Published in final edited form as:

J Appl Stat. 2018 ; 45(4): 697–713. doi:10.1080/02664763.2017.1290789.

Detecting T Cell Activation Using a Varying Dimension Bayesian Model

Zicheng Hu^{a,*}, Jessica N Lancaster^a, Lauren I R Ehrlich^{a,†}, and Peter Müller^{b,†}

^aDepartment of Molecular Biosciences, School of Natural Sciences, The University of Texas at Austin, Austin, TX, 78712, U.S.A

^bDepartment of Mathematics, School of Natural Sciences, The University of Texas at Austin, Austin, TX, 78712, U.S.A

Abstract

The detection of T cell activation is critical in many immunological assays. However, detecting T cell activation in live tissues remains a challenge due to highly noisy data. We developed a Bayesian probabilistic model to identify T cell activation based on calcium flux, a dramatic increase in intracellular calcium concentration that occurs during T cell activation. Because a T cell has unknown number of flux events, the implementation of posterior inference requires trans-dimensional posterior simulation. The model is able to detect calcium flux events at the single cell level from simulated data, as well as from noisy biological data.

Keywords

Bayesian; Indo-1; MCMC; Pseudo prior; T cell activation; Two photon microscopy

1. Introduction

In this study, we consider the problem of detecting T cell activation at the single cell level from noisy calcium concentration data. We developed model-based statistical inference to provide a full probabilistic description of the rise in intracellular calcium concentration during the T cell activation. The implementation involves a varying dimension parameter vector, which gives rise to a challenging trans-dimensional posterior simulation. The model allows accurate detection of T cell activation events based on calcium concentration data obtained using a fluorescent calcium indicator. It also provides inference for parameters that describe the calcium flux events during T cell activation. In summary, we provided a case

*Corresponding author. zicheng@utexas.edu.

†The two authors contributed equally to this work.

Disclosure Statement

The authors declare no potential conflict of interest.

Supplemental online material

Additional 'Supplemental online material' may be found in the on-line version of this article: 'Supplementary materials to: Detecting T Cell Activation Using a Varying Dimension Bayesian Model'. Section A of the material describes the experimental procedure of the motivating problem. Section B of the material describes the Gibbs sampler used to sample the posterior distribution.

study that illustrates how trans-dimensional posterior simulation can be implemented in a relatively straightforward fashion to address an important biological inference problem.

T cells play a central role in the adaptive immune system. Upon antigen recognition, T cells are activated to carry out a variety of immunological functions that result in pathogen clearance and the generation of immunologic memory. Detection of T cell activation is critical for clinical tests that monitor patients' immune systems [20, 22]. Detection of T cell activation events is also an important tool in biological research to query mechanisms of immune activation [15, 18].

Calcium flux, a dramatic increase in the intracellular Calcium concentration, is a hallmark of T cell activation [19]. Therefore, it is often used as a proxy for T cell activation in biological assays. Intracellular calcium can be measured with a widely used ratiometric fluorescence calcium indicator, Indo-1 [7]. The emission peak of Indo-1 shifts from 475 nm to 405nm upon Ca^{2+} binding when excited at 350nm. Thus, relative intracellular calcium concentrations can be calculated as the ratio of Indo-1 emission at 405 to emission at 475. Indo-1 is often successfully used in *in vitro* assays where the activation of cells is highly synchronized and the excitation and emission light is not scattered by surrounding tissues [4, 8].

Multi-photon microscopy enables researchers to image cells that are located deep within live tissue [5, 14]. This technology creates the opportunity to monitor T cell activation events *in situ*, but is accompanied by several challenges. Since both the excitation and emission light for Indo-1 travels through thick live tissues, the Indo-1 signals are noisy. In addition, unlike most *in vitro* assays, cells encountering endogenous antigens in live tissues do not undergo synchronized activation. Therefore, inference on T cell activation must be made for each single cell. Due to the high noise to signal ratio and the non-synchronized cell behavior, it is difficult to reliably identify calcium flux events with simple peak finding methods. Researchers currently rely on manual inspection of calcium data and correlation with other parameters, such as changes in cell velocity, to identify T cell activation events [1, 18].

Here, we developed a Bayesian probabilistic model to identify calcium flux events at the single cell level from noisy Indo-1 signal data. We introduced latent variables as indicators of calcium flux. Since the number of calcium flux events over the time course is unknown, we face an inference problem with a random number of parameters. Such a problem gives rise to challenging posterior simulation problems. We will implement a method known as pseudo prior mechanism [2] to handle the problem. The resultant model is able to detect calcium flux events accurately from simulated data, as well as from biological data from images of live tissues.

The rest of this report is structured as follows. In Section 2 we describe an example problem that motivated the development of the model. In Section 3 we describe the construction of the model. The sampling model, prior probability model, the pseudo prior mechanism and the posterior distribution are described in detail. In Section 4 we test the model with simulated data, as well as with biological data in which the initial time of calcium fluxes are

known. In Section 5 we provide an example of the model's application by analyzing an example dataset of the motivating problem.

2. A Motivating Problem

We first describe the biological problem that motivated development of the model. One specialized example of T cell activation that can be measured by calcium flux is activation of developing T cells during negative selection, a process by which developing T cells that are activated by endogenous self-antigens in the thymus undergo programmed cell death. This process is critical for eliminating auto-reactive T cells and preventing autoimmunity [12]. To understand the mechanism and outcome of T cell activation in the thymus, an experimental setup has been developed to allow visualization of T cell activation *in situ*. OT-I T cells, which express a transgenic T cell receptor with specificity for the ovalbumin antigen, are incubated with live thymic slices that express high levels of a secreted form of the ovalbumin antigen. Upon encountering the ovalbumin antigen in the thymic slice, OT-I thymocytes are first activated and then undergo programmed cell death in response to this model self-antigen [10]. Cultured thymic slices containing OT-I T cells were imaged by two-photon microscopy every 15 seconds for 15 minutes (Figure 1A). OT-I T cells were labeled with the calcium indicator dye Indo-1 before loading on the thymic slice, allowing us to measure the relative calcium concentration in each OT-I T cell using the fluorescence emission ratio of Indo-1. Figure 1B shows an example of the relative calcium concentration data from a single T cell in Figure 1A. See Supporting Information A for a detailed description of the experimental setup and the image processing procedure.

The experimental setup allows us to investigate how genetic defects or exogenously added compounds affect the activation of developing T cells. This requires not only the accurate detection of each calcium flux event from Indo-1 emission ratios, but also the estimation of parameters that describe the characteristics of the calcium flux, such as the duration of each flux event. Due to the high noise to signal ratio of the data, it is difficult to identify the calcium flux by manual inspection or by simple peak finding methods, as can be seen in Figure 1B. In addition, since each T cell encounters the endogenous antigens in a stochastic manner, the calcium flux of T cells is not synchronized. As a result, initial times of calcium flux need to be estimated at the single cell level. This problem motivated us to develop a full probabilistic model to automate the detection of calcium flux events in T cells.

3. Methods

3.1 Sampling model

The change of intracellular calcium concentration during T cell activation has been well characterized. Upon T cell activation, the calcium signal increases dramatically in a short period of time. After the T cell leaves the activation state, the intracellular calcium concentration decays exponentially [11].

Based on the characteristics of the T cell activation, we introduced latent variables to indicate the state (activation/ resting) of a T cell. In the activation state, the mean calcium concentration is modeled as a constant. In the resting state, the mean calcium concentration

is modeled as an exponential decrease (Figure 2A). In summary, mean calcium concentration can be expressed as

$$EY(t) = a + b \left(I_{(ST=A)} + I_{(ST=R)} e^{-c(t-T)} \right), \quad (1)$$

where EY is the mean calcium concentration, t is time, a is the baseline calcium concentration, b is the magnitude of the calcium flux upon T cell activation, c describes the rate of the exponential decrease to the resting state, and ST is in $\{A, R\}$ indicates the state of the T cell with $ST = A$ for the activation state and $ST = R$ for the resting state. Also, T is the time at which the T cell leaves the preceding activation state and I_A is the indicator function for event A .

Since a T cell can go through multiple T cell activation events, we concatenate the profiles of multiple activation events (Figure 2B).

$$EY(t) = a + b \sum_{i=1}^{N-1} I_{(T_i < t < T_{i+1})} \left(I_{(ST_i=A)} + I_{(ST_i=R)} e^{-c(t-T_i)} \right), \quad (2)$$

where N is the total number of states the T cell experienced during the time of data recording, $ST = (ST_i, i = 1, \dots, N)$ is now a vector of states, and T is a vector of the switch times. Since the number of activation states is unknown, N is a random variable that determines the dimension of the T and ST .

Note that T_1 may be negative when the beginning of the first state happens before the start of data recording. Similarly, T_N can be beyond the end of data recording when the end of the last state happens after the end of the data recording. For example, T_1 is negative in Figure 2B because the first state happens before the beginning of the data, and T_6 is beyond the last time point.

We complete the sampling model by assuming independent normal residuals

$$Y(t) = EY(t) + \varepsilon, \quad \varepsilon \sim N(0, SY). \quad (3)$$

Here SY is the standard deviation of the error term ε .

3.2 Priors

The varying number of parameters and the segmented nature of EY in equation (2) complicates inference using a maximum likelihood approach. This leads us to consider Bayesian inference. We assign normal and gamma priors for a , b , c , and SY^{-2} , using

$$a \sim N(E_a, S_a), \quad b \sim N(E_b, S_b), \quad c \sim \text{Ga}(a_c, b_c), \quad SY^{-2} \sim \text{Ga}(a_s, b_s). \quad (4)$$

The priors for a, b and SY^{-2} are chosen as (conditionally) conjugate priors for the normal sampling model.

We used T , the transition times between states to describe the sampling model (equation 2). However, it is more intuitive to assign prior distributions to the length of states dt_i (Figure 2B), with a straightforward mapping between the two. Note the special case of the first state, whose length is $dt_1 + dt_2$. We assume an exponential prior

$$dt_i \sim \text{Exp}(l_d) \quad (5)$$

Since the R and A states happen alternatingly, once the first state ST_1 is known, the rest states are known. We assume a Bernoulli prior distribution

$$P(ST_1=A)=p_a \quad (6)$$

In summary, $\phi = (E_a, S_a, E_b, S_b, a_c, b_c, a_s, b_s, l_d, p_a)$ are fixed hyperparameters that specify the prior distributions.

3.3 Pseudo prior and trans-dimensional MCMC

Because the number of calcium flux events for a T cell is unknown we face an inference problem with a random dimension parameter vector. Specifically, the dimension of $\mathbf{dt} = (dt_1, \dots, dt_N)$ changes with the number N of activation events. Several approaches have been proposed to implement posterior simulation for distributions with varying dimension parameter vectors. The two most popular methods are the pseudo prior method [2], and the reversible jump method [6]. Here, we choose the pseudo prior method due to its ease of implementation in this particular model.

We first fix a maximum dimension of $\mathbf{dt}$ to be $K - N$. Next, we assign independent exponential priors to dt_i , $i = 1, \dots, K$. All dt_i are included in the parameter vector. However, only the dt_i that falls within the time range of the data will actually feature in the likelihood function (equation 3). For example, in Figure 3A, dt_7 and dt_8 fall outside the time range of the data and are not used in the evaluation of the likelihood function. However, under another set of $\mathbf{dt}$ as in Figure 3B, dt_7 and dt_8 may be included in the likelihood evaluation. Importantly, in both cases the prior, and thus the posterior distribution includes dt_7 and dt_8 . Only in the earlier case they are hypothetical parameters that are not currently used in the likelihood evaluation. We refer to the prior $P(dt_7, dt_8, \dots, dt_K)$ in that case as a pseudo prior, that is, a prior probability model for parameters that are not needed for the likelihood evaluation. They are only included in anticipation of possible changes in N beyond 7. Mathematically, throughout the posterior probability model includes the full vector $\mathbf{dt} = (dt_1,$

..., dt_K), thus avoiding any complication related to variable dimension parameter vectors. As a result, posterior simulation can be implemented using any commonly used posterior simulation techniques, including, for example, Gibbs sampling.

3.4 Posterior distribution

For reference we state the joint posterior distribution. The likelihood function can be expressed as $P(Y|EY, SY) \propto \prod_{i=1}^n N(Y(t_i)|EY(t_i), SY)$, where n is the number of time points in the data, $N(y|\mu, \sigma)$ is a normal density with moments (μ, σ) evaluated at y , and EY is evaluated using equation (2). The joint prior is

$$P(a, b, c, ST, dt) = N(a|E_a, S_a) N(b|E_b, S_b) \text{Ga}(c|a_c, b_c) \prod_{k=1}^K \text{Exp}(dt_k|l_d) \text{Ga}(SY^{-2}|a_s, b_s) \text{Ber}(p_a)$$

where $\text{Ga}(y/a, b)$ and $\text{Exp}(y/\lambda)$ indicate the gamma and exponential density functions. $\text{Ber}(p)$ indicates the Bernoulli probability function. The joint posterior distribution is then simply the product

$$P(a, b, c, ST, dt|Y) \propto P(Y|EY, SY)P(a, b, c, ST, dt) \quad (7)$$

We implemented posterior Markov chain Monte Carlo simulation using a Gibbs sampler. See the section B of the Supplementary Online Material for details.

4. Model Testing

4.1 Testing the model using simulated data

To test the performance of the proposed Bayesian model, we simulated data based on the sampling model described in Section 3.1, and then carried out inference under the proposed Bayesian model using diffuse priors. Figure 4A is an example of the simulated data. T cells are activated at time points between 5–7.5 minutes and 12.5–15 minutes. Black dots in Figure 4B show the estimated mean of Y at each time point. Figure 4C shows the probability that the T cell is activated at each time point, which is defined as the fraction of iterations in which an activated state has been assigned to the time point in the MCMC sampling.

By thresholding the probability of activation, each time point can be classified into two states, activated or resting. The performance of the classifier can be evaluated by the receiver operator characteristic curve (ROC, Figure 4D) that plots sensitivity and specificity as a function of the threshold in the classification.

To test how noise affects the performance of the model, we simulated data with different noise to signal ratios, defined as $\frac{SY}{b}$. The data are then fitted with the Bayesian model and the ROC is generated as in Figure 4D. Area under the ROC (AUC) is calculated to evaluate performance of the model. We generated data with noise to signal ratios from 0.2 to 3. At each ratio, simulation and inference was done 50 times. Figure 4E shows the boxplot of 50 AUC at each noise to signal ratio. We found that the Bayesian model classifies the time points into an activated or resting state almost perfectly when the noise to signal ratio is lower than 0.8. Performance decreases gradually as the ratio increases. However, the

Bayesian model still maintains good performance (median AUC =0.8) even when the noise is 1.6 times higher than the signal.

4.2 Model testing using biological data

We then tested whether or not the Bayesian model is able to detect calcium flux from data acquired from live tissues. SIINFEKL, a peptide from the ovalbumin protein, can activate OT-I T cells in the thymus [9]. By introducing the activating peptide at a defined time, we were able to synchronize the timing of calcium flux. This allows us to test if our Bayesian model is able to estimate the initial time of calcium flux at the single cell level accurately.

OT-I thymocytes were first stained with the calcium indicator, Indo-1, and then incubated with live thymic slices. Thymic slices were imaged by two-photon microscopy at 15 second time intervals for a total of 15 minutes. Around time point 15 (3.75 minutes after initiating imaging) SIINFEKL peptide was added to the thymic slice. As expected, the mean calcium concentration of the OT-I T cell population first increased upon SIINFEKL addition and then gradually declined (Black dots in Figure 5A). Due to the synchronized behavior of calcium concentration, initial flux time of OT-I cells can be estimated easily at the population level (arrow in Figure 5A). However, signals from individual cells were highly noisy (Black circles in Figure 5A), making it difficult to estimate the flux time at the single cell level.

Inference under the proposed model was made based on calcium concentration data from each single cell. Figure 6A–C shows two examples of the inference. Probability of activation is calculated as in Figure 4C. We use the first time point in which the probability of activation is higher than a cutoff as an estimator of the initial calcium flux time. We estimated the initial calcium flux time using a range of cutoffs. Figure 6D shows the boxplot of estimated initial calcium flux times for each single cell using different cutoffs. The dashed line in Figure 6D shows the initial flux time estimated at the population level. We found that the estimation is robust with respect to the choice of cutoff. The estimations are most accurate when the cutoff is 0.5, suggesting that the estimated probability of activation is unbiased.

To benchmark the performance of our model, we compared our model with the wavelet thresholding method, which is widely used to remove noise from data [3]. After denoising the calcium concentration data with wavelet thresholding method, we used a hidden Markov model to estimate the probability of each time point being in the resting state, the activation state or the transition state between the activation state and the resting state. We then estimated the initial activation time point using different cutoffs (Figure 6E). See the section C of the Supplementary Online Material for the detailed description.

We found that the estimations from the wavelet thresholding method are most accurate when the cutoff is 0.8, suggesting that the estimated probability of activation is biased towards higher values. The standard deviation of the estimated initial activation time is larger in the wavelet method (SD=0.92 when cutoff = 0.8) compare to our model (SD=0.53 when cutoff=0.5). In summary, the estimation of the calcium flux time made by our method is less biased and more precise comparing to the wavelet thresholding method.

5. Results

We report posterior inference for datasets acquired from two experimental conditions of the motivating example. In one experimental condition, OT-I cells were loaded on RIP-OVA^{hi} thymic slices that endogenously express a secreted form of ovalbumin (OVA+); In the other experimental condition, OT-I cells were loaded on thymic slices that do not express ovalbumin (OVA−). We first used inference under the proposed Bayesian model to identify the calcium flux events in the motivating problem. Calcium concentration data from 44 single cells in the OVA+ dataset and 32 cells in the OVA− dataset were analyzed. Based on the results in Figure 6D, we used 0.5 as the cutoff to classify activation versus resting states. That is, time points that have a probability of activation above 0.5 are classified as in the activated state. Figure 7A shows an example from the OVA+ data set in which two calcium flux events were detected by the model. Figure 7B shows an example from the OVA− data set in which no calcium flux events were detected by the model.

Based on the inference under the Bayesian model, we were able to derive summary statistics that describe calcium flux events in the experimental settings of the motivating problem. We found that in the presence of the secreted ovalbumin antigen, 81% of the cells undergo calcium flux at least once during the data recording. 32% of the cells undergo 2 or 3 calcium flux events. The duration of calcium flux events varies from less than one minute to more than 6 minutes (Figure 7C). On average, OT-I cells undergo calcium flux 5.2 times per hour on RIP-OVA^{hi} slices and each flux event lasts 3.3 minutes. During the 15 minute time course, OT-I T cells had elevated calcium levels 28.8% of the time (Table 1). In the absence of Ovalbumin (OVA−), most (79%) OT-I T cells did not undergo calcium flux, as expected. Interestingly, 21% of the cells underwent calcium flux at least once, suggesting that non-ovalbumin endogenous self-antigens that activate the OT-I TCR are present in the thymus. However, most calcium flux events were transient, lasting less than 1 min (Figure 7D). This suggests that those non-ovalbumin antigens have a low affinity for the OT-I T cell receptor, consistent with the presence of positively-selecting ligands native to the thymic slice[12, 18]. On average, in response to endogenous non-ovalbumin antigens, OT-I thymocytes undergo calcium flux 2.6 times per hour and each flux event lasts 0.8 minutes. During the 15 minute time course, OT-I T cells underwent calcium fluxes 5.5% of the total time in OVA− slices (Table 1).

6. Discussion

In this study, we constructed a probabilistic model to identify T cell activation based on calcium concentration data. We used a latent variable vector (ST) to indicate the state (Activation or Resting) of a T cell. Based on the profile of calcium concentration during T cell activation, the mean calcium concentration is modeled as a constant at the activation state. At the resting state, the calcium is modeled as an exponential decay. We completed the sampling model by assuming independent normal residues. Since the number of activation events in the data for a given T cell during the imaging duration is not known, we used a pseudo-prior method to deal with the varying number of parameters. The performance of our model was tested using simulated data with different noise to signal ratios. We found that the model is able to detect T cell activation from highly noisy data (noise to signal ratio up to

1.6). We then tested our model using experimental data in which the initial activation time was known. We found that our model is able to identify the initial activation time point of T cells that are located in thick live tissue based on calcium concentration data. Finally, we provided an example of the model's application by using the model to detect thymocyte activation in the presence or absence of cognate antigens from data acquired in live mouse thymic sections.

Multiple methods have been proposed to detect the time of calcium influx [16, 17, 23]. However, these methods lack the ability to make inference for other biologically relevant parameters. For example, the magnitude and duration of calcium influx are important determinants of T cell function [12]. To our knowledge, our model is the first fully parameterized model proposed for longitudinal calcium concentration signal data. In addition to the accurate detection of calcium flux time, our model is able to make inference for other biologically relevant parameters, such as the magnitude of calcium influx, the duration of T cell activation and the kinetics of calcium concentration decrease after the T cell activation.

In our current model, non-informative priors are given to the base line a and flux amplitude b for each cell. This is due to the fact that the Indo-1 ratio depends on the microscope configuration of the experiment. The Indo-1 calcium ratio is calculated by dividing the Indo-1 calcium-bound emission signal at 405nm to the calcium-free emission signal at 475nm. The voltage settings for the photon multipliers detecting emission light at 405nm and 475nm can dramatically change the Indo-1 ratio. In addition, cell loading of the Indo-1 dye can vary among experiments, altering the dynamic range of ratios observed. Therefore no prior information of the base line and the flux amplitude is available before the inference. One pitfall of using non-informative prior for baseline is that the model is unable to tell the difference between an all-activation time course from an all-resting time course. If a T cell has been in the resting state for a long period, the Indo-1 ratio data of the cell will be a constant plus noise. Based on our current model, the data could be fitted as activation state throughout the time course, as well as fitted as resting state with the beginning of the resting state (T_1) being a large negative number. To solve this problem, a hierarchical model can be used. We can assume that each cell has its own baseline, and all the baselines in the same image follows a normal distribution centered at $\hat{a}$, which has a non-informative prior. This way, a constant signal will be fitted as an activation state if it is significantly higher than the estimated $\hat{a}$. Since the hierarchical model requires fitting the data for all cells at once, it requires much higher computational power and memory, making it unsuitable for analyzing large data sets. An alternative solution is doing a side-by-side control experiment in which all T cells stay at the resting state. The prior information of a can be extracted from the control experiment.

In early analyses of calcium signaling within thymic slices, T cells were sorted into activated or non-activated groups based on manual inspection of Indo-1 ratio traces. Activated patterns were manually identified based on the appearance of high magnitude calcium signal fluxes [1, 18]. In a previous study, calcium flux events of OT-I cells undergoing negative selection in the thymus were characterized. SIINFEKL peptide was exogenously added to the thymic slices, and 60% of cells were designated to have high calcium, with up to 80% designated to

high calcium when OVA-presentation was restricted to dendritic cells [13]. In WT thymic slices, which did not express OVA, 20% of T cells experienced high calcium. These results are similar to our own findings, which were obtained using Rip-OVA^{hi} thymic slices, demonstrating that the responses of OT-I T cells to a high affinity, secreted endogenous antigen are similar to their response when the same antigen is added exogenously. Another strength of our approach is that while the aforementioned two-photon imaging studies relied on signal thresholds to classify thymocytes as activated or not activated, our method does not require a priori knowledge of the calcium threshold. Thus, our Bayesian model of cell calcium flux improves the detection and throughput of T cell activation.

Calcium flux is often accompanied by decrease in migration velocity of the T cell. Therefore, previous studies often use both calcium and motility information to detect calcium flux [1, 18]. While our model only require calcium concentration data to detect calcium flux, analysis of cell motility can be included as a follow up step to further eliminate false positive errors.

Our model is constructed based on calcium flux patterns during T cell activation. When a T cell is activated, the intracellular calcium concentration increases dramatically, followed by a gradual exponential decay. Since calcium flux occurs in other types of cells, such as neuron cells, muscle cells and B cells, following the same general pattern [21], our model can be used to detect activation events in these cell types as well. Thus, the Bayesian model described can be easily adopted for many biological applications.

7. Software

Software in the form of R code, together with datasets containing the simulated data and data derived from live imaging and complete documentation is available on Github in <https://github.com/hzc363/BayesianFlux>.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding

This work was supported by the NIH/NIAID (R01AI104870; to L.I.R.E.).

References

1. Bhakta NR, Oh DY, Lewis RS. Calcium oscillations regulate thymocyte motility during positive selection in the three-dimensional thymic environment. *Nature immunology*. 2005; 6:143–151. [PubMed: 15654342]
2. Carlin BP, Chib S. Bayesian Model Choice via Markov Chain Monte Carlo Methods. *Journal of the Royal Statistical Society*. 2013; 57:473–484.
3. Donoho DL, Johnstone IM, Kerkycharian G, Picard D. Density estimation by wavelet thresholding. *Ann Statist*. 1996; 24:508–539.

4. Ehrlich LIR, Ebert PJR, Krummel MF, Weiss A, Davis MM. Dynamics of p56lck translocation to the T cell immunological synapse following agonist and antagonist stimulation. *Immunity*. 2002; 17:809–822. [PubMed: 12479826]
5. Ehrlich LIR, Oh DY, Weissman IL, Lewis RS. Differential Contribution of Chemotaxis and Substrate Restriction to Segregation of Immature and Mature Thymocytes. *Immunity*. 2009; 31:986–998. [PubMed: 19962328]
6. Green PJ. Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination. *Biometrika*. 1995; 82:711–732.
7. Grynkiewicz G, Poenie M, Tsien R. A new generation of Ca²⁺ indicators with greatly improved fluorescence properties. *The Journal of Biological Chemistry*. 1984; 260:344–3450.
8. Herold KC, Burton JB, Francois F, Poumian-ruiz E, Glandt M, Bluestone Ja. Activation of human T cells by FcR nonbinding anti-CD3 mAb, hOKT3 γ 1 (Ala-Ala). *The Journal of Clinical Investigation*. 2003; 111:409–418. [PubMed: 12569167]
9. Hogquist K, Jameson S, Heath W, Howard J, Bevan M, Carbone F. T cell receptor antagonist peptides induce positive selection. *Cell*. 1994; 76:17–27. [PubMed: 8287475]
10. Hubert FX, Kinkel Sa, Davey GM, Phipson B, Mueller SN, Liston A, Proietto AI, Cannon PZF, Forehan S, Smyth GK, Wu L, Goodnow CC, Carbone FR, Scott HS, Heath WR. Aire regulates the transfer of antigen from mTECs to dendritic cells for induction of thymic tolerance. *Blood*. 2011; 118:2462–72. [PubMed: 21505196]
11. Huse M, Klein LO, Girvin AT, Faraj JM, Li QJ, Kuhns MS, Davis MM. Spatial and Temporal Dynamics of T Cell Receptor Signaling with a Photoactivatable Agonist. *Immunity*. 2007; 27:76–88. [PubMed: 17629516]
12. Klein L, Kyewski B, Allen PM, Hogquist Ka. Positive and negative selection of the T cell repertoire: what thymocytes see (and don't see). *Nature reviews Immunology*. 2014; 14:377–91.
13. Melichar H, Ross JO, Herzmark P, Hogquist Ka, Robey Ea. Distinct temporal pattern of T cell receptor signals during positive versus negative selection in situ. *Science Signaling*. 2004; 6:1–33.
14. Mempel TR, Henrickson SE, Von Andrian UH. T-cell priming by dendritic cells in lymph nodes occurs in three distinct phases. *Nature*. 2004; 427:154–9. [PubMed: 14712275]
15. Pesic M, Bartholomäus I, Kyratsous NI, Heissmeyer V, Wekerle H, Kawakami N. 2-photon imaging of phagocyte-mediated T cell activation in the CNS. *J Clin Invest*. 2013; 123:1192–1201. [PubMed: 23454769]
16. Pnevmatikakis EA., Merel J., Pakman, A., Paninski, L. 2013 Asilomar Conference on Signals, Systems and Computers. IEEE; Nov. 2013 Bayesian spike inference from calcium imaging data; p. 349-353. Available at <http://ieeexplore.ieee.org/document/6810293/>
17. Pnevmatikakis EA, Soudry D, Gao Y, Machado TA, Merel J, Pfau D, Reardon T, Mu Y, Lacefield C, Yang W, Ahrens M, Bruno R, Jessell TM, Peterka DS, Yuste R, Paninski L. Simultaneous Denoising, Deconvolution, and Demixing of Calcium Imaging Data. *Neuron*. 2016; 89:299. Available at <http://linkinghub.elsevier.com/retrieve/pii/S0896627315010843>.
18. Ross JO, Melichar HJ, Au-Yeung BB, Herzmark P, Weiss A, Robey Ea. Distinct phases in the positive selection of CD8+ T cells distinguished by intrathymic migration and T-cell receptor signaling patterns. *Proceedings of the National Academy of Sciences of the United States of America*. 2014; 111:E2550–8. [PubMed: 24927565]
19. Smith-Garvin JE, Koretzky Ga, Jordan MS. T cell activation. *Annual review of immunology*. 2009; 27:591–619.
20. Stegeman, Ca, Tervaert, JW., Huitema, MG., Kallenberg, CG. Serum markers of T cell activation in relapses of Wegener's granulomatosis. *Clinical and experimental immunology*. 1993; 91:415–420. [PubMed: 8443965]
21. Takahashi A, Camacho P, Lechleiter JD, Herman B. Measurement of Intracellular Calcium. *Physiological Reviews*. 1999; 79:1089–1125. [PubMed: 10508230]
22. Tsuchiya T, Ohshima K, Karube K, Yamaguchi T, Suefuji H, Hamasaki M, Kawasaki C, Suzumiya J, Tomonaga M, Kikuchi M. Th1, Th2, and activated T-cell marker and clinical prognosis in peripheral T-cell lymphoma, unspecified: Comparison with AILD, ALCL, lymphoblastic lymphoma, and ATLL. *Blood*. 2004; 103:236–241. [PubMed: 12958063]

23. Vogelstein JT, Watson BO, Packer AM, Yuste R, Jedynak B, Paninskik L. Spike inference from calcium imaging using sequential Monte Carlo methods. *Biophysical Journal*. 2009; 97:636–655. [PubMed: 19619479]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 1.

Data of the motivating problem. Live thymic slices were imaged using a two photon microscope every 15 seconds for 15 mins. (A) The image of a single time point. T cells are labeled with Indo-1. The Indo-1 signal is color coded with white indicating high Calcium concentration and purple indicating low Calcium concentration. Scale bar represents $20\mu\text{m}$. The color-coded cell track shows the path of T cell movement with purple indicating earlier time points and yellow indicating later time points. (B) An example of calcium concentration (measured as Indo-1 signal ratio) data from a single T cell.

Figure 2.

A schematic diagram showing the sampling model. (A) Sampling model for a single T cell activation event. The black curve represents the mean of the calcium concentration(EY). The EY is modeled as a constant in activation state (state = A), and an exponential decrease in the resting state (state = R). a represents the baseline of the data. b represents the amplitude of calcium increase once the T cell is activated. c represents the rate of decrease in the resting state. T represent the transition time between states (B) Sampling model for multiple T cell activation events. Notations are the same as (A). dt_i represent the duration of each state. Gray dots represent calcium concentration data.

Figure 3.

Schematic diagrams showing the implementation of the pseudo prior method in the proposed Bayesian model. A fixed number of K dt_i are included in the Bayesian model. K is greater than the maximum number of T cell activation events a T cell can have in the time range of the data. The prior of all dt_i is included in the joint posterior distribution even when some of them do not contribute to the evaluation of the likelihood. (A) An example scenario in which a set of dt is proposed in the Gibbs sampling. Only dt_1 to dt_6 falls within the time span of the data and will contribute to the likelihood evaluation. (B) Another example scenario in which a different set of dt is proposed. dt_1 to dt_8 falls within the time range and will contribute to the likelihood.

Figure 4.

Results of model testing using the simulated data. (A) An example of data simulated based on the sampling model with a noise to signal ratio of 0.5. (B) Data in (A) was fitted by the Bayesian model. Black dots represent the estimated mean calcium concentration (EY) at each time point. (C) Estimated probability that the cell is in an activation state at each time point. The probability is estimated using the ratio between the number of MCMC iterations in which the time point is classified to be in activation state and the total number of MCMC iterations. (D) Based on the probability in (C), each time point can be classified into activation or resting states. The performance of the classifier is characterized by a receiver operator characteristic (ROC) curve. (E) Simulation and inference was done 50 times at each noise to signal ratio from 0.2 to 3. Area under the curve (AUC) of ROC curve is calculated after each simulation study. Each Boxplot represents 50 AUC at each noise to signal ratio.

Figure 5.

An overview of the biological data used to test the model. Indo-1 stained OT-I T cells were loaded on thymic slices. Thymic slices were then imaged every 15 seconds for 15 mins. At time point 15 (3.75 minutes after the start of the imaging), SIINFEKL is added to the thymic slice to activate the OT-I T cells. (A) The calcium concentration data derived from the images. Black dots represent the average calcium concentration of all T cells in the image at each time point. Black circles represent the calcium concentration in each single T cell at each time point. The black arrow indicates the estimated initial flux time at the population level. (B) Images of a thymic slice before and after SIINFEKL addition. Indo-1 signal is

color coded, with white indicating high calcium concentration and purple indicating low calcium concentration. The color-coded line shows the path of T cell movement with purple indicating earlier time points and yellow indicating later time points.

Figure 6.

Results of model testing using biological data. The Bayesian model is used to make inference on the biological data described in Fig 5. (A) Two examples of calcium concentration data fitted by the Bayesian model. Black dots represent the estimated mean calcium concentration. Black circles represent the data. (B) Probability of activation estimated by the model from the data above. (C) The estimated number of activation states occurring within the data range. *dt* is sampled from the posterior. The number of activation states is calculated based on *dt*. (D) Boxplots showing the estimated initial calcium flux time of each cell at different cutoffs. The first time points which have a probability of activation above the cutoff is used to estimate the initial flux time. The dashed line represents the initial calcium flux time point estimated at the population level. (E) Boxplots showing the estimated initial calcium flux time of each cell at different cutoffs using wavelet thresholding and hidden Markov model. The first time points which have a probability of activation above the cutoff is used to estimate the initial flux time. The dashed line represents the initial calcium flux time point estimated at the population level.

Figure 7.

The analysis of the motivating example using the model. (A–B) Examples of the calcium concentration data and estimated probability of activation of single OT-I thymocytes in OVA + (A) and OVA– (B) thymic slice datasets. Black circles represent the data. Black dots represent the estimated mean calcium concentration at each time point. Black arrows in (A) indicate the initial time of calcium flux events. (C–D) Histograms showing the number of calcium flux events detected from each cell and the length of each flux events in the OVA+ (C) and OVA– (D) datasets.

Table 1

A summary of T cell activation events detected from the motivating example data sets.

	OVA– slice	OVA+ slice
Total number of cells analyzed	32	44
Total number of time points analyzed	1302	2430
Total number of time points undergoing calcium flux	72	701
Total number calcium flux events	14	53
Percent of time undergone calcium flux	5.5%	28.8%
Percent of cells undergone calcium flux	81.1%	21.3%
Frequency of calcium flux	2.6/hour	5.2/hour
Average length of calcium flux	0.8 minutes	3.3 minutes