell movement pattern depends on the T cell's specificity for the infection, the T cell's distance from the infected cell, or the presence of an infection in the body. We also address the question of whether the tissue topology is the key determining factor

in T cell movement, and we try to reconstruct the tissue landscape at each imaged site, based on the movement coordinate data.

We find that CD8+ T cells do exhibit different movement patterns based on their specificity for the infecting parasite and that our support vector machine trained on two types of T cell movement data can distinguish between the two cell types with 72% accuracy. Hierarchical clustering reveals several subtypes of T cell movement even in a homogeneous set of CD8+ T cells specific for malaria antigen. We also observe zones of highly traveled areas in videos of the liver, which might be important information for overlaying T cell movement and tissue topology. Overall, we demonstrate several approaches and quantitative, computational tools that are applicable not only to the movement data of cells, but of animals, or any other moving object.

2.2 Implementation

We have analyzed the T cell movement coordinate datasets of several mouse experiments. Experiments were done by Ian Cockburn and Rogerio Amino at Johns Hopkins University (Baltimore, USA) and Pasteur Institute (Paris, France), respectively. These experiments involved in vivo imaging of fluorescently labeled *Plasmodium yoelii* parasites, the causative agents of malaria, and CD8+ T cells of various epitope specificities (Figure 2.1). All the imaging was done by confocal spinning disc microscopy in anaesthetized mice over a time period of 30-60 minutes. Two- or three- dimensional position coordinates of labeled cells nearby a malaria-infected cell and the parasite were acquired at time steps of 30 seconds to four minutes, as described by [8]. We focused on three datasets that we will refer to as Datasets 1, 2, and 3, as defined below.

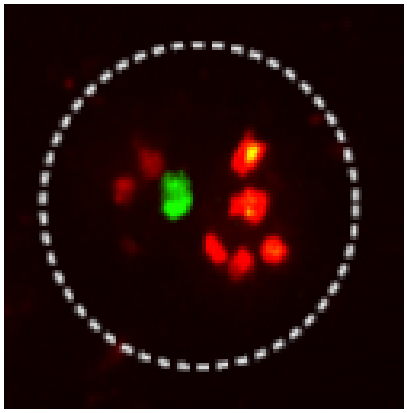

Figure 2.1: Image of malaria-specific PyTCR cells labeled with red fluorescent membrane dye (PKH-26) congregating around a green fluorescent protein expressing malaria parasite inside an infected liver cell. Dashed circle denotes a hypothetical outline of an infected hepatocyte (radius 40 μm). Image courtesy of Ian Cockburn.

2.2.1 Dataset 1: PyTCR and OT-1 cells in infected mice

The first dataset we analyzed consisted of movement coordinates of two different types of CD8+ T cells in malaria infected mice: one specific for the *Plasmodium yoelii* (Py) epitope SYVPSAEQI, and the other specific for the ovalbumin epitope, which is unrelated to malaria. We will refer to these two cell types as PyTCR (*Plasmodium yoelii* T cell receptor transgenic) and OT-1 (ovalbumin T cell receptor transgenic). The position coordinates of these two distinctively labeled cell types around a single infected liver cell were acquired every two minutes for a total of 38 minutes. This experimental setup is suitable for testing whether CD8+ T cells specific and CD8+ T cells not specific for malaria exhibit different movement patterns when injected together into malaria-infected mice.

2.2.2 Dataset 2: PyTCR cells in infected mice

The second dataset contained the movement coordinates of only PyTCR cells with malaria infection in the liver imaged at time intervals of four minutes for a total of 50 minutes. This dataset is useful for clustering a seemingly homogeneous set of T cells based on their movement attributes.

2.2.3 Dataset 3: Parasite specific T cells in infected and uninfected mice

The third dataset was a result of mouse experiments where OT-1 cells were injected into mice, then the mouse was immunized with irradiated malaria parasites that expressed a mutated protein containing the chicken ovalbumin epitope. Six days later the mouse either was challenged with ovalbumin-expressing malaria parasites or remained unchallenged right before the imaging of the liver. This dataset is useful for classifying T cell movement patterns based on whether there is or is not an infection in the body.

2.2.4 Movement Attributes

To characterize T cell movement for classification as discussed in Sections 2.5 and 2.6 (support vector machine classifier and hierarchical clustering) we calculated several attributes for each cell in Dataset 1 (PyTCR and OT-1 cells) based on the two- or three-dimensional coordinate points. The attributes are total path, maximum, minimum, and average movement, maximum, minimum, average and sum of all angles towards the parasite, total displacement length, total displacement vector, and the *meandering index*, which is the ratio of the total path length to the displacement (Figure 2.2).

To include angles related to T cell movement in our analysis, we used the following calculations for each T cell. The procedures to calculate turning angles of moving T cells are as follows.

Find the average coordinate of the parasite.

$$P =< P_x, P_y, P_z > \tag{2.1}$$

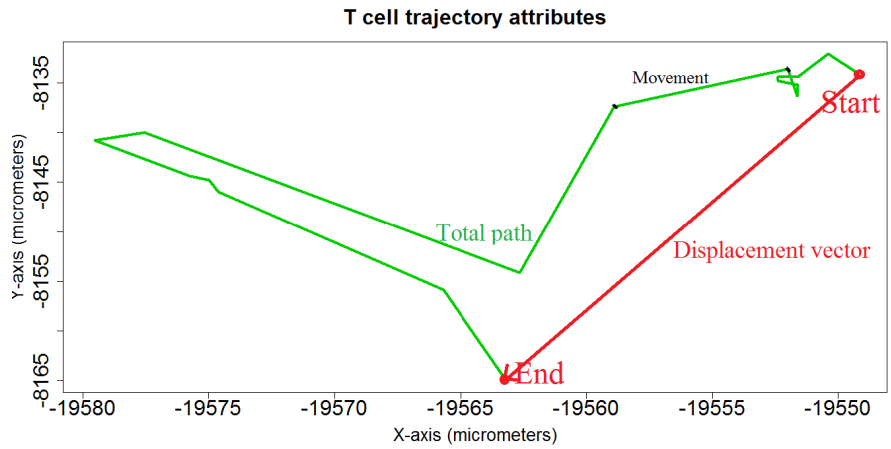

Figure 2.2: Movement attributes of T cells. The distance between two consecutive positions of a T cell was considered a movement, and the sum of all movements for a T cell was its total path length. The distance and the vector between the start and end points of a T cell are called its displacement and displacement vector.

Calculate parasite-T cell distance vector, PT_n, at each time point n.

$$PT_n =< T_{x,n} - P_x, T_{y,n} - P_y, T_{z,n} - P_z > \tag{2.2}$$

Calculate the T cell's movement vector, TT_n, between each time point n.

$$TT_n =< T_{x,n} - T_{x,n+1}, T_{y,n} - T_{y,n+1}, T_{z,n} - T_{z,n+1} > \tag{2.3}$$

Calculate the angle between the parasite-T cell distance vector and T cell's movement vector at time point n.

$$\theta_{PT-TT,n} = \arccos(PT_n \cdot TT_n / ||PT_n|| ||TT_n||) \tag{2.4}$$

where $||x||$ denotes Euclidean distance. Calculate the angle between the T cell's movement vector at time point n and its movement vector at the subsequent time point, $n+1$, as

$$\theta_{TT-TT,n} = \arccos(TT_n \cdot TT_{n+1} / ||TT_n|| ||TT_{n+1}||) \tag{2.5}$$

so that the scalar distance from parasite can be represented by

$$PT_n = (T_{x,n} - P_x)^2 + (T_{y,n} - P_y)^2 + (T_{z,n} - P_z)^2. \tag{2.6}$$

2.2.5 Spatial Analysis

To reconstruct a sense of the tissue topology based on the frequency of T cell presence at various sites in the imaged area of the liver, we used a clustering algorithm to create *zones* based on T cell positions with proximity to each other for Datasets 1 (PyTCR and OT-1 cells) and 2 (PyTCR cells). An implementation of the DBSCAN algorithm in Python was utilized for clustering T cell locations across all time points in each video dataset [11]. No differentiation was made between T cell locations based on whether the cell belonged to a class of parasite specificity, and the T cell locations were not grouped by the cell that they belonged to. The original testing and visualization of various clustering algorithms was performed in the ELKI platform [1]. The ELKI platform is a powerful tool for assessing clustering methods most likely to yield biologically relevant results, particularly through the visualization and availability of a multitude of clustering techniques.

The parameters used for all DBSCAN clustering experiments were an epsilon of 10 and a minimum points of 2. Minimum points is a parameter that specifies how many points are needed to meet the requirements of clustering with one another in order to form a viable cluster. The epsilon is a value regarding the maximum distance to look from a point for nearby candidate points to join a cluster. These parameters were selected based on visual assessment of T cell clusters in order to increase their biological relevance in the liver tissue.

2.2.6 Classifier: Support Vector Machine

We implemented the Python interface of LIBSVM [6] with Python 2.7.1, making use of the LIBSVM cross-validation function in model training. The above described distance, angle, and zone attributes were used as input into the SVM to train a linear or radial basis function (RBF) model using 3 to 10-fold cross-validation and the no shrinking heuristics option. This was applied to Datasets 1 (PyTCR and OT-1 cells) and 3 (infected and uninfected).

2.2.7 Hierarchical clustering

After calculating movement attributes including average movement length, total path length, displacement, and angle between movements for each cell in our collection of videos, we provided this information to a hierarchical clustering algorithm for Datasets 2 (PyTCR cells) and 3 (infected and uninfected). Hierarchical clustering and heatmap generation was done using JMP Genomics 6.0 (SAS Institute Inc. 2012) with Wards minimum variance method. All attributes were used for clustering. We additionally used the R-based software packages (*ggplot* and *plyr*) to visualize a series of T cell trajectories on their own, or in relation to the parasite for Dataset 2 (PyTCR).

2.3 Results and Discussion

2.3.1 T cell movement with respect to distance from parasite

Most T cells within Dataset 1 (PyTCR and OT-1 cells) tend to move towards the parasite regardless of cell type. The direction of a trajectory was considered as moving towards the parasite if the angle given by Equation 2.4 was acute. 82% of both OT-1 and PyTCR cells were found to be traveling towards the parasite for more than half the duration of their monitored movement. The distribution of angles between the two cell types were not significantly different based on nonparametric Kruskall-Wallis

rank sums test at an alpha level of 0.05. This finding using solely angle data initially suggested that OT-1 and PyTCR behavior are similar in that they will all migrate towards the infection site. However, there may be a bias in our data since the parasite was located in the center of the frame, so cells further from the parasite are not included in the small (200x200x50 μm) field of view. Additionally, this dataset only contained cells with at least six timepoints, which may have excluded cells moving quickly in and out of the frame.

A binomial test is applicable if we consider the orientation data as a Bernoulli trial where "moving towards the parasite" is the success. This statistical method gives the likelihood of observing the probabilities reported in Figure 2.3 when the expected probability is 50%. We found the likelihood of observing these probabilities is less than 2.5% for both OT-1 and PyTCR cells and both distance ranges greater and less than 40 microns. In other words, it is unlikely we are observing higher probabilities of moving towards the parasite by all cells by random chance. Therefore, the statistics support the possibility of a general bias in T cell behavior to move towards the parasite.

We found that (parasite-specific) PyTCR cells move towards the parasite 15% more often than non-specific OT-1 cells at distances greater than 40 μm from the parasite (Figure 2.3, $P = 0.027$, Kruskall-Wallis rank sums). However, no difference in migration towards the infected cell was noted when T cells are within 40 μm of the parasite; the latter length approximates the maximum size of an infected hepatocyte in mice (Figure 2.3, $P = 0.579$, Student's t-test). Interestingly, irrespective of the distance, both parasite-specific and non-specific cells had increased preferance to move to the infected cell (more sharp turning angles than expected by chance, $P < 0.025$, binomial test). In conclusion, these results provide the basis of classifying T cells using a combination of their distance and orientation to the parasite as attributes.

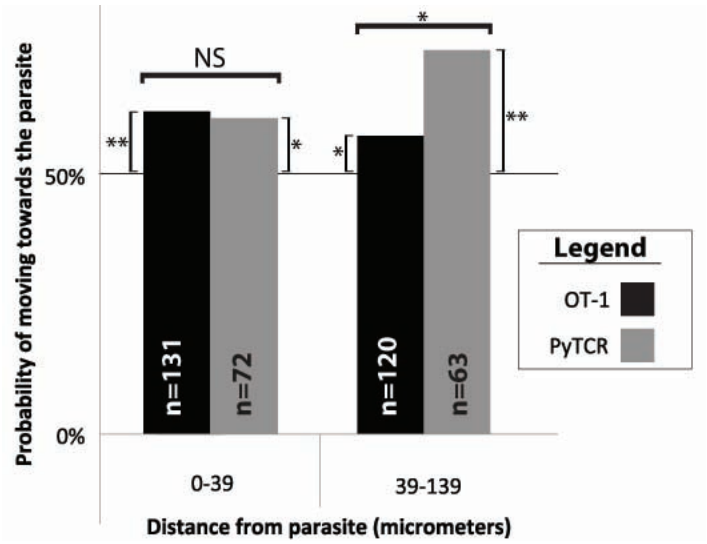

Figure 2.3: Direction trends of PyTCR (malaria-specific) and OT-1 (non-specific for malaria) cells (Dataset 1) based on distance from infection site. The direction of T cell movement at each time point was grouped by distance from parasite: up to 39 μm (near a parasite-infected cell, see Figure 2.1) or farther than 40 μm (away from a parasite-infected cell). The probability of moving towards the parasite is calculated by dividing the number of angles less than 90^0 relative to the direction of the infection site by total number of data points at that distance interval. Both cell types head towards the site of infection at a probability greater than 50%, suggesting a general tendency for T cells to move towards the parasite. However, at distances greater than 40 μm, malaria-specific (PyTCR) cells will travel towards the parasite more often than nonspecific (OT-1) cells. The number of data points analyzed at each bin is indicated by n. Distance bins with less than 3 data points are not included. NS = not significant. $*P < 0.05$ and $**P < 0.002$ (unpaired two-tailed t-test, Kruskall-Wallis rank sums, binomial test).

2.3.2 Spatial Analysis

Our spatial clustering analysis showed areas of highly traveled sites in the imaged tissue of Datasets 1 (PyTCR and OT-1 cells) and 2 (PyTCR cells), while most of the surveyed area was untouched by T cells. Interestingly, some features of the underlying tissue topology, like *highways* of immune cell traffic can be extrapolated from these figures of points traveled by T cells (Figure 2.4). Further analysis of high density T cell videos could lead to more highly resolved pictures of the tissue

that could potentially predict complete highways of immune cell movement and the chemoattractant concentration gradient.

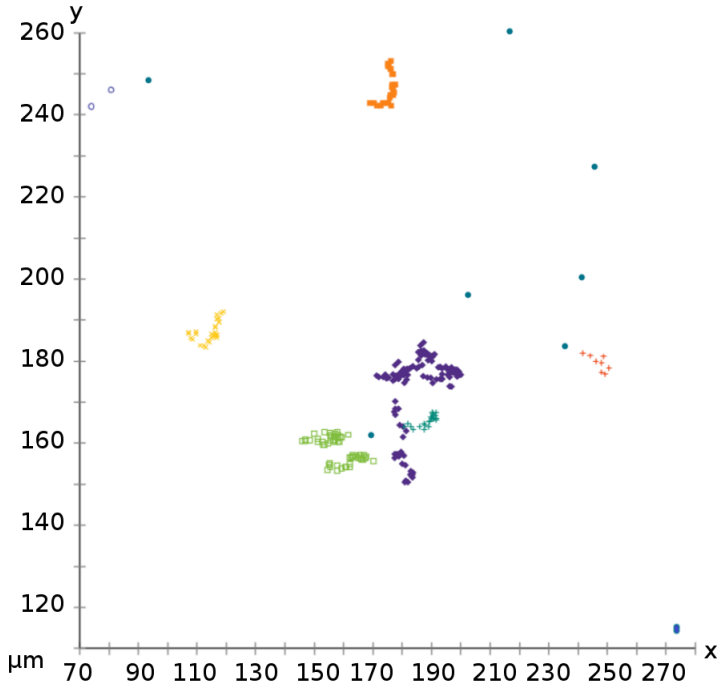

Figure 2.4: Colors represent density based clusters of T cell positions at all timepoints from a parasite-infected mouse liver (Dataset 2). Blue circles are cell positions that did not cluster. These colors correspond to categories of the *zone* attribute used in the SVM, which highlight potential areas of intense traffic denoted by many dots and a single color that may include one or more individual cells. While this dataset was too sparse to visualize complete *highways*, this method could be used to predict junctions of sinusoids or the presence of recruiting cells in the liver with a larger dataset.

2.3.3 Classifier: Support Vector Machine

We used a support vector machine to assess whether attributes of T cell movement can be used to differentiate PyTCR from OT-1 T cells in parasite-infected mice (Dataset 1) or parasite-specific T cells in infected or noninfected mice (Dataset 3). A support vector machine functions as a classifier by mapping the vector data to a higher dimensional space using a kernel function and then by finding a hyperplane that optimally separates the data by maximizing the margin between the plane and

each class [30]. Support vector machines are a common means of supervised (binary) classification that have recently been used in a study by Goodson et al. in 2011 to analyze movement patterns of mouse sperm. Goodson et al. used a support vector machine to classify image data of sperm cells moving over time into different movement classes [14]. To our knowledge, we are the first to apply such a method to the understanding and classification of T cell movement.

We started with a baseline assessment of classification by feeding two basic attributes into the support vector machine: total path length traversed by a cell and the distance traveled for the first jump. Using a linear kernel and 3-fold cross validation (using the -v cross validation option of LIBSVM) on our dataset of 29 cells (18 malaria specific and 11 malaria non-specific) our support vector machine (SVM) model gave 64% accuracy. Next, we used the same SVM parameters but added several more distance attributes, then several angle based attributes. All distance attributes (path length, average jump, mininum jump, maximum jump, net x displacement vector, net y displacement vector, net z displacement vector, total displacement) gave an accuracy of 53%, total path length and all angle attributes (mean angle, minimum angle, maximum angle, sum of angles, mean angle to parasite, minimum angle to parasite, maximum angle to parasite, and sum of angles to parasite) gave an accuracy of 64%, and total path length and zoning information gave an accuracy of 57%.

Since the dataset is from a single imaging space, cells move in and out of the frame over the course of the experiment. In an attempt to normalize cells that remain in frame with cells that move in and out of frame, we divided cell data into overlapping *chunks* of six timepoints, the minimum number of time points required to keep a cell in the dataset. We generated the same distance and angle attributes from these chunks of cell time and obtained a 66.9% accuracy with 3-fold cross validation. However, since the total number of data points increased when separated into these chunks, we discovered that this level of cross validation was underfitting the model. At 10-fold cross validation with all distance and angle attributes for time chunks, we observed a

70.6% accuracy. Finally, when we added zoning information into this model, the 10-fold cross validation accuracy increased to 72%. Nevertheless, when we investigated single instances of training on 9/10ths of the data and testing on 1/10th of the data, accuracies varied from 72% to 96% (see Table 2.2). This could be due to different data distributions in the training and testing folds. Since this dataset was quite small, we did not try subsampling within the data.

For the dataset of cells in malaria-infected mice versus uninfected mice, we generated the same set of distance and angle attributes, excluding those attributes using information about parasite location (mean angle to parasite, minimum angle to parasite, maximum angle to parasite, and sum of angles to parasite), on overlapping chunks of six time points for these cells. Cells with fewer than 6 timepoints were excluded. These attributes were used to train a radial basis function (RBF) SVM model, with which we observed an accuracy of 87.9%. Looking closer at our data, however, we discovered that the percent of data points labeled 'infected' was about 87%. To avoid biasing the classifier to the larger quantity of infected cells, we tried using the built-in weight function of LIBSVM, but this also gave an accuracy of 87%. As a further check, we pseudo-randomly sampled the infected cells for a subset equal to the number of uninfected cells in the dataset. This subset gave a cross-validation accuracy of 69.8% under the same conditions. Since the LIBSVM cross validation option is intended for training, we next made our own split of the chunk data with equal numbers of infected and uninfected cells into 10-folds. We trained with 9/10ths of this data and tested with 1/10th of this data in single instances, and were surprised by the high accuracies this gave (Table 2.3). The large discrepancy between the 69.8% of LIBSVM's cross validation option and the accuracies around 99% that we observed with single instance train and test sets may be due to an unrepresentative random sampling of infected cell datapoints. Although the number of infected and uninfected cells is the same with this method, the number of timepoints for each cell varies. Our dataset of 6 timepoint chunks has about twice as many datapoints for the infected

class as the uninfected class. Nevertheless, our confusion matrix and high accuracy show a good separation of the data with an RBF kernel function (see Table 2.3).

Table 2.1: Summary of support vector machine results. Accuracies vary depending on input attributes, folds used for cross validation, and whether whole cells or overlapping *chunks* of 6 timepoints were used. Cross-validation folds are given, meaning the dataset was divided into 3 or 10 equal parts (folds), trained on *n-1* folds, and tested on 1 fold. Accuracy denotes the percent of the testing dataset that was correctly classified by the support vector machine. All distance attributes include path length, average jump, mininum jump, maximum jump, net x displacement vector, net y displacement vector, net z displacement vector, and total displacement. All angle attributes include mean angle, minimum angle, maximum angle, sum of angles, mean angle to parasite, minimum angle to parasite, maximum angle to parasite, and sum of angles to parasite).

Cross Val.	**Data Type**	**Attributes**	**Accuracy**
3	whole cell	path length, 1st distance	64%
3	whole cell	all distance attributes	53%
3	whole cell	path length, all angle attributes	64%
3	whole cell	path length, zones	57%
3	6 timepoints	all distance, angle attributes	66.9%
10	6 timepoints	all distance, angle attributes	70.6%
10	6 timepoints	all distance, angle attributes,zones	72%

Our application of machine learning to the datasets where there are two distinct cell types (malaria-specific vs. nonspecific) or two different experimental conditions (infection or no infection) present yielded a confirmation of those differences in the movement patterns of cells. Despite the small size of the first dataset our support vector machine predicted the right cell type in at least 72% of the cases (Table 2.1). This result is encouraging for further training on parasite-specific and non-specific cell movement datasets to better outline and understand what differences it makes for a cell to have a receptor for the parasite that is causing an on-going infection. A related classification was based on two experimental conditions, namely having a malaria infection in the liver or not. The support vector machine trained on the malaria-specific T cell movement coordinate data in these two conditions was able to predict which cell movement takes place in an infected versus an uninfected mouse with a 70% accuracy. This result indicates that just like between malaria-specific and non-specific

cells during malaria infection, there are differences between the movement attributes of malaria-specific cells in infection and infection-free environments. Since we did run into variation and discrepancies between LIBSVM cross validation accuracy and single instance train and test accuracies, it would be beneficial to include boosting in future analyses of this dataset to reduce bias and hopefully obtain more consistent accuracies.

Table 2.2: Summary of support vector machine results: OT-1 and PyTCR data. The overall accuracy using 10-fold RBF and linear cross-validation was 67.7% and 72%, respectively. The number of correctly or incorrectly classified testing *chunks* of cell timepoints are given in the true and false categories for each cell type; the SVM appears to correctly classify at least 72% of the testing dataset. (**T**=true, **F**=false).

Model	**Accuracy**	**T-OT1**	**F-OT1**	**T-PyTCR**	**F-PyTCR**
RBF	64%	16	9	0	0
RBF	72%	18	7	0	0
Linear	80%	14	3	6	2
Linear	96%	20	1	4	0
Linear	80%	16	4	4	1
Linear	84%	16	4	5	0
Linear	72%	15	6	3	1
Linear	84%	20	4	1	0
Linear	92%	18	2	5	0
Linear	80%	15	5	5	0
Linear	88%	17	2	5	1
Linear	76%	14	4	5	2

Previous studies of the role of receptor-ligand interactions in T cell migration have shown that the T cells' antigen-specific receptor (TCR) is involved in regulating the motility of the cell based on the presence or absence of the specific parasite antigen on nearby endothelial cells [35]. The biological explanation of non-specific cells behaving just as different from infection-specific T cells as if there was no infection in the surveyed tissue is consistent with our findings.

Table 2.3: Summary of support vector machine results: infected (**inf**) versus uninfected (**uni**). The overall 10-fold RBF cross-validation accuracy was 69.8%. The number of correctly or incorrectly classified testing *chunks* of cell timepoints are given in the true and false categories for each cell type; the SVM appears to correctly classify at least 97% of the testing dataset in these iterations. (**T**=true, **F**=false).

Model	**Accuracy**	**T-uni**	**F-uni**	**T-inf**	**F-inf**
RBF	99.2%	132	2	238	1
RBF	100%	99	0	282	0
RBF	97.9%	112	5	225	2
RBF	99.7%	114	0	256	1
RBF	99.7%	115	0	281	1
RBF	100%	96	0	259	0
RBF	99.2%	114	0	271	3
RBF	99.5%	121	1	61	1
RBF	99.7%	134	1	221	0
RBF	99.4%	131	2	220	0

2.3.4 Hierarchical Clustering

Hierarchical clustering of the dataset containing the movement coordinates of only malaria-specific (PyTCR) cells resulted in seven very distinct clusters and numerous sub-clusters of cells based on their movement. Plotting the tracks of cells in different clusters showed that some cells tend to make small movements while others tend to move in very long jumps over the same time interval (Figure 2.5).

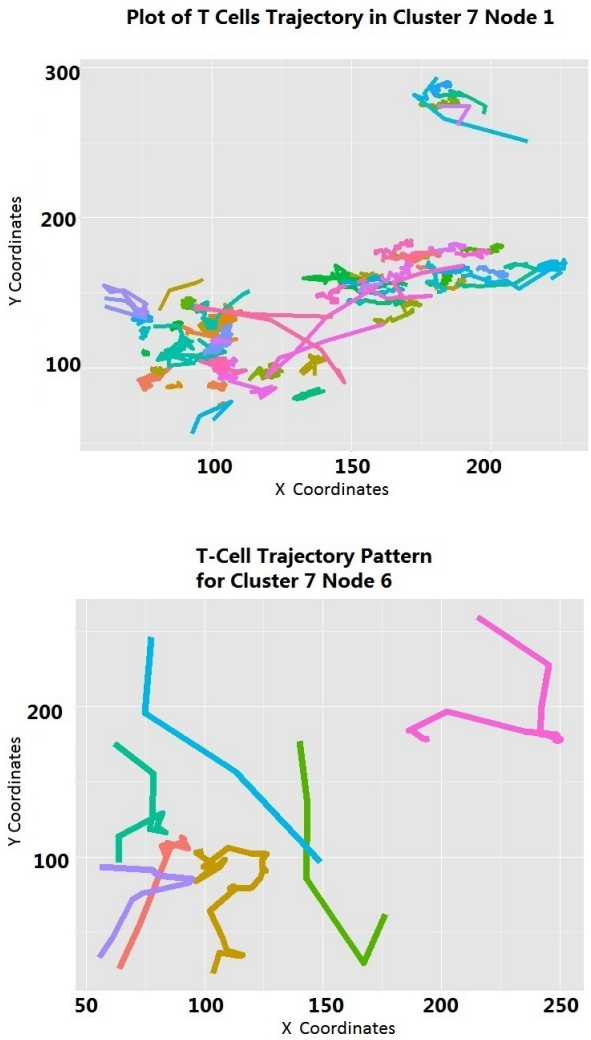

Figure 2.5: Each of these images contains malaria-specific CD8+ T cell paths that belong to different clusters based hierarchical clustering of different movement patterns (Section 3.4). We see that T cells tend to move either in small jumps and remain localized (top panel) or in large jumps and traverse large distances (bottom panel). Colors represent different cells.

Other clusters have cells with a mixture of long or short movements, but interestingly they have very small overall displacement. This means that in many cases the cells move a lot, but due to sharp angles between jumps they change their direction of movement drastically and end up near their starting point. This suggests that there are different movement types and perhaps search strategies even within the same set of cells, which have the same antigenic specificity.

We additionally did Ward's hierarchical clustering using all the attributes for malaria-specific T cells in the malaria-infected versus uninfected mice (Figure 2.6). We can see patterns in the attributes that reveal what some clusters were based on, but the clusters do not appear to correspond to infected or uninfected cells (as visualized by the the red and blue bars to the left of the heatmap). In Ward's hierarchical clustering, Euclidean distance for each attribute is used in an unsupervised manner to separate the clusters while the support vector machine uses a supervised approach to find an optimized hyperplane to separate the data. The difference between supervised and unsupervised learning may be the reason that the SVM appears to find differences where the clustering does not.

The technological advancements in visualization techniques in the fields of biology and immunology enable us to record the movement of cells in vivo. Beyond visual assessment, qualitative and small-scale statistical analysis, we can take advantage of the large and information-dense movement coordinate data of these videos to find underlying common cell movement patterns, which are understudied. In this paper, we have demonstrated the use of applying hierarchical and spatial clustering and machine learning algorithms to analyze cell track datasets.

2.4 Conclusions and Future Work

This study has tackled the novel interdisciplinary problem of classifying T cell movement track data using computational tools, and provides a detailed outline of our approaches and results. Future work in T cell movement classification will benefit

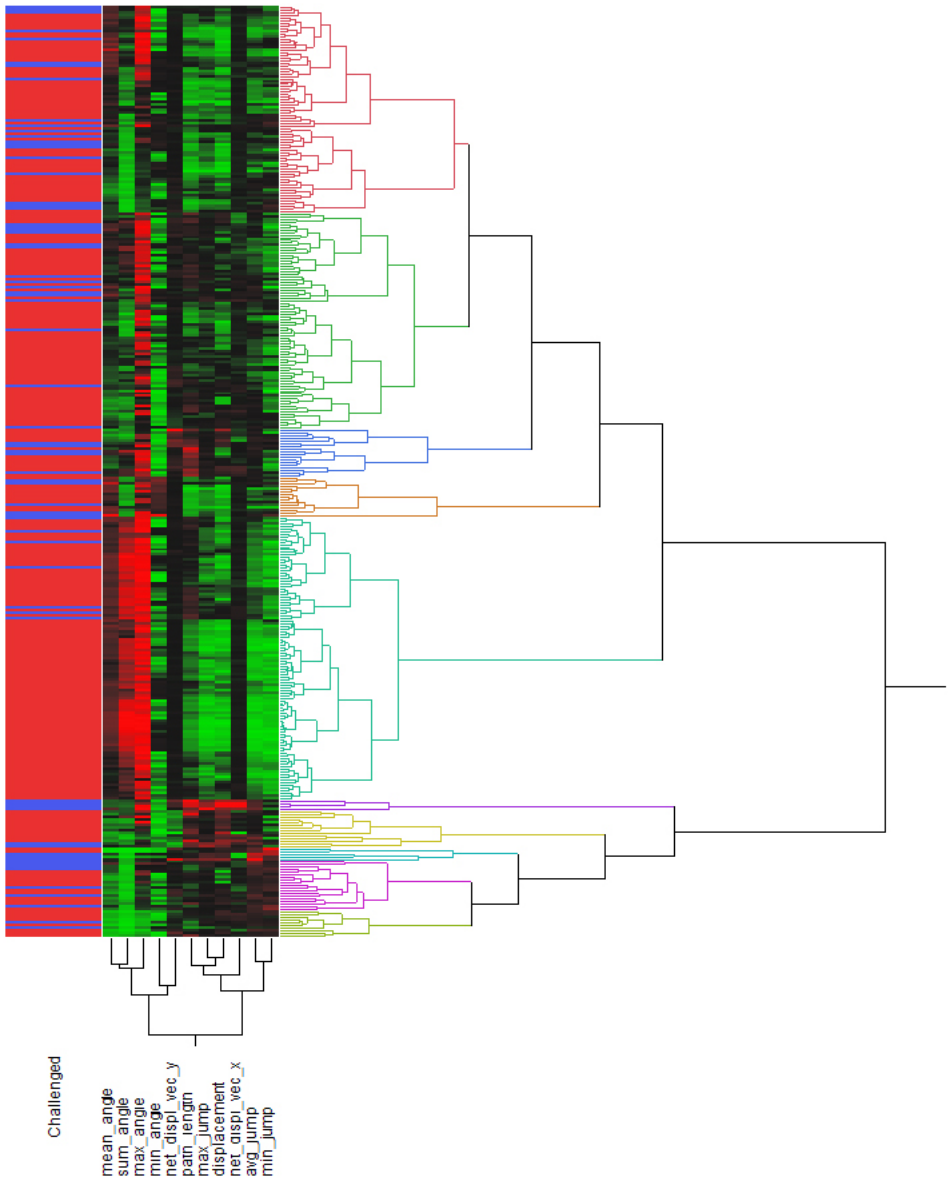

Figure 2.6: Hierarchical clustering of parasite-specific T cell movement patterns in an infected mouse (red bars) and an uninfected mouse (blue bars) based on mean angle, sum of angles, max. angle, min. angle, net y displacement, path length, max jump, total displacement, net x displacement, avg. jump, and min. jump. Clustering patterns are visible in the attribute similarities, but do not appear to correlate with whether a CD8+ T cell is from an infected or uninfected mouse.

from high quality images that are being generated more and more by experimentalists. With the availability of the underlying tissue topology information in videos of the liver, it is possible to map the movement patterns onto the landscape and test crucial hypotheses about the main determining factors of T cell movement in the tissue. Comparing T cell movements between different tissues, and also movements of different cell types using our methods contributes to our understanding of the immune system and opens up a new perspective on the classification of cell types.

Chapter 3

Exploring the possible effects of T-cell-extrinsic and T-cell-intrinsic factors in driving the clustering of T cells at sites of infection

This chapter represents a manuscript in progress by the author.
R. K. Kelemen, I. A. Cockburn, R. Amino and V. V. Ganusov. Discriminating between T-cell-extrinsic and T-cell-intrinsic mechanisms of CD8 T cell clustering during liver-stage malaria in mice. PLoS Computational Biology.

3.1 Introduction

In Chapter 1 we compared mathematical models that represented three alternative hypotheses about the driving forces of CD8 T cell clustering during liver-stage malaria in mice. We have shown that in case of malaria-specific T cells the model with T-cell-dependent recruitment to infected cells predicts the distribution of cluster sizes

6 hours post T cell transfer (PTCT) with best quality. In the case of non-malaria-specific T cells the models with T-cell-dependent recruitment into or retention in the clusters did not explain the observations significantly better than our null hypothesis of random T cell movement. Co-transferring OT-1 cells with PyTCR cells, however, enhanced their clustering behaviour and our density-dependent (DD) recruitment model fit this data the best. We also used our three basic models to characterize the clustering mechanisms of endogenous CD8 T cells during liver-stage malaria in previously immunized mice, and concluded that T-cell-density-dependent recruitment and retention explain the observed cluster size frequencies similarly and reject the null model. In this chapter we will address our assumption that cluster sizes are at equilibrium 6 hours PTCT using measured cluster size changes 4-8 hours PTCT as well as simulations using our DD recruitment model. We will also refine our analysis of the datasets mentioned above with further models and include an additional dataset that contains cluster sizes produced by pertussis toxin treated PyTCR cells. In particular, we will question the exclusivity of T-cell-dependent mechanisms in explaining the observed large clusters formed by certain T cell types and test models that rely on T-cell-independent causes to the clustering process. Using a mixed cluster modeling framework, where two cell types are distinguished by different entry and possibly recruitment rates, we are able to search for evidence for two subpopulations even in a T cell population with specificity for the same parasite antigen. This model also allows to estimate separately the contributions of malaria-specific and non-malaria-specific T cells in the recruitment to infected cells in the experiment of co-transfer.

3.2 Materials and methods

3.2.1 Data

This study utilized datasets generated by the experiments of Cockburn et al, which are described in more detail in [8], and are summarized in Figure 1.1. Our six datasets of observed cluster size distributions contain the counts of the following fluorescently labeled cell types: PyTCR cells alone, OT-1 cells alone, PyTCR cells when transferred together with OT-1 cells, OT-1 cells when transferred together with PyTCR cells, pertussis toxin treated PyTCR cells and endogenous CD8 T cells. The first five datasets are the products of imaging mouse livers 6 hours PTCT, while the endogenous T cell clusters were stained and imaged 24 hours post infection. The third and fourth datasets are counts of two different cell types at each observed infected hepatocyte, and can be combined to obtain the total size of each mixed cluster. The T cell populations of the experiments we analyzed, PyTCR, OT-1 and PTx cells are all activated T cells that likely carry similar receptors for adhesion molecules expressed on endothelial cells, which help them immobilize in the blood flow and enter the inflamed tissue. PyTCR and OT-1 cells express G-protein-coupled receptors (GPCRs) that allow them to sense signals from chemoattractants, such as chemokines. On PTx cells GPCRs have been immobilized in the inactivate conformation due to treatment with pertussis toxin, therefore most of the PyTCR cells in this population were insensitive to chemoattractants. CD8 T cells also express T-cell receptors, which they use to recognize their target antigen that is displayed on cells infected by the parasite they originate from, and TCRs possibly also influence the duration of stay at infected cells. PyTCR and PTx cells possess TCRs specific to one short peptide sequence (epitope) in the *P. yoelii* proteome, while the OT-1 cell population expresses a TCR specific to an epitope absent in the *P. yoelii* proteome. Our analysis also included endogenous CD8 T cells 10 days after radiation attenuated *P. yoelii* immunization, which were labeled by a fluorescent dye specific to CD8α receptors found on all CD8 T cells.

Thus, this dataset most likely includes clusters with a mixture of T cell types, some specific, some non-specific to malaria.

3.2.2 Mathematical models

Our three basic models used in Chapter 1 are the random entry/exit (Random EE) model, the density-dependent (DD) recruitment model and the density-independent (DI) exit model. These models, as well as the differential equations framework for all of our single-cell-type models are described in detail in Section 1.2.2 in Chapter 1.

Alternative decreased per capita exit rate model

While the density-independent exit rate model is used to describe a mechanism where the per capita exit rate from a cluster decreases as the number of T cells in it increases, there are alternative ways to express this relationship. Our alternative decreased exit rate models also consider cluster formation as a result of constant entry rates, $\lambda_k = \lambda$, and the total exit rate can be either $\mu_k = k\mu e^{-\alpha k}$, where the per capita exit rate exponentially decreases with increasing cluster size, or $\mu_k = k\mu k^{-\alpha}$, where the proportion of decrease as a function of the number of cells, k, is scaled by α .

Gamma-distributed entry rate models

We can also hypothesize that T cells do not have a constant entry rate into clusters around infected cells, but that there is a variation around a mean entry rate, according to a Gamma distribution, which allows for the creation of all of the cluster sizes observed. We can then express the probability of a cluster of size k as an integral of P_k values over all possible entry rates times the probability densities of those Gamma

distributed entry rates, as follows

$$\begin{aligned} P_k &= \int_0^\infty P_k(\lambda) \cdot g(\lambda; \alpha, \beta) d\lambda \\ &= \int_0^\infty \frac{\lambda^k}{\mu^k k!} e^{-\frac{\lambda}{\mu}} \cdot \frac{\beta^\alpha \lambda^{\alpha-1} e^{-\lambda\beta}}{\Gamma(\alpha)} d\lambda, \end{aligned}$$

where α is the shape parameter and β is the rate parameter of the Gamma distribution. After integration we obtain the probability of a cluster of size k

$$P_k = (\frac{1}{\mu} + \beta)^{-(\alpha+k)} \cdot \frac{\beta^\alpha \Gamma(\alpha + k)}{\mu^k k! \Gamma(\alpha)}, \tag{3.1}$$

where $\Gamma(\alpha) = (\alpha - 1)!$

Model with two subpopulations of infected cells associated with distinct non-density-dependent entry rates

This model has the underlying assumption that the observed extreme sizes of T cell clusters around infected cells are a result of two cluster types that have two different entry/exit rate ratios associated with them. This would be possible if infected cells attracted different numbers of T cells due to differences in T-cell-extrinsic factors, such as signaling phenotypes or locations in the liver. Therefore, the probability of a cluster of k T cells is a sum of the two probabilities, each with its own θ_0 parameter and multiplied by the fraction of the cell population belonging to that sub-population.

$$P_k = f \cdot P_k(\theta_{01}) + (1 - f) \cdot P_k(\theta_{02}), \tag{3.2}$$

where $\theta_0 = \frac{\lambda_0}{\mu_0}$ and $P_k(\theta_0) = \frac{\theta_0^k}{k!} e^{-\theta_0}$.

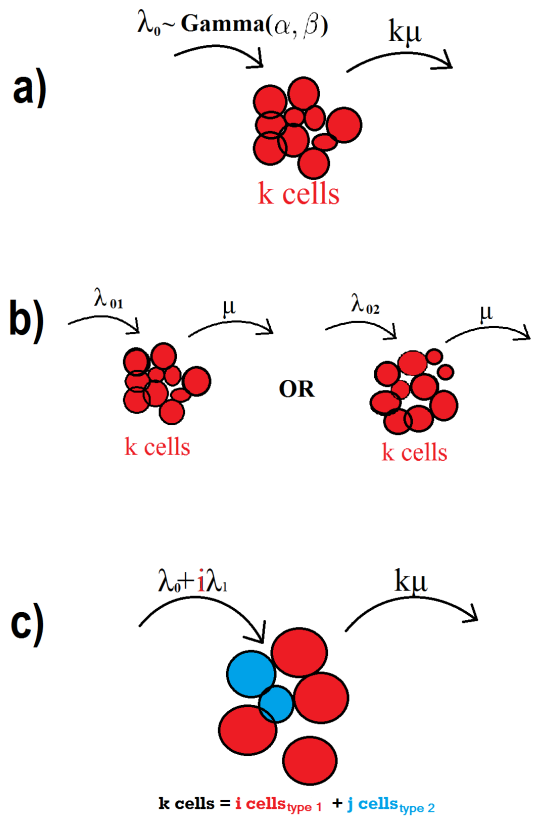

Figure 3.1: Cartoon representation of three models introduced in this chapter. a) shows the Gamma distributed entry rate model, where α and β are the rate and shape parameters for the distribution that the T cell entry rates for each infected cell are drawn from. b) represents the two entry rate model, where a fraction f of the infected hepatocyte population is associated with one entry rate, and $1 - f$ fraction is associated with another entry rate. c) demonstrates a case under the mixed cluster modeling framework, where the recruitment rate is dependent on the density of only cell type 1.

Mixed models containing two T cell types

To mathematically represent the experiments involving both PyTCR and OT-1 cells in malaria-infected mice and to test hypotheses about how these two T cell types find infected cells and form mixed clusters around them we have developed a two-T-cell-type model framework. This framework can also be used to obtain the likelihood of two T-cell-subsets within a certain imaged T cell population that are associated with distinct entry to exit rate ratios. We define our state variables in the form of P_{ij},

which is the probability of a cluster containing i cells of type 1 and j cells of type 2, and denote the entry and exit rates by a T cell of type x into such a cluster as λ_{ij}^{x} and μ_{ij}^{x} respectively (Figure 1.1). Then our system of differential equations is

$$\begin{aligned}
\frac{dP_{00}}{dt} &= -(\lambda_{00}^{t_1} + \lambda_{00}^{t_2})P_{00} + \mu_{10}^{t_1}P_{10} + \mu_{01}^{t_2}P_{01}, \\
\frac{dP_{01}}{dt} &= -(\lambda_{01}^{t_1} + \lambda_{01}^{t_2})P_{01} + \lambda_{00}^{t_2}P_{00} + \mu_{11}^{t_1}P_{11} + \mu_{02}^{t_2}P_{02} \\
\frac{dP_{ij}}{dt} &= -(\lambda_{ij}^{t_1} + \lambda_{ij}^{t_2})P_{ij} + \lambda_{(i-1)j}^{t_1}P_{(i-1)j} + \lambda_{i(j-1)}^{t_2}P_{i(j-1)} + \mu_{(i+1)j}^{t_1}P_{(i+1)j} + \mu_{i(j+1)}^{t_2}P_{i(j+1)},
\end{aligned} \tag{3.3}$$

Then some of the ways to define the entry rates into a cluster with i cells of type 1 and j cells of type 2 by a T cell of type x are:

$\lambda_{ij}^{x} = \lambda_0$, where the entry rate is random and shared by both cell types,

$\lambda_{ij}^{x} = \lambda_0 + \lambda_1 \cdot (i + j)$, where both cell types enter and recruit both cell types with the same rate,

$\lambda_{ij}^{x} = \lambda_0 + \lambda_1 \cdot i$, where both cell types have the same basic entry rate, but only T cells of type 1 recruit to clusters,

$\lambda_{ij}^{x} = \lambda_{0x} + \lambda_1 \cdot j$, where both cell types have type-specific basic entry rates, and only T cells of type 2 recruit to clusters,

$\lambda_{ij}^{x} = \lambda_0 + \lambda_{1type1} \cdot i + \lambda_{1type2} \cdot j$, where both cell types have the same basic entry rates, and recruit to clusters with different rates,

$\lambda_{ij}^{x} = \lambda_0 + \lambda_{1type1} \cdot i + \lambda_{1type2} \cdot j$, where both cell types have the same basic entry rates, and recruit to clusters with different rates,

$\lambda_{ij}^{x} = \lambda_0 + \lambda_1 \cdot i$, if $x = type1$, where both cell types have the same basic entry rates, and T cells of type 1 recruit only T cells of the same type to clusters. Similarly, we can define the exit rates out of mixed clusters to have various dependences on the cells already in the cluster.

3.3 Results

3.3.1 Entry and exit rates observed 4-8 hours PTCT suggest that cluster formation is near steady state at the time of imaging

In our initial analysis we assumed that the system of cluster formation 6 hours PTCT is at a steady state and entry and exit rates do not disturb the probabilities of the cluster sizes. Our first reason for this assumption was that at the time of imaging, 6 hours after transferring florescently labeled T cells into infected mice, there has been time for the formation of clusters, even of size 25. When observed between 4 and 8 hours post T cell transfer, relatively few T cells entered or exited the 40 μm circles around the parasites (on average 0.3 and 0.2 cells/hour respectively), the maximum cluster size change over the observed 4 hours was 2 cells and the mean change was 0.31 cells. The directions of the cluster size changes in the observed time interval were mostly positive but not statistically different from random (Fisher sign test $p > 0.05$). When using our DD recruitment model, which reaches near steady state at 6 hours PTCT (μ_0=3/hour, as shown in Figure 3.1), to sample pairs of 32 cluster sizes at 4 and 8 hours PTCT iteratively 100,000 times, we obtained a mean change in average cluster sizes of 0.25 with a standard deviation of 0.73. The change in mean cluster sizes that was experimentally observed, 0.31 cells, was well within one standard deviation of the theoretical expected change in a system that reaches steady state at 6 hours PTCT.

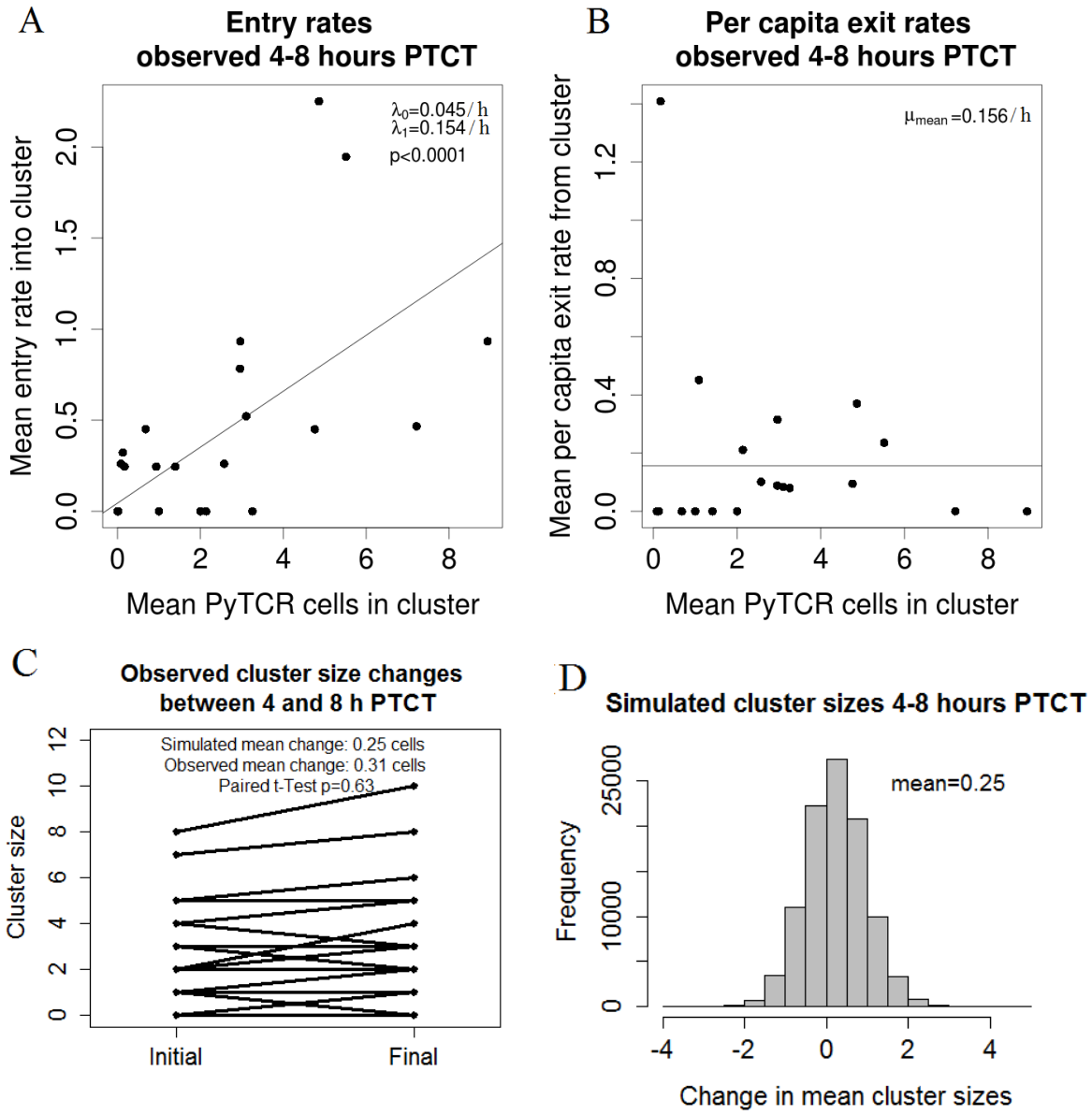

Figure 3.2: Cluster size changes measured 4-8 hours PTCT are small, but significantly depend on the mean number of cells in the cluster.(A and B) Mean entry and per capita exit rates as a function of the mean cells in the clusters between 4 and 8 hours PTCT. A regression line is shown in (A), and the estimated basic entry rate and recruitment rate are shown. F-test yielded a significant dependence on cluster size. (B) shows the mean per capita exit rate, which was found to be independent of cluster size (F-test p=0.5). (C) 32 observed clusters show cluster size change that is not significantly different from an expected mean change of 0.25 cells. (D) When simulating cluster size changes 4-8 hours PTCT using our DD recruitment model with $\mu_0 = 3$/hour, we have found a mean change of 0.25 with the displayed distribution.

Based on the significant dependence of the mean T cell entry rate on the mean cluster size in the observed data ($p < 0.0001$ with F-test, Figure 3.2 A), we were able to substitute these rates into our best fitting DD recruitment model to test if the parameters observed 4-8 hours post T cell transfer could have given rise to large clusters in 6 hours (Figure 3.3). While the estimated cluster size probabilities at 6 hours PTCT deviate within 50% of the observed frequencies for clusters smaller than four cells, this predicted distribution has too low probabilities for large clusters. When using the obtained distribution of cluster sizes at 6 hours PTCT to iteratively sample cluster sizes around 130 infected cells (consistently with the number of imaged infected hepatocytes) we found no occurrence of clusters with 25 cells or more in 10^6 re-samplings.

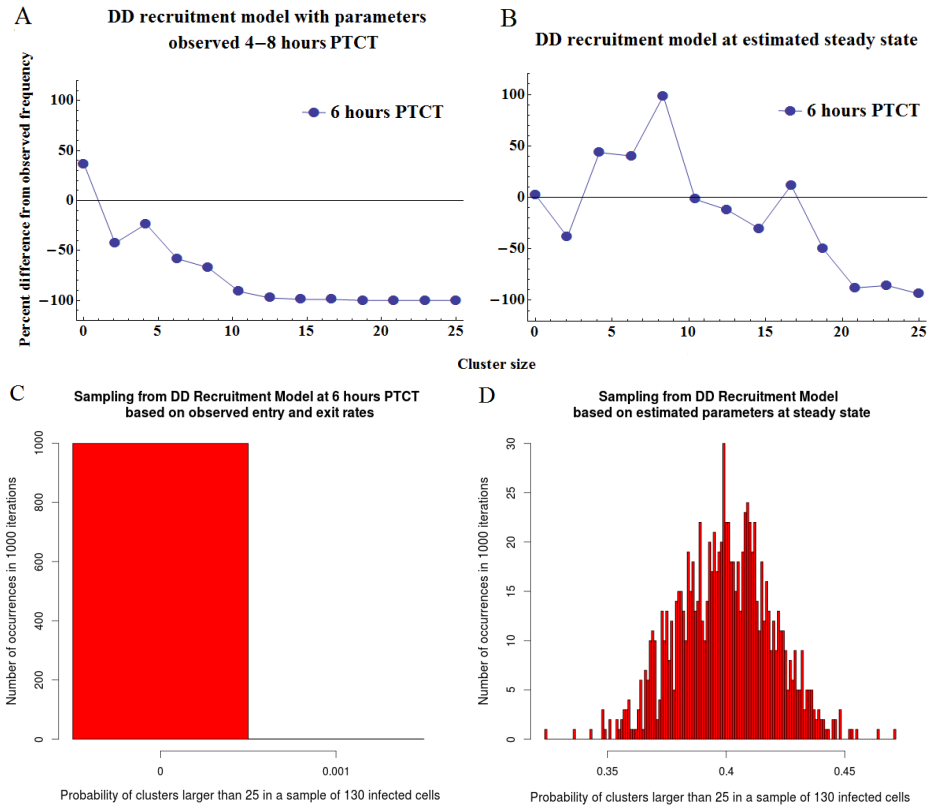

Figure 3.3: Probabilities of cluster sizes were estimated with our DD recruitment model using entry and exit rates observed 4-8 hours PTCT or using our steady state parameters $\theta_0 = 0.2$, $\theta_1 = 0.88$, and percent differences from the observed frequencies were calculated. (A) shows that using parameters observed 4-8 hours PTCT the distribution at 6 hours PTCT has too low probabilities for large clusters. (B) demonstrates the percent differences from observed if we use the estimated steady state distribution. (C and D) We sampled 1000 times a 130 clusters from our obtained distributions 6 hours PTCT and repeated the process 1000 times. Shown are the histograms of the probabilities of seeing clusters larger than 25 cells. The DD recruitment model cannot give rise to any clusters of those sizes with the parameters observed 4-8 hours PTCT, while at steady state the average probability of large clusters is 0.4.

On the contrary, when iteratively sampling 130 infected cells from the fitted steady state of the DD recruitment model with estimated parameters (θ_0=0.20 and θ_1=0.88) we found on average a 40% chance of observing a cluster of size 25 or larger. Assuming a DD recruitment model with the entry and exit rates observed 4-8 hours PTCT the cluster formation process would not have reached an equilibrium in as long as 48

hours. The observed entry and exit rate ratios, $\theta_0 = 0.28$ and $\theta_1 = 0.98$, however, are close to the estimated steady state parameter values of 0.20 and 0.88. This suggests that it is likely that the cluster formation process had reached its steady state between 4 and 8 hours PTCT and the entry and exit rates that Cockburn *et al.* were observing were minor changes at steady state.

3.3.2 If cluster sizes are not at equilibrium, DD recruitment model still fits PyTCR data best

When we set aside the assumption that the observed cluster size frequencies are at equilibrium and fitted models to the "PyTCR alone" dataset using numerical solutions, the DD recruitment model still provided the same likelihood as at steady state, and had the best AIC score among all models tested for a given exit rate (Table 3.1).

Table 3.1: DD recruitment model still fits best when using various fixed exit rates for fitting the "PyTCR alone" dataset using numerical solutions at 6 hours PTCT. We fixed the per capita exit rates, μ_0 values, to those shown in the top row, and used numerical solutions in a MLE approach to find best fit and estimate entry/exit rate parameters, of which the ones corresponding to the DD recruitment model are shown. AIC scores were then calculated for each model's numerical fit with exit rates fixed to the various values shown.

	μ_0=0.06/h	μ_0=0.15/h (observed)	μ_0=3/h
$\mathbf{AIC_{DD\ recruitment}}$	**389**	**389**	**392**
$\mathbf{AIC_{Random\ EE}}$	772	772	740
$\mathbf{AIC_{DI\ exit}}$	741	701	479
$\mathbf{AIC_{DD\ exit,\ exponential}}$	749	714	448
$\mathbf{AIC_{DD\ exit,\ power\ law}}$	746	699	452

To be certain that the relatively poor fit of the hypothesis that cluster growth is solely due to a retentive property of T cells in clusters is not due to the simplicity of our DI exit model, we tested alternative models with density-dependent exit (see Section

(3.2.2)). These models, however, also lagged behind the DD recruitment model, by more than $\triangle = 80$, and more than $\triangle = 40$ at steady state. Curiously, we have found that the estimated entry/exit rate ratios (θ_0, θ_1) in the DD recruitment model were not robust to changes in the per capita exit rate and were inversely proportional to the provided exit rates in the numerical fitting. As seen in Table 3.2, the estimated entry/exit ratios decreased as the fixed per capita exit rates were increased. Using the steady state solutions to fit data at 6 hours PTCT, thus, we obtain the minimal estimates of θ_0 and θ_1 parameters. Interestingly, the recruitment rate of each T cell was consistently 4.3 times the basic entry rate into clusters, which therefore was true for the ratios, θ_0 and θ_1, of those parameters to the exit rate (Table 3.2). It is notable that the λ_0 and λ_1 values estimated from the total entry rates observed 4-8 hours PTCT in 32 clusters also approximately have a ratio of 4 (Figure 3.2).

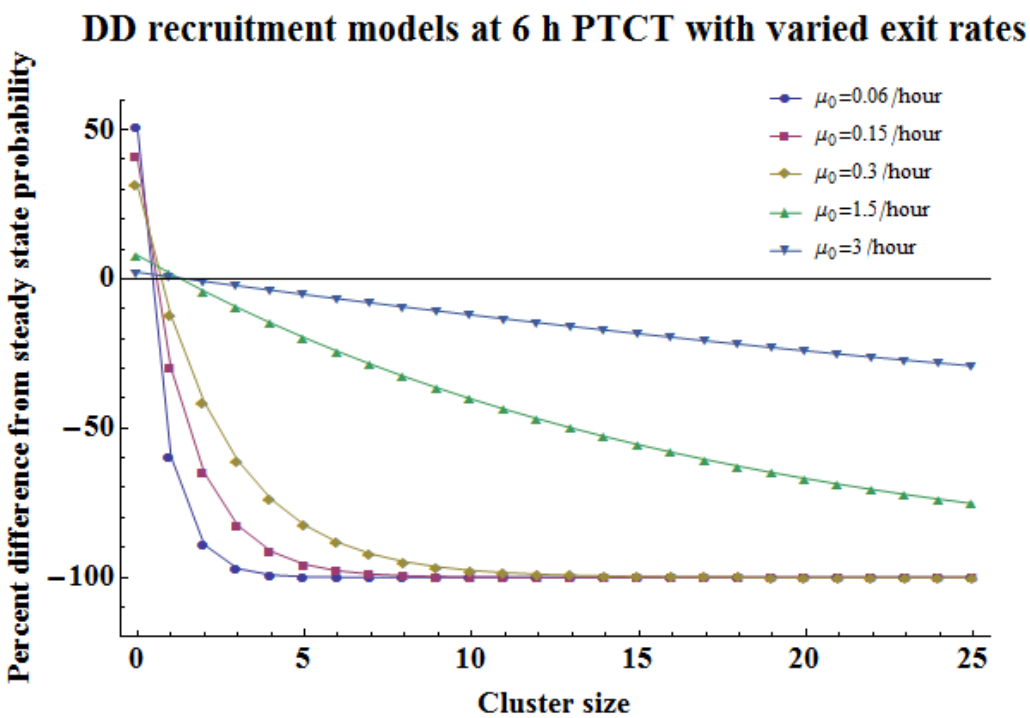

Figure 3.4: Percent differences from steady state probabilities as the per capita exit rate, μ_0, is fixed to the values indicated. The DD recruitment model was used to obtain the cluster size distributions at 6 hours PTCT. As apparent in the figure, a per capita exit rate of $\mu_0 = 3/hour$ results in cluster size frequencies close to the estimated steady state.

Table 3.2: Estimated entry/exit rate ratio parameters of the DD recruitment model decrease as the exit rate used for model fitting is increased. At steady state, which is approached closely in 6 hours PTCT if μ_0=3/hour, the estimated rates are minimal. $\theta_1/\theta_0 \approx$4.3 consistently in all cases.

	μ_0=0.06/h	μ_0=0.15/h (observed)	μ_0=0.3/h	μ_0=1.5/h	μ_0=3/h
Estimated θ_0	1.44	0.63	0.40	0.22	0.20
Estimated θ_1	6.45	2.86	1.80	0.95	0.88

Figure 3.4 illustrates that in order to approach closely the steady state distribution of cluster sizes in the first 6 hours PTCT, the per capita exit rate has to be at least 3/hour.

3.3.3 Tests of alternative hypotheses that assume variance in T cell entry rates due to T-cell-extrinsic or T-cell-intrinsic factors

Activated CD8 T cells migrate to sites of inflammation, such as the liver during the initial stage of malaria, with the help of blood flow, adhesion molecules on the endothelium, chemoattractants expressed by infected cells and immune cells [4], and their likelihood of remaining at infected cells is most likely influenced by whether they associate with their target antigens. The combinations of functional receptors that the imaged T cell populations possess to sense the above molecular information (see Section 3.2.1) are likely key determinants of the observed distributions of cluster sizes around infected hepatocytes. Previously we have shown that the hypothesis with most statistical support based on its mathematical model is that malaria-specific T cells as well as non-malaria-specific T cells have a constant ratio of entry and exit rates (estimated to be around 0.2), but malaria-specific T cells can additionally have entry rate increase per one T cell in the cluster (estimated ratio to exit rate is 0.88 per cell in cluster)[8]. Here we are adding a new distribution of clusters to

our analyses, containing GPCR-disabled PyTCR cells (PTx cells), which suggests that without the ability to sense chemoattractants, PyTCR cells cannot have a T-cell-density-dependent increase in their entry rates into clusters around infected cells (Figure 3.5).

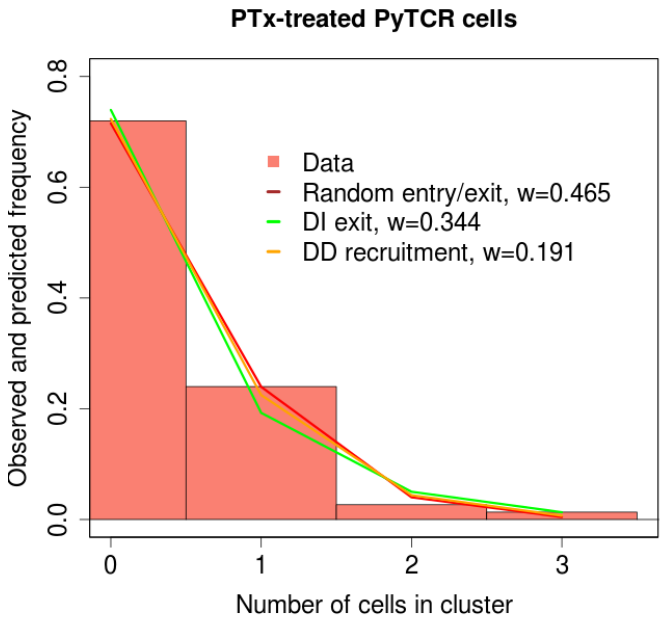

Figure 3.5: Fitting our three basic models, which are described in section 1.2.2 to the PTx data we find not enough support from the DI exit and DD recruitment models to reject the Random entry and exit model. "w" stands for AIC weights, which give the likelihood of a model relative to other models ($w < 0.01$ suggests unlikely model).

This section introduces and tests alternative hypotheses that could potentially give rise to the observed cluster distributions. It is possible that the accumulation of large numbers of T cells at certain infected hepatocytes while not at others is simply due to variation in T-cell-extrinsic factors that determine the entry rate, such as blood flow, accessibility within the tissue structure, chemokine secretion by the infected cell or nearby unlabeled immune cells, or abundance of adhesion molecules on surrounding endothelial cells. We also refine our hypothesis that attributes large

clusters to T-cell-intrinsic factors, such as the ability to sense and execute density-dependent recruitment, and extend it to two T cell subpopulations that have possibly different entry rates and contributions to the recruitment into clusters around infected cells.

T-cell-extrinsic factors cannot consistently explain cluster distributions of all cell types

We created two models to test the hypothesis that the observed cluster size distributions of all datasets can be explained by T-cell-extrinsic conditions that give rise to variation among infected hepatocytes in T cell entry rates. Our continuous model assumes a density-independent entry rate into clusters that is drawn from a gamma distribution to allow variation around a mean entry rate, and a more discrete model, where two separate non-density-dependent entry rates are each associated with a certain fraction of the infected hepatocytes (see Equations (3.1) and (3.2)). When fitting the model in which the entry rates into clusters are drawn from a gamma distribution we found that it described the PyTCR alone, OT-1 alone and PTx data similarly to their respective best-fitting models. We chose these three datasets because they are results of experiments, in which single cell types were used and therefore should allow for distinguishing between T-cell-intrinsic and T-cell-extrinsic effects with the least amount of uncontrolled effects. The mean per capita entry rates for these three cell types, however, were highly variable (4.7, 0.49 and 1/hour for PyTCR, OT-1 and PTx respectively, with the exit rate fixed to 3/hour), and were not consistent with an underlying hypothesis that the mean entry rates are determined by T-cell-extrinsic factors, such as blood flow, and only the variances in entry rates into clusters are different possibly due to different sensitivities of PyTCR, OT-1 and PTx cells to chemoattractants. For instance, some infected hepatocytes might have higher rates of chemokine secretion than average, but if one imaged T cell population does not have the capacity to sense them, then the variance around the mean entry rate would be smaller for that population. Therefore we tested a

model where the mean entry rate was shared among the three cell types, and only the variances differed among them.

As seen in Figure 3.6 and Table 3.3, this hypothesis does not fit as well as the DD recruitment model for PyTCR or the Random entry/exit for OT-1 and PTx cells, and when compared to a similarly constrained DD recruitment model, where θ_0 is shared among the three datasets and only the recruitment rates are allowed to vary, we see an AIC score difference of 40 between the two types of constraints. Table 3.3 shows that if we allow the mean entry rates to be different in the three datasets to accommodate different abilities of basic T cell migration and constrain the degree of T-cell-extrinsic variance among entry rates, the resulting model also fits the data poorly ($\triangle$=35).

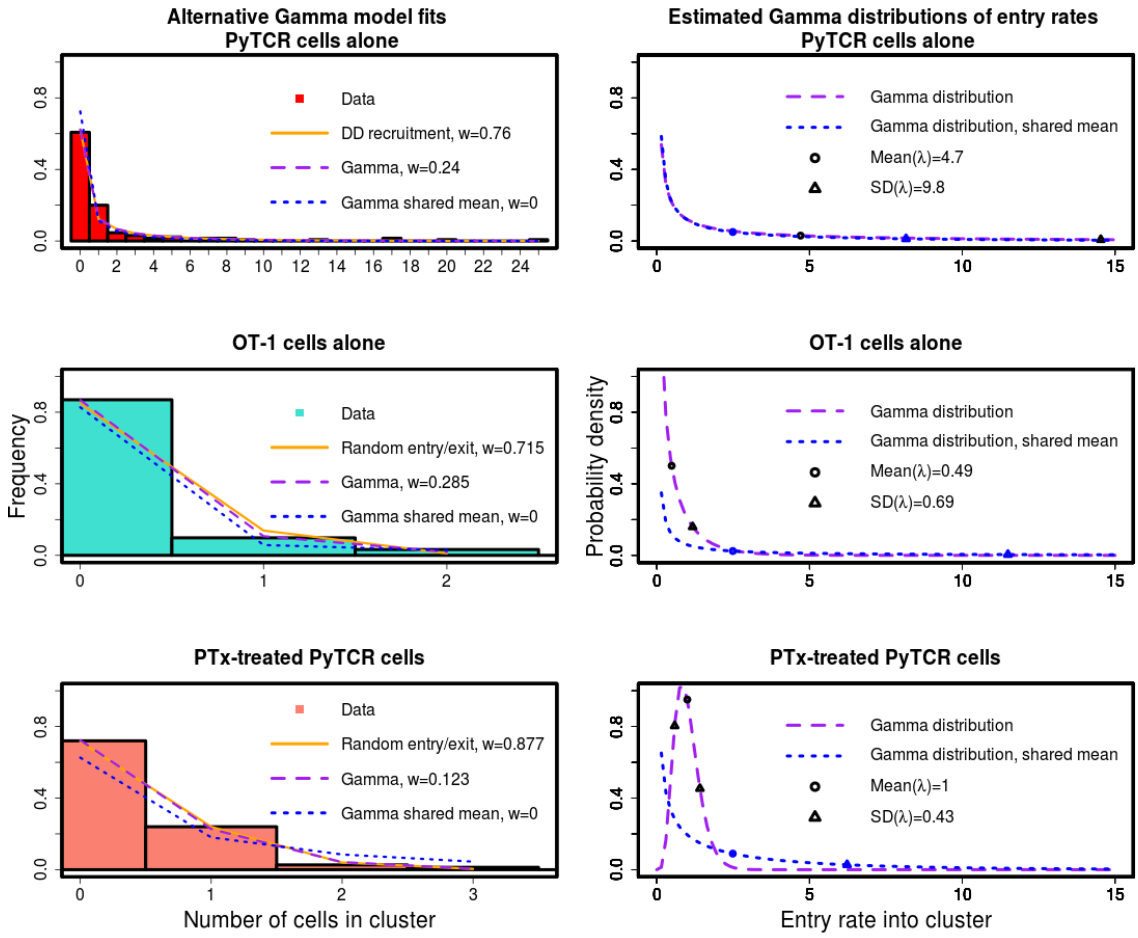

Figure 3.6: Gamma distributed entry rate models fit PyTCR, OT-1 and PTx datasets well, but with high discrepancies between estimated distributions of entry rates of the cell types. When mean entry rate is shared among three cell types the comparison of fits and AIC weights of these models to the respective best models shows no statistical support for the hypothesis that variance in entry rate alone can explain small and large clusters. Exit rates were fixed to 3/hour.

Table 3.3: Gamma distributed entry rates explain PyTCR, OT-1 and PTx data well, but yield highly different means and variances, which suggest different effect of T-cell-extrinsic factors on different cell types. When we constrained either the mean or variance of entry rates consistently over datasets, AIC values showed no support for those models. Likelihood ratio test was also performed and found a significant increase in fit quality if added T-cell-type-specific means and variances of entry rates. The exit rate was fixed to 3/hour when estimating the entry rate parameters, and the table shows the ratio of estimated entry rates to the exit rate. A discrete model with two subsets of entry rates into clusters did not explain these three datasets well, as shown by the high AIC values. The value f stands for the fraction of infected cells with entry and exit rate ratio, θ_{0f_1}, thus $1-f$ is the fraction of infected cells with rates ratio θ_{0f_2}.

Model	**PyTCR cells alone**	**OT-1 cells alone**	**PTx cells**	**AIC**
Gamma distributed entry (DE)	mean(θ_0)=1.6, sd(θ_0)=3.3	mean(θ_0)=0.16, sd(θ_0)=0.23	mean(θ_0)=0.3, sd(θ_0)=0.14	**598**
Gamma DE, shared mean	mean(θ_0)=0.83, sd(θ_0)=1.9	mean(θ_0)=0.83, sd(θ_0)=3	mean(θ_0)=0.83, sd(θ_0)=1.2	639
Gamma DE, shared variance	mean(θ_0)=1.1, sd(θ_0)=2.13	mean(θ_0)=0.57, sd(θ_0)=2.13	mean(θ_0)=0.93, sd(θ_0)=2.13	634
Two entry rates	θ_{0f_1}=10.6, θ_{0f_2}=0.46, $f=0.11$	θ_{0f_1}=0.46, θ_{0f_2}=0, $f=0.35$	θ_{0f_1}=1.07, θ_{0f_2}=0.29, $f=0.052$	625
Two entry rates, shared fraction of hepatocytes, estimated $f=0.11$	θ_{0f_1}=10.5, θ_{0f_2}=0.46	θ_{0f_1}=0.82, θ_{0f_2}=0.08	θ_{0f_1}=0.09, θ_{0f_2}=0.37	622
DD recruitment, shared θ_0	θ_0=0.2, θ_1=0.88	θ_0=0.2, θ_1=0.087	θ_0=0.2, θ_1=0.22	**599**
Best model estimates for each dataset	θ_0=0.2, θ_1=0.88	θ_0=0.16,	θ_0=0.34	**600**

These results provided essentially no support for the hypothesis that large clusters are solely results of variation in T-cell-extrinsic factors around infected hepatocytes, which might be sensed to different degrees by various CD8 T cell types, thus resulting in some T cell populations' tendency to form big clusters, and some T cell populations' lack of that. However, based on the relatively good explanation of the single cell type data by the Gamma distributed entry rate model where both the mean and variance of entry rates are allowed to be T cell type specific, we cannot reject that the formation of clusters is driven by variation in T-cell-extrinsic factors affecting both the mean entry rate and the variation from it, which is sensed to different degrees by the three T cell populations. While the continuous version of our variable entry rate model fits the three datasets well when allowing T-cell-type-specific means and variances to be estimated, our discrete model, where an estimated fraction of infected hepatocytes has a certain T cell entry rate associated with them, and the rest of them have a different estimated entry rate, does not provide a good overall fit of the three datasets ($\triangle$ >20).

When fitting the cluster distribution of the endogenous T cell population with models based on variation between infected cells in T-cell-extrinsic conditions we find that the hypotheses with T-cell-density-dependent retention or recruitment to clusters have a higher relative likelihood (Figure 3.7).

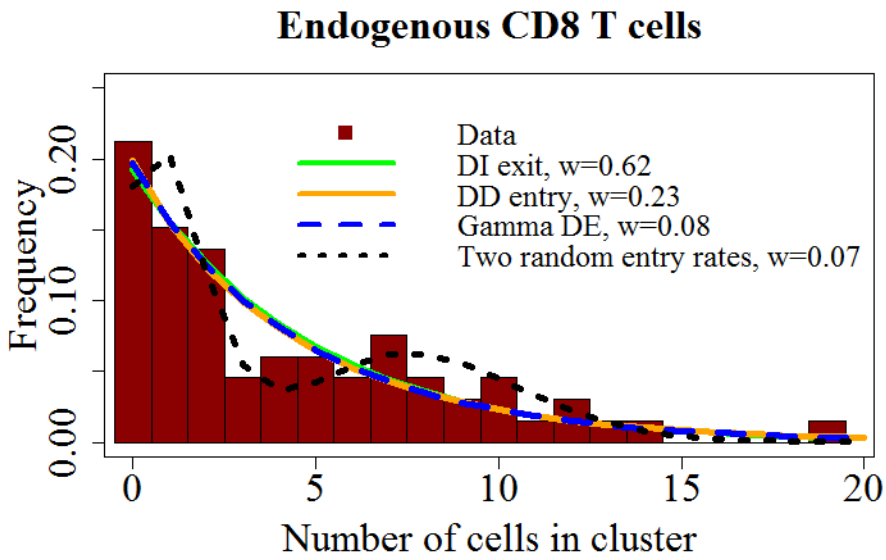

Figure 3.7: Endogenous CD8 T cells do not show strong evidence for T-cell-extrinsic reasons to observed cluster distribution, based on our Gamma distributed entry rate model and model with two subsets of infected cells with different T cell entry rates associated with them. "w" stands for AIC weights, which give the likelihood of a model relative to other models ($w < 0.01$ suggests unlikely model).

As shown above, based on the datasets containing single cell types of PyTCR, OT-1 and PTx PyTCR, we cannot distinguish between T-cell-intrinsic mechanisms, in this case T-cell-density-dependent recruitment, and T-cell-extrinsic causes that might have given rise to T cell clusters as large as 25 cells around infected liver cells. We therefore turned to comparing entry rate parameters of the same T cell type, OT-1, in two different experiments, one where only OT-1 cells were transferred into mice, and one where OT-1 and PyTCR cells were transferred together. Parameters were estimated through either assuming T-cell-extrinsic causes of variance in entry rates, using our Gamma model, or assuming T-cell-intrinsic mechanisms, with our DD recruitment model. As seen in Chapter 1, OT-1 cells have drastically different cluster size distributions in these two experiments, and when transferred together with PyTCR cells, OT-1 clustering is enhanced. In order to fit the OT-1 cluster distribution in the co-transfer experiment similarly well as the DD recruitment model

does, the mean and variance of entry rates are estimated to have a 10-fold increase from those in the single transfer experiment (Table 3.4). In contrast, the DD recruitment model estimates a basic entry to exit rate ratio of 0.16 and 0.22 for OT-1 cells when alone and when co-transferred, respectively, and the per capita recruitment rate to exit rate ratio increases from 0 to 0.79 when PyTCR cells are present as well.

Table 3.4: Estimated mean and variance of entry rates by OT-1 cells into clusters increase by nearly 10-fold when co-transferred with PyTCR cells, when assuming a Gamma model, where entry rates only vary because of T-cell-extrinsic reasons. In the density-dependent recruitment model, however, OT-1 cells have similar estimated basic entry rates in both experiments, and only the T-cell-density-dependent recruitment rates are changed. During fitting of the Gamma model, we used exit rates of 3/hour.

Model	**OT-1 alone**	**OT-1 mixed**	**AIC**
Gamma DE	mean(θ_0)=0.16, sd(θ_0)=0.23	mean(θ_0)=1.1, sd(θ_0)=2	**230.6**
Gamma DE, shared mean	mean(θ_0)=2, sd(θ_0)=6.5	mean(θ_0)=2, sd(θ_0)=4	245.2
DD recruitment	θ_0=0.16, θ_1=0	θ_0=0.22, θ_1=0.79	**230.8**

These results suggest that the presence of malaria-specific, PyTCR cells either increases the OT-1 cells' sensibility to T-cell-extrinsic signals by 10-fold, or that the basic entry rate by OT-1 cells remains nearly constant throughout the two experiments, and only the T-cell-density-dependent recruitment rate into clusters is increased when co-transferred with PyTCR cells.

T-cell-intrinsic differences in recruitment to clusters are apparent in PyTCR and OT-1 cells in experiments of co-transfer

The fluorescently labeled T cell populations that were imaged by Cockburn *et al.* were CD8 T cells including either a variety of antigen-specificities, such as endogenous

CD8 T cells, or were a subset of T cells specific to one antigen, as in the case of PyTCR, PTx PyTCR and OT-1 cells. Even CD8 T cells activated by the same antigen have been shown to be polyfunctional and of various phenotypes [24] in terms of chemokine secretion and receptor expression. Pertussis toxin treatment of lymphocytes is known not to be 100% efficacious [31], and thus a subset of PyTCR cells in the treated samples might have retained their GPCR activity. Therefore, in both PyTCR and PTx PyTCR populations, it is possible that there are further subsets of the imaged population that differ in their basic entry rates into clusters and possibly in their abilities to send or sense recruitment signals. To obtain the relative likelihood of multiple T cell entry and perhaps recruitment rates giving rise to the observed cluster size distributions we formulated a mixed cluster modeling framework (see Section 3.2.2). These models distinguish between two T cell types and compute probabilities, P_{ij}, for clusters with i cells of one type and j cells of another type (Figure 1.1). The recruitment rate into a mixed cluster with i and j cells can then be defined using various i- or j-dependent expressions in the model in order to assume various contributions from each cell type. Hence, these mixed cluster models allow for variation between entry and possibly recruitment rates within the labeled T cell population, which are mostly T-cell-intrinsic.

We have found very little evidence for two subpopulations of T cells with distinct entry rates within the populations of PyTCR, OT-1, PTx cells and even within the endogenous CD8 T cell population using our mixed cluster models. Likelihood ratio tests of nested models were performed and in all cases suggested that the addition of subpopulation-specific parameters did not increase the fit of the models significantly ($p > 0.2$ for all T cell populations).

Previously we have shown that when transferred into mice together with PyTCR cells, the OT-1 cell number distribution around infected cells suggests a density-dependent recruitment of cells into clusters. However, it is important to consider the two cell types together in our analysis of these experimental results, as the recruitment suggested by our model comparisons might not be the effect of OT-1 cells.

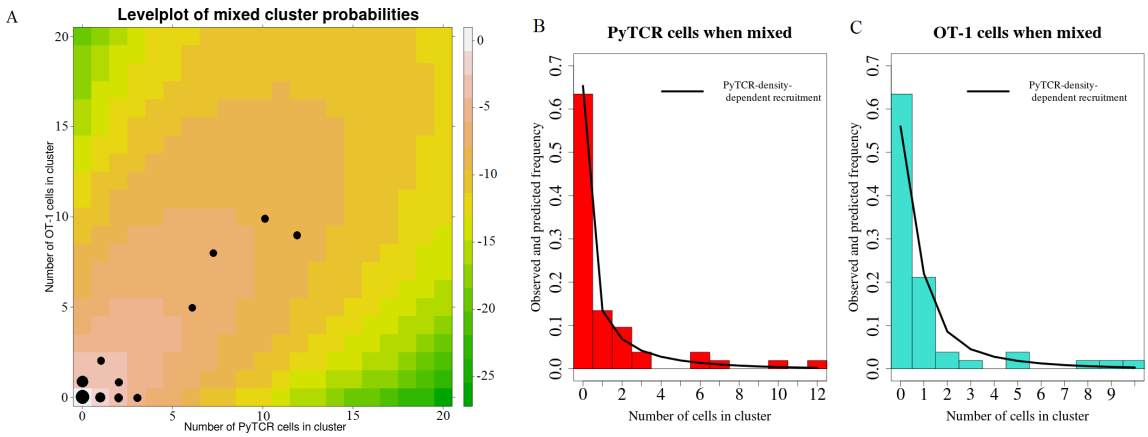

Figure 3.8: (A) Estimated probability distribution of mixed cluster sizes according to the best fitting model, where only PyTCR cells recruit, with data overlaid using points whose diameters are proportional to the number of clusters observed. Legend shows the natural logarithm of the probability values. (B and C) Estimated and observed probabilities of PyTCR and OT-1 cell numbers around infected cells when transferred together.

Therefore, we used our mixed modeling framework to test which T cell type might contribute more to the recruitment into clusters around infected cells.

We have found that the PyTCR-density-dependent recruitment model is the best hypothesis to explain the observed mixed cluster distribution. When distinguishing between the similarly fitting models of PyTCR cells recruiting versus PyTCR and OT-1 cells both recruiting, likelihood ratio test of the two nested models showed that adding the extra parameter of OT-1 recruitment rate does not create a significantly better fit ($p > 0.25$ in likelihood ratio test). Similarly, the addition of basic entry rates specific to each cell type does not significantly increase the fit to the data compared to a shared entry rate ($p > 0.25$ in likelihood ratio test). Consistently with our parameter estimated for the PyTCR alone dataset, the per capita recruitment rate of PyTCR cells in this mixed experiment is approximately four times of the basic per capita entry rate of T cells into clusters (Table 3.5). Figure 3.8 displays the fit of the best model, in which the recruitment rate into mixed clusters only depends on the number of PyTCR cells that they contain. The levelplot in Figure 3.8A is used to demonstrate the coincidence of high probability mixed clusters with those that are

observed, while the plots B and C in Figure 3.8 show the fit of estimated probabilities for each cell type's numbers in the mixed clusters.

Table 3.5: This table lists the basic entry/exit rate values $\theta_0 = \lambda_0/\mu_0$, which are shared by both cell types. Likelihood ratio found no significant increase in fit quality when estimating cell-type-specific entry rates. Additional entry rates, due to density-dependent recruitment, divided by exit rates are indicated by $\theta_1 = \lambda_1/\mu_1$. For instance, "PyTCR" in the column of θ_1 means that the recruitment rate/exit rate is exerted by PyTCR cells and extends onto both cell types, while "PyTCR:OT-1" denotes the recruitment rate/exit rate by PyTCR cells that is sensed by OT-1 cells. The mixed cluster model that defines the recruitment rates into clusters to be only dependent on the number of PyTCR cell density predicts the observed mixed clusters the best, as shown by AIC scores. The addition of an OT-1 dependent recruitment gives a fit with a similar AIC value, but likelihood ratio test of these nested models shows that this more complex model does not significantly increase the quality of the fit ($p > 0.25$). Interestingly, the estimated recruitment rate is also about 4.3 times the estimated basic entry rate.

Model	θ_0 **values**	θ_1 **values**	**AIC**
PyTCR cells recruit	0.21	0.82	**260.0**
OT-1 cells recruit	0.22	0.83	267.8
PyTCR and OT-1 cells recruit with same rate	0.17	0.43	264.3
PyTCR and OT-1 cells recruit with different rates	0.3	PyTCR=1.1, OT-1=-0.33	**259.2**
PyTCR and OT-1 cells recruit with different rates towards different cell types	0.32	PyTCR:PyTCR=1.17, PyTCR:OT-1=0.9, OT-1:OT-1=-0.3, OT-1:PyTCR=0.47	264.0

3.4 Discussion

The migration of CD8 T cells into inflamed non-lymphoid tissues is known to be regulated by the interactions between receptors on T cells and specific molecules secreted or displayed by other immune cells, endothelial cells or infected cells [5]. The recent availability of in vivo imaging of fluorescently labeled immune cells has allowed the study of their movement and behavior in non-inflamed as well as inflamed conditions [8, 16, 23]. Using this technique Cockburn *et al.* found that CD8 T cells tend to form clusters as large as 25 cells around infected hepatocytes during liver-stage malaria in mice. To understand the processes that allow some subsets of CD8 T cells to aggregate in high numbers around infected cells, while not other subsets, we turned to statistically comparing mathematical models that incorporated assumptions about the micro-environments around infected cells as well as the T cells' ability to send and sense signals. Our analysis of the cluster size changes measured between 4 and 8 hours PTCT suggested that it is very likely that at 6 hours PTCT, which is the time of recording the cluster size frequencies, they can be assumed to be near steady state. We showed that the DD recruitment model, which was found to be the best fitting model of the formation of large clusters in Chapter 1, gave the best AIC values among the tested basic models even if we don't assume the system to be at equilibrium. The estimates of entry to exit rate ratios did depend on whether we assumed steady state or not, and by using steady state solutions to fit data we estimated the minimal ratios, which are 0.2 and 0.88 for the basic entry rate and the per capita recruitment rate respectively. Interestingly, the estimated per capita recruitment rate into clusters exerted by PyTCR cells in both the single-transfer and the co-transfer experiments was consistently around 4.3 times the estimated basic per capita entry rate. Our tests of hypotheses with T-cell-extrinsic reasons for the observed distributions of T cell cluster sizes around infected hepatocytes demonstrated that the only way for these T-cell-independent factors to explain the behavior of all T cell populations in Cockburn *et al.*'s experiments is if we allow drastic T-cell-type-specific differences in

their effects. The example of non-malaria-specific OT-1 cells clearly shows that the presence of malaria-specific T cells induces a change in their sensitivity for finding and remaining in clusters of T cells. If we try to explain both of these experiments, OT-1 cells alone and in co-transfer, with the T-cell-extrinsic hypothesis, we have to interpret a 10-fold increase in their mean entry rates and in their sensed T-cell-extrinsic variation of entry rates as the effect of the presence of PyTCR cells. It is known that integrins, receptors on T cells that facilitate migration to sites of infection via adhesion to endothelial cells, can be activated through a conformational change, which can also be enhanced by nearby chemoattractants [2]. The ability of PyTCR cells to use their T cell receptors to recognize parasite antigen upregulates their chemokine secretion, which possibly induces integrin and chemokine receptor activation and expression in other T cells that sense them. The parameter estimates of our T-cell-extrinsic model might be interpreted through the above biological scenario, where large clusters are formed around infected cells that attract stronger than others, and the sensitivity to the stronger signals is highest in PyTCR cells and in T cells that they are co-transferred with. On the other hand, if we compare the DD recruitment model's estimated entry and recruitment rates associated with OT-1 cells when they are transferred alone or together with PyTCR cells, we see a consistent level of entry rates by these cells around 0.2, when divided by the exit rate, as well as a consistent lack of recruitment of other T cells to clusters. Our mixed cluster models revealed that assuming T-cell-intrinsic mechanisms of cluster formation, the statistically best fitting hypothesis is that only PyTCR cells contribute, in a density-dependent manner, to the recruitment of even non-malaria-specific T cells to infected cells. To distinguish between T-cell-extrinsic causes or T-cell-dependent recruitment that might underlie the cluster formation process further experiments would be needed. If T cell entry rates associated with each infected hepatocyte are independent of T cell numbers around it then we would expect them to stay constant over time, even though they would be of different magnitudes. However, if entry rates to infected cells are dependent on the cluster sizes already around

them, then we would expect to see an increase of entry rates over time. Even if the environment would change through the hours of imaging, causing increase in entry rates in the T-cell-independent scenario, we would expect to be able to distinguish between change dependent on time versus change dependent on numbers of T cells around each infected hepatocyte. Co-transferring non-malaria-specific OT-1 cells with PyTCR cells that have been treated with pertussis toxin might result in only OT-1 cells forming large clusters, in which case we could confirm that the presence of antigen-specific cells, possibly through chemokine secretion, makes irrelevant T cells more sensitive to environmental signals. However, if OT-1 cells would not form large clusters when co-transferred with non-clustering, GPCR-inactivated PyTCR cells, that would further emphasize that recruitment to a cluster is dependent on the density of antigen-specific T cells. Based on the consistency of our DD recruitment parameter estimates across all datasets of transferred T cells (Tables 1.4 and 3.8), and the more biologically likely explanation of T-cell-density-dependent recruitment to clusters, however, we suggest that the essential role of antigen-specific T cells in clustering most likely lies in their cumulative signaling to other T cells once they associate with infected cells. Since it is likely that T cell clustering increases the likelihood of parasite killing [4, 8] how CD8 T cell clusters are formed and how this process could be enhanced are questions worth deciphering.

Bibliography

[1] Achtert, E., Goldhofer, S., Kriegel, H., Schubert, E., and Zimek, A. (2012). Evaluation of clusterings– metrics and visual support. *Proc. of the 28th Int. Conf. on Data Eng. (ICDE)*. 22

[2] Alon, R. and Dustin, M. L. (2007). Force as a facilitator of integrin conformational changes during leukocyte arrest on blood vessels and antigen-presenting cells. *Immunity*, 26(1):17–27. 62

[3] Boissonnas, A., Fetler, L., Zeelenberg, I. S., Hugues, S., and Amigorena, S. (2007). In vivo imaging of cytotoxic T cell infiltration and elimination of a solid tumor. *J. Exp. Med.*, 204(2):345–356. 14

[4] Bromley, S. K., Mempel, T. R., and Luster, A. D. (2008). Orchestrating the orchestrators: chemokines in control of T cell traffic. *Nat. Immunol.*, 9(9):970–980. 49, 63

[5] Campbell, D. J., Debes, G. F., Johnston, B., Wilson, E., and Butcher, E. C. (2003). Targeting T cell responses by selective chemokine receptor expression. *Semin. Immunol.*, 15(5):277–286. 61

[6] Chang, C.-C. and Lin, C.-J. (2011). LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`. 23

[7] Chtanova, T., Han, S. J., Schaeffer, M., van Dooren, G. G., Herzmark, P., Striepen, B., and Robey, E. A. (2009). Dynamics of T cell, antigen-presenting cell,

and pathogen interactions during recall responses in the lymph node. *Immunity*, 31(2):342–355. 14

[8] Cockburn, I. A., Amino, R., Kelemen, R. K., Kuo, S. C., Tse, S. W., Radtke, A., Mac-Daniel, L., Ganusov, V. V., Zavala, F., and Menard, R. (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. iv, 18, 38, 49, 61, 63

[9] Cockburn, I. A., Chen, Y. C., Overstreet, M. G., Lees, J. R., van Rooijen, N., Farber, D. L., and Zavala, F. (2010). Prolonged antigen presentation is required for optimal CD8+ T cell responses against malaria liver stage parasites. *PLoS Pathog.*, 6(5):e1000877. 14

[10] Egen, J. G., Rothfuchs, A. G., Feng, C. G., Winter, N., Sher, A., and Germain, R. N. (2008). Macrophage and T cell dynamics during the development and disintegration of mycobacterial granulomas. *Immunity*, 28(2):271–284. 14

[11] Ester, M., Kriegel, H., Sander, J., and Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining (KDD-96)*, pages 226–231. 22

[12] Filipe-Santos, O., Pescher, P., Breart, B., Lippuner, C., Aebischer, T., Glaichenhaus, N., Spath, G. F., and Bousso, P. (2009). A dynamic map of antigen recognition by CD4 T cells at the site of Leishmania major infection. *Cell Host Microbe*, 6(1):23–33. 14

[13] Germain, R. N., Robey, E. A., and Cahalan, M. D. (2012). A decade of imaging cellular motility and interaction dynamics in the immune system. *Science*, 336(6089):1676–1681. 17

[14] Goodson, S. G., Zhang, Z., Tsuruta, J. K., Wang, W., and O'Brien, D. A. (2011). Classification of mouse sperm motility patterns using an automated multiclass support vector machines model. *Biol. Reprod.*, 84(6):1207–1215. 27

[15] H Akaike, K Tanabe, E. P. G. K. (1998). *Selected Papers of Hirotugu Akaike*. Springer, New York. 8

[16] Harris, T. H., Banigan, E. J., Christian, D. A., Konradt, C., Tait Wojno, E. D., Norose, K., Wilson, E. H., John, B., Weninger, W., Luster, A. D., Liu, A. J., and Hunter, C. A. (2012). Generalized Lvy walks and the role of chemokines in migration of effector CD8+ T cells. *Nature*, 486(7404):545–548. 14, 17, 61

[17] Hoffman, S. L., Isenbarger, D., Long, G. W., Sedegah, M., Szarfman, A., Waters, L., Hollingdale, M. R., van der Meide, P. H., Finbloom, D. S., and Ballou, W. R. (1989). Sporozoite vaccine induces genetically restricted T cell elimination of malaria from hepatocytes. *Science*, 244(4908):1078–1081. 14

[18] Johnson, J. B. and Omland, K. S. (2004). Model selection in ecology and evolution. *Trends Ecol. Evol. (Amst.)*, 19(2):101–108. 8

[19] K. P. Burnham, D. R. A. (2002). *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach*. Springer-Verlag. 8, 9

[20] Kastenmuller, W., Brandes, M., Wang, Z., Herz, J., Egen, J. G., and Germain, R. N. (2013). Peripheral prepositioning and local CXCL9 chemokine-mediated guidance orchestrate rapid memory CD8+ T cell responses in the lymph node. *Immunity*, 38(3):502–513. 15

[21] Kohlmeier, J. E., Miller, S. C., Smith, J., Lu, B., Gerard, C., Cookenham, T., Roberts, A. D., and Woodland, D. L. (2008). The chemokine receptor CCR5 plays a key role in the early memory CD8+ T cell response to respiratory virus infections. *Immunity*, 29(1):101–113. 15

[22] Mrass, P., Takano, H., Ng, L. G., Daxini, S., Lasaro, M. O., Iparraguirre, A., Cavanagh, L. L., von Andrian, U. H., Ertl, H. C., Haydon, P. G., and Weninger, W. (2006). Random migration precedes stable target cell interactions of tumor-infiltrating T cells. *J. Exp. Med.*, 203(12):2749–2761. 14

[23] Pai, S., Danne, K. J., Qin, J., Cavanagh, L. L., Smith, A., Hickey, M. J., and Weninger, W. (2012). Visualizing leukocyte trafficking in the living brain with 2-photon intravital microscopy. *Front Cell Neurosci*, 6:67. 61

[24] Precopio, M. L., Betts, M. R., Parrino, J., Price, D. A., Gostick, E., Ambrozak, D. R., Asher, T. E., Douek, D. C., Harari, A., Pantaleo, G., Bailer, R., Graham, B. S., Roederer, M., and Koup, R. A. (2007). Immunization with vaccinia virus induces polyfunctional and phenotypically distinctive CD8(+) T cell responses. *J. Exp. Med.*, 204(6):1405–1416. 15, 58

[25] Romero, P., Maryanski, J. L., Corradin, G., Nussenzweig, R. S., Nussenzweig, V., and Zavala, F. (1989). Cloned cytotoxic T cells recognize an epitope in the circumsporozoite protein and protect against malaria. *Nature*, 341(6240):323–326. 2, 15

[26] Sano, G., Hafalla, J. C., Morrot, A., Abe, R., Lafaille, J. J., and Zavala, F. (2001). Swift development of protective effector functions in naive CD8(+) T cells against malaria liver stages. *J. Exp. Med.*, 194(2):173–180. 5

[27] Schmidt, N. W., Butler, N. S., Badovinac, V. P., and 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):e1000998. 14

[28] Schmidt, N. W., Butler, N. S., and Harty, J. T. (2011). Plasmodium-host interactions directly influence the threshold of memory CD8 T cells required for protective immunity. *J. Immunol.*, 186(10):5873–5884. 17

[29] Schmidt, N. W., Podyminogin, R. L., Butler, N. S., Badovinac, V. P., Tucker, B. J., Bahjat, K. S., Lauer, P., Reyes-Sandoval, A., Hutchings, C. L., Moore, A. C., Gilbert, S. C., Hill, A. V., Bartholomay, L. C., and Harty, J. T. (2008). Memory CD8 T cell responses exceeding a large but definable threshold provide long-term immunity to malaria. *Proc. Natl. Acad. Sci. U.S.A.*, 105(37):14017–14022. iv, 15, 16, 17

[30] Solem, J. (2012). *Programming Computer Vision with Python: Tools and Algorithms for Analyzing Images*. Oreilly and Associate Series. O'Reilly Media, Incorporated. 27

[31] Spangrude, G. J., Sacchi, F., Hill, H. R., Van Epps, D. E., and Daynes, R. A. (1985). Inhibition of lymphocyte and neutrophil chemotaxis by pertussis toxin. *J. Immunol.*, 135(6):4135–4143. 58

[32] Sung, J. H., Zhang, H., Moseman, E. A., Alvarez, D., Iannacone, M., Henrickson, S. E., de la Torre, J. C., Groom, J. R., Luster, A. D., and von Andrian, U. H. (2012). Chemokine guidance of central memory T cells is critical for antiviral recall responses in lymph nodes. *Cell*, 150(6):1249–1263. 15

[33] Thiberge, S., Blazquez, S., Baldacci, P., Renaud, O., Shorte, S., Menard, R., and Amino, R. (2007). In vivo imaging of malaria parasites in the murine liver. *Nat Protoc*, 2(7):1811–1818. 3

[34] Waite, J. C., Leiner, I., Lauer, P., Rae, C. S., Barbet, G., Zheng, H., Portnoy, D. A., Pamer, E. G., and Dustin, M. L. (2011). Dynamic imaging of the effector immune response to listeria infection in vivo. *PLoS Pathog.*, 7(3):e1001326. 14

[35] Ward, S. G. and Marelli-Berg, F. M. (2009). Mechanisms of chemokine and antigen-dependent T-lymphocyte navigation. *Biochem. J.*, 418(1):13–27. 30

[36] Weiss, W. R., Sedegah, M., Beaudoin, R. L., Miller, L. H., and 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. iv, 16

Vita

Reka Katalin Kelemen was born in Budapest, Hungary, and moved to the United States at the age of eighteen to pursue her Bachelor of Science degree. She enrolled at Iowa State University in the newly formed undergraduate program of bioinformatics and computational biology, which she completed in August of 2011. During her undergraduate studies she participated in several research projects in her interdisciplinary field, one of which was organized by the National Institute for Mathematical and Biological Synthesis in Knoxville, Tennessee, and introduced her to mathematical modeling of biological systems. In early 2011 she applied to the Graduate School of Genome Science and Technology, a joint program by the University of Tennessee, Knoxville, and the Oak Ridge National Laboratories, and later accepted a graduate teaching assistantship there. During her graduate studies she gained experience teaching recitation sessions in general genetics, and continued learning about computational and mathematical analysis of biological data. In 2012 she joined the research group of Vitaly V. Ganusov, a theoretical immunologist in the Microbiology Department at the University of Tennessee, Knoxville, and she worked under his guidance on her thesis project for the Master of Science degree. In 2014 she accepted a PhD position at the Vienna Graduate School of Population Genetics in Austria, where she will continue her graduate studies.