Quantifying naive T cell dynamics in mice using a novel fluorescent division reporter system

Minor research project for the Master of Bioinformatics and Biocomplexity at Utrecht University

Work of Eva Lukas along with Thea Hogan, Benedict Seddon and Andrew J. Yates

Daily supervisor: Prof. Andrew Yates
Examiner: Prof. Rob de Boer

Abstract

Lymphocyte dynamics and their potential to self-renew post maturation have frequently been studied using tracking techniques like heavy isotopes. Label is incorporated into newly synthesized DNA during treatment and lost upon cell death. We here introduce a novel method of labeling using yellow fluorescent protein. Upon tamoxifen administration, cells undergoing division as marked by Ki67 transcription induce expression of fluorescence. In contrast to isotope labeling, offspring of fluorescent cells will also express YFP, eliminating dilution effects. We show two models describing either YFP data only or YFP and Ki67 data collectively. We conclude estimates for the lifetime of peripheral naive T cells, loss rate of Ki67 expression and of possible self-renewal. Comparison with previously reported estimates support our hypothesis that YFP labeling is a reliable system for modeling the kinetics of lymphocyte populations.

Lay summary

The dynamics of cells, namely their proliferation or turnover rate, are commonly studied and quantified using a labeling agent. If a cell divides in a certain period, DNA synthesis enables the incorporation of the agent, e.g. deuterium. One disadvantage of these assays is the need to account for dilution of the reporter through division in the follow-up phase. Here we explore a new reporter mouse that induces the permanent expression of yellow fluorescent protein (YFP) in dividing cells while the drug tamoxifen is administered. YFP labeling is inherited by cells' offspring following division, avoiding the need to account for dilution effects. We use this system to study the dynamics of naive CD4 and CD8 T cells in mice, and compare its predictions with those from other studies. We also explore a variety of methods of inference in an attempt to minimise the effect of mouse-to-mouse variation in labelling dynamics and measured cell numbers. Our estimates of kinetic parameters are comparable to previously reported values and allow us to conclude that the system of YFP tagging is suitable for the investigation of T cell dynamics.

Introduction

Arising from hematopoietic stem cells in the bone marrow, precursors of T cells migrate to the thymus starting off neither expressing TCR, CD4 or CD8 (double negative thymocytes). Positive and negative selection follow, resulting in mature single-positive CD4 (mSP4) and CD8 (mSP8) thymocytes which are exported to circulation. Naive T cells are awaiting pathogen exposure in the periphery. The lack of activation through antigen encounter is assumed to result in cell decay within a few months. Upon pathogen recognition, effector T cells may arise, possibly eventually differentiating into memory T cells [1]. This study focuses on the quantification of kinetics of naive T cells.

A consensus has emerged that naive CD4 and CD8 T cell numbers in mice are sustained predominantly by the thymus, with rates of proliferative self-renewal so low, such that the majority of naive T cells will never divide [2–5], while proliferation of naive T cells is more appreciable in humans [3]. Naive T cell numbers can be sustained into old age in the face of natural waning of influx from the thymus [4, 6, 7] and even following thymectomy [3, 8], but in either setting there is no clear evidence for any compensatory increase in cell division [5, 7, 9]. Instead, it appears that slow increases in the propensity of naive T cells to survive with their post-thymic age, either through selective or adaptive mechanisms, help to sustain their numbers long-term in both mice and humans [5, 10–12].

Many studies of lymphocyte dynamics use DNA labelling methods, in which an identifiable agent such as bromodeoxyuridine (BrdU), or deuterium derived from deuterated glucose or heavy water, is taken up by dividing cells. Tracking the increase in frequency of labeled cells during a period of label administration, and its decline afterwards, allows one to measure rates of division and turnover (loss) within that population [13]. However, one issue with these assays is that the loss of label derives not only from the influx of unlabeled cells from a precursor population, but also from dilution of the label by cell division.

In this study, we introduce a novel mouse model which uses heritable stamping of cells to report both historical and ongoing cell division events. In this mouse model, any cells that express Ki67 during a period of administration of Tamoxifen are induced to permanently express yellow fluorescent protein (YFP), which is inherited across further cell divisions (Figure 1). Ki67 is a nuclear protein that is expressed during the cell cycle for 3-4 days after mitosis [4, 14, 15]. At subsequent times, cells expressing YFP are the descendents of cells that divided during the window of tamoxifen treatment. The YFP reporter therefore acts in an analogous way to pulse-chase DNA labeling assays, although without dilution of label within cells in the chase period. These reporter mice also allow a live readout of cell division, by conjugating a red fluorescent protein (mCherry) to Ki67 (Figure 1).

The aims in this study were two-fold. First, we wanted to assess the utility of this division reporter system for modeling cellular dynamics. To do this we focused on validating the system by modeling naive CD4 and CD8 T cell dynamics in mice, since the behaviour of these populations have been characterized extensively in earlier research. Second, we wanted to explore how to maximize the information gained from YFP and Ki67 labelling timecourses, by implementing different strategies for dealing with potential sources of noise and bias in the data.

Results

Estimates of mean lifespans of naive T cells using reporter mice align with previous studies

We studied the dynamics of YFP^+ T cells in 61 mice aged between 42 and 170 days that underwent 5 days of tamoxifen treatment (Figure S.1). Mice were sacrificed at a range of timepoints across the following 90 days, and T cell subsets were harvested from thymi, lymph nodes and spleens. We then measured numbers of mature single positive CD4 (mSP4) and CD8 (mSP8) thymocytes, and of naive ($CD44^{low}$ $CD62L^{high}$) CD4 and CD8 T cells pooled from spleen and lymph nodes, and the expression of YFP and Ki67-mCherry within each of these populations (Figure S.2). Numbers and Ki67 expression within these subsets in the reporter mice were indistinguishable from age-matched wild-type controls (not shown), indicating that tamoxifen treatment and YFP induction had no detectable effect on naive T cell dynamics.

Previous studies have shown that the expected lifespans of naive CD4 and CD8 T cells increase slowly with their post-thymic age [5, 10, 11], and hence also with host age. However in this experiment the ages of mice at treatment were distributed across a wide range, and non-sequentially with respect to age at sacrifice. We therefore did not expect to see a strong signal of cell-age effects in this experiment, and as-

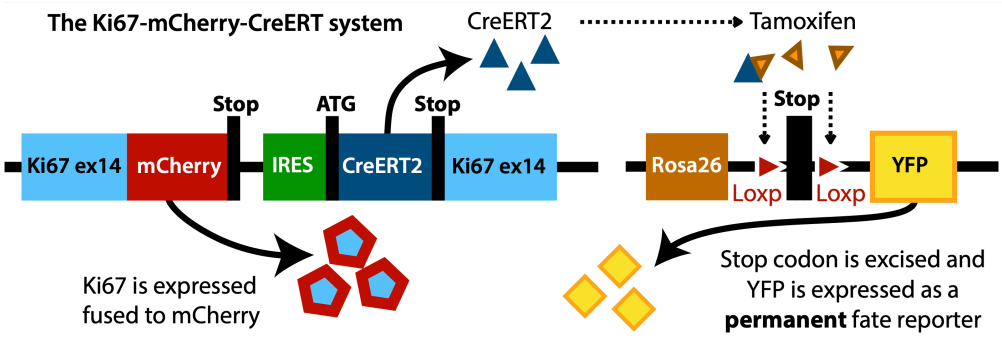


Figure 1: The Ki67 reporter system. (A) Scheme of the mouse model that allows us to indelibly label dividing cells and their progeny with yellow fluorescent protein (YFP) and follow recent cell division events with a fluorescent Ki67 reporter (mCherry). Expression of mCherry and Rosa26 EYFP reporter in total thymus, following four daily injections of tamoxifen.

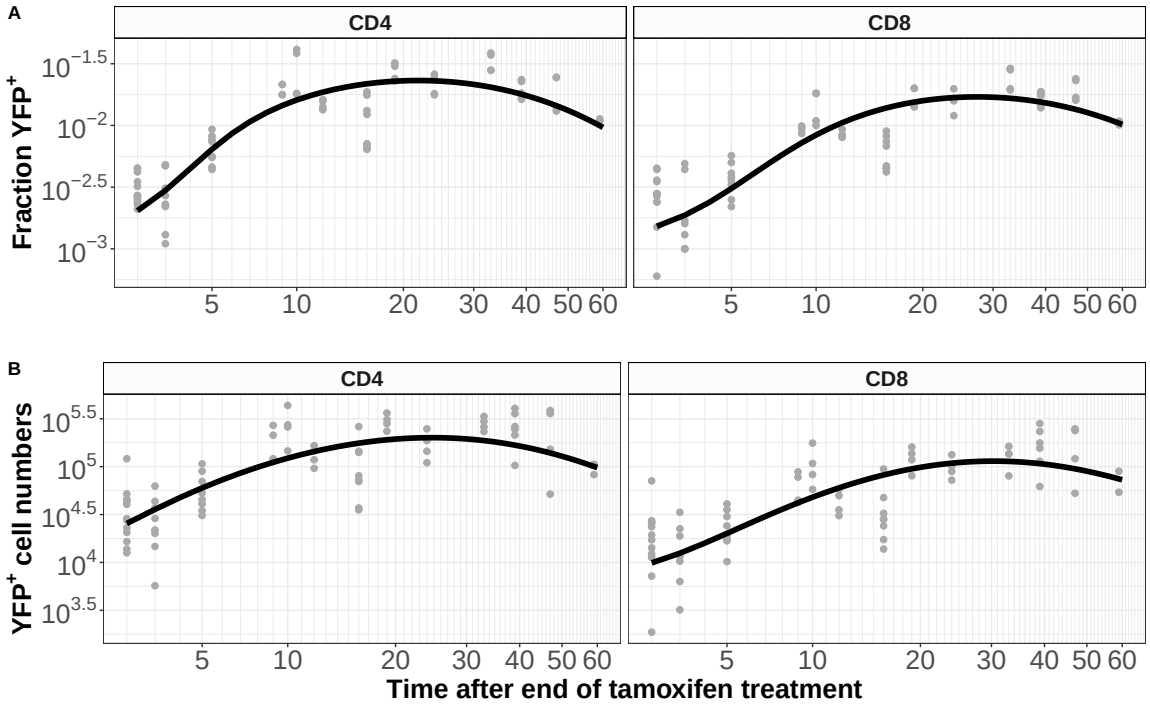


Figure 2: A simple homogeneous model. Best fits for Equation 1 (**A**) describing the time-variation of the numbers of YFP$^+$ naive T cells and Equation 2 (**B**), representing the proportion of naive cells that are YFP-expressing. Estimated parameters are summarized in Table A.1.

sumed instead that naive CD4 or CD8 T cells could both be treated as homogeneous with respect to their rates of loss through death or onward differentiation, μ. We constructed a model describing the kinetics of labeled cells within the thymic- and peripheral naive T cell populations, stratifying each by Ki67 expression (illustrated in Figure 3).

The consensus view is that naive T cells self-renew rarely, if at all [2–4, 7, 10]. We therefore began by exploring a reduced model in which the division rate p was set to zero. In this model, any YFP expression within the naive population derives from cells that divided in the thymus during tamoxifen treatment and were subsequently exported. We assume that mSPs leave the thymus at a constant *per capita* rate Θ, motivated by the observation that the total rate of thymic output scales linearly with thymocyte numbers [16]. Labeled naive cells (L) are lost at a *per capita* rate μ through death or differentiation. μ

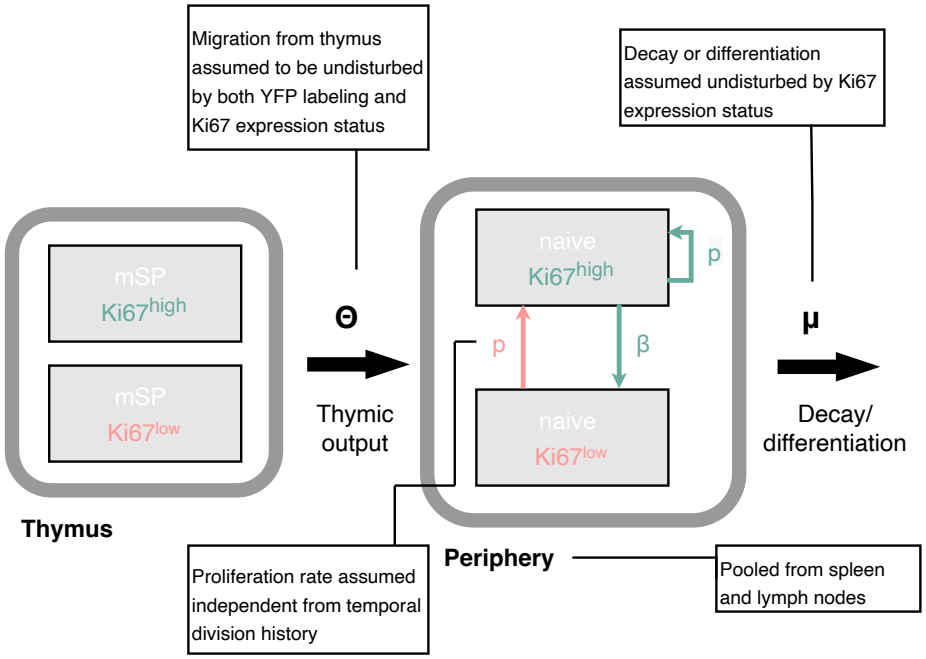


Figure 3: A simple homeogeneous model of naive T cell dynamics. New Ki67$^{high/+}$ and Ki67$^{low/-}$ naive cells are exported from the thymus at a rate assumed to be proportional to the numbers of Ki67$^{+/-}$ mSPs. Once exported, naive T cells are lost at rate μ and divide at rate p. Division is accompanied by induction of Ki67 expression which persist for a mean time $1/\beta$ before cells return to Ki67^{-} state. See Equation 4 and Equation 5.

is assumed to be identical for labeled and unlabeled cells. If the numbers of YFP^{+} mSP cells over time is $Y(t)$, then

$$\frac{dL}{dt} = \Theta Y(t) - \mu L. \tag{1}$$

Possibly partially due to low efficiency of YFP induction, the measured numbers of labeled cells were somewhat noisy (Figure 2A, gray points). We hypothesize that subset frequencies of label may be subject to smaller variation as compared to absolute counts of cells. We therefore also explored an alternative approach, modeling the kinetics of the proportion of naive cells that expressed YFP, $\ell = L/N$, where N is the total number of naive CD4 or CD8 T cells (Figure 2B, gray points). We observed that total naive T cell numbers showed very little variation across the experiment (Figure S.2), and so we could reasonably assume that they were at or close to steady state ($dN/dt = 0$). In that case, Equation 1 can be recast as

$$\frac{d\ell}{dt} = \frac{\Theta Y(t)}{N} - \mu\ell = \Theta Z(t) - \mu\ell, \tag{2}$$

where $Z(t)$ is the number of YFP^{+} cells within the mSP compartment normalized by the number of peripheral naive T cells.

The process of induction and loss of YFP^{+} cells in the thymus is a complex dynamic occurring across a chain of proliferating precursors, back to lymphoid progenitor cells in the bone marrow. Rather than modeling this process explicitly, we described the timecourses of total YFP^{+} mSP numbers ($Y(t)$) and the frequency of YFP^{+} cells within mSP cells over the number of naive T cells ($Z(t)$) using the following empirical functions:

$$Y(t) = Y_0 - at^b e^{-ct}, \tag{3a}$$

$$Z(t) = ke^{at}\left(1 + e^{v(t-\tau)}\right)^{-\frac{a+b}{v}}. \tag{3b}$$

For mSP4 and mSP8s, Equation 3a and Equation 3b were fitted to their appropriate timecourses, using least squares and log- and logit-transformed YFP^{+} cell numbers and frequencies, respectively (Figure S.3). These fitted functions were then used in Equation 1 and Equation 2, which in turn were solved numerically and fitted to the observed timecourses of numbers and frequencies of YFP^{+} naive T cells, respectively (Figure 2; see p.10 for details).

The two methods yielded similar estimates of the mean residence times (μ^{-1}) with comparable levels of uncertainty (Figure 4, center panels, blue points). Residence time estimates for both CD4 and CD8 T

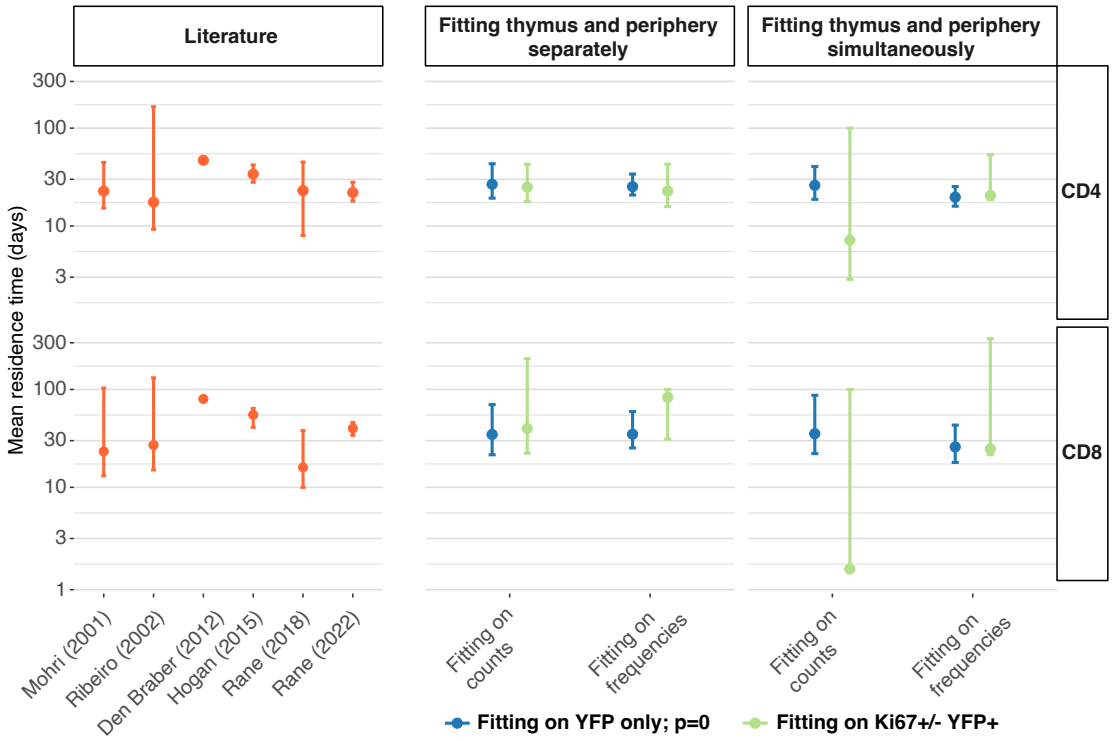


Figure 4: Summary of estimates of mean residence times of naive CD4 and CD8 T cells in adult mice, derived from different methods. **Left panels**; summary of published estimates. **Center panels**; estimates derived from Ki67 reporter mice, using pre-determined, fitted empirical descriptions of the dynamics of labeled thymocytes as model inputs. **Right panels**; estimates derived from simultaneous fitting of the kinetics of labelled thymocytes and labeled naive T cells. In the center and right panels, blue points show estimates derived from the simpler model that assumes no proliferation among naive T cells; green points show estimates from the larger model that explicitly tracks Ki67 within thymocytes and naive T cells and allows for proliferation. Bars indicate 95% confidence intervals.

cells agreed well with those previously derived from BrdU or deuterium labeling [3, 17, 18] and from busulfan chimeric mice [4, 5, 10], shown for reference in Figure 4 (left panels). The cell count and cell frequency methods also generated very similar estimates of the *per capita* rate of output of cells from the thymus (Θ; Table A.1). We conclude that a simple homogeneous model without cell proliferation, and with pre-determined empirical functions describing the influx of labeled cells from the thymus, yields residence times of naive T cells consistent with those derived from other labeling methods. We also conclude that fitting on either the total numbers of YFP$^+$ naive T cells, or their frequency, yields similar point estimates with comparable precision.

Defining rates of naive T cell proliferation by combining YFP labeling with live readouts of Ki67 expression

Expression of YFP indicates that a cell or its ancestor divided during tamoxifen treatment, but does not tell us about a cell's subsequent propensity for division. Since Ki67 persists in the nucleus for several days following cell division [4, 14, 15], its expression among naive T cells could derive from cells that divided in the thymus and were recently exported, and/or from post-thymic self-renewal. Incorporating the dynamics of Ki67 expression in the models of YFP$^+$ cell dynamics might then allow us to distinguish these processes and test the consensus that naive T cells divide rarely, if at all, in mice.

Before extending the model we examined the kinetics of Ki67 expression within the fluorecent subpopulations and within the general populations of mSP or naive phenotype (Figure 5). At early timepoints we see that YFP$^+$ naive T cells (yellow points) are significantly enriched for Ki67$^+$ cells relative to the population as a whole (grey). The observation is consistent with the intuitive result that the YFP$^+$ population is predominantly comprised of recent thymic emigrants (RTEs), many of which divided recently in the thymus, where levels of proliferation are high (Figure S.2). At later time points we see that Ki67

expression among naive YFP$^+$ cells approaches the population average. This is counterintuitive at first. If naive T cells divide rarely, we would expect all naive YFP$^+$ cells to be mature at late times, and to have lost any Ki67, initially derived from intrathymic division. In contrast, we expect substantial numbers of unlabeled RTEs, with high levels of inherited Ki67 after the treatment window is closed. Therefore, at later time points, we would expect the proportion of YFP$^+$ cells that express Ki67 to be lower than the population average. The convergence in frequencies we however observe (Figure 5), has two potential explanations. One is that mature naive T cells do indeed divide at significant levels. The other is that YFP$^+$ cells continue to be exported from the thymus long after tamoxifen treatment, perhaps due to labeling of stem-like precursor cells in the bone marrow. Indeed we did observe low but significant levels of label within the mSP4 and mSP8 populations more than two months after tamoxifen administration ended (Figure S.3, points), which were captured by the functions we used to define these kinetics (Figure S.3, black lines).

To estimate rates of division of naive T cells, we expanded the basic model (Equation 1) to track Ki67^{low} and Ki67^{high} cells (L^- and L^+, respectively; Equation 4). Division induces Ki67 expression, and the return to the Ki67^{low} state is assumed to occur with first order kinetics at rate β. Thus β^{-1} represents the mean residence time in the Ki67-expressing state, which is determined by both the intrinsic decay rate of the protein and the threshold of expression used in the flow cytometry defining the boundary between the high and low states. As before, we assume that YFP-labeling has no effect on cell dynamics, with all cells dividing and decaying at equal rates (p and μ respectively; see Figure 3).

$$\begin{aligned}\frac{dL^+}{dt} &= \Theta K^+(t) - L^+(\mu+\beta) + p(2L^- + L^+),\\ \frac{dL^-}{dt} &= \Theta K^-(t) - L^-(\mu+p) + \beta L^+,\end{aligned} \tag{4}$$

The empirical functions $K^{+/-}(t)$ describe the timecourses of the numbers of YFP$^+$Ki67$^{high/low}$ mSP cells, respectively. The fractions of cells expressing Ki67 within YFP$^+$ mSP4 and mSP8 cells were roughly constant (Figure S.4). If we denote this constant f^+, we define $K^+(t) = Y(t)f^+$ and $K^-(t) = Y(t)(1-f^+)$ (shown in Figure S.6), where $Y(t)$ is the same function as used above for the timecourse of YFP$^+$ mSP4 or mSP8 cells (Equation 3a).

We also recast these dynamics in terms of cell frequencies, with ℓ^+ denoting the frequency of Ki67^{high} cells within the YFP$^+$ population, now called N_{YFP};

$$\begin{aligned}\frac{d\ell^+}{dt} &= \frac{\Theta}{N_{YFP}}\left(K^+(t) - Y(t)\ell^+\right) - \beta\ell^+ + 2p(1-\ell^+),\\ \frac{dN_{YFP}}{dt} &= \Theta Y(t) + N_{YFP}(p-\mu).\end{aligned} \tag{5}$$

We then took a similar approach to the one described above, simultaneously fitting L^+ and L^- (defined by Equation 4), or ℓ^+ and N_{YFP} (Equation 5) using log-transformation, to their appropriate timecourses (fits shown in Figure 6a and Figure 6b respectively).

The estimated residence times agreed with previous studies and with our estimates derived from the simpler model neglecting possible proliferation (Figure 4; central panels, green points/bar). Our point estimates of interdivision times were significantly higher than the lifetime of naive T cells (Table A.2) (see our residence time estimates, meta-literature suggests a maximum of three months [1]), supporting the consensus that naive T cells in mice rarely divide [2–4, 7, 10]. As expected, estimating the extra parameter p slightly increased the uncertainty in the mean residence time (Figure 4; central panels, compare blue and green confidence intervals). We also found no difference in the estimated residence times using the two methods above, suggesting that the signal to noise ratios involved in the two methods of using the experimental data were similar.

Both methods yielded estimates of the Ki67 lifetime (β^{-1}) that were broadly consistent with previous studies. Fitting on YFP$^+$Ki67$^{high/low}$ counts (Equation 4) yielded a lifetime of 1.6d (95% CI: 1.0 – 2.3) for naive CD4 T cells, and 3.4d (2.1 – 4.8) for naive CD8 T cells. Fitting using the Ki67$^+$ frequency (Equation 5) yielded similar estimates (1.9d (1.4 – 2.7) and 3.8d (2.5 – 5.4) for CD4 and CD8, respectively). All parameters are shown in Table A.2.

Allowing for uncertainty in the label content in the thymus

The approaches above assume that the timecourses of label availability in the thymus are described exactly by the fitted functions $Y(t)$ and $Z(t)$. Neglecting uncertainty in these functions will likely overestimate our confidence in the parameters describing naive T cell dynamics. To address this issue, we fitted the labeling data from the thymus and periphery simultaneously. To do this we formed a simple

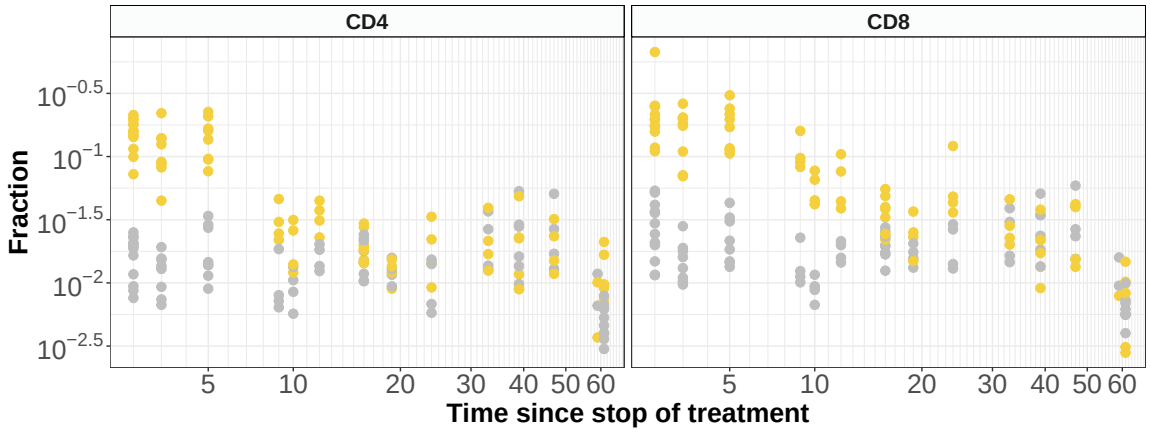


Figure 5: Frequencies of Ki67-expressing naive CD4 and CD8 T cells over time since tamoxifen treatment, among YFP^+ cells (yellow) and the population as a whole (gray). Unexpectedly, Ki67 expression fractions converge at later time points. This may be explained by stem-like precursor YFP labeling feeding labeled RTEs into the peripheral population at later timepoints.

multiplicative joint likelihood for these timecourses, motivated by the lack of correlations between them (Figure S.5).

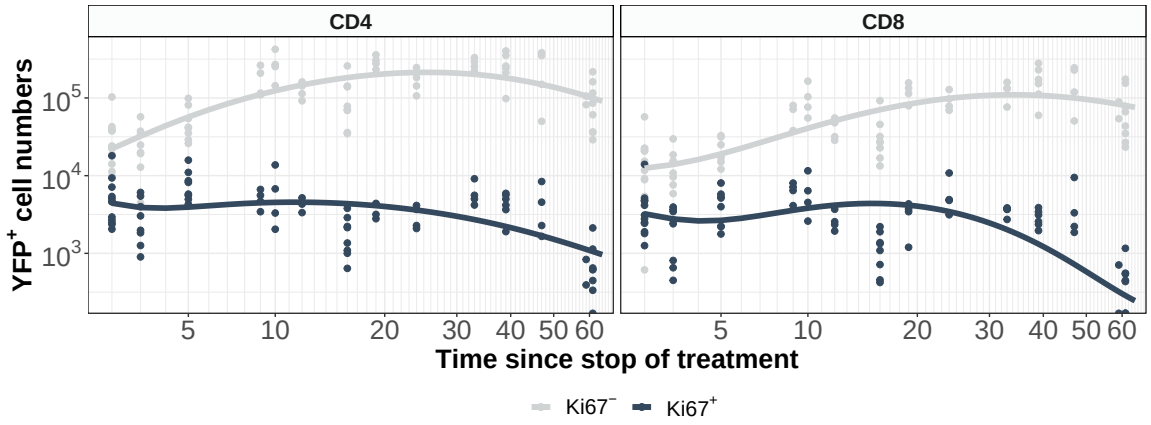
We refitted Equation 1 along with Equation 3a simultaneously (dark data and model show fluorescence in the periphery and lighter dots and lines show thymus occurences ($Y(t)$) and thymic/naive ratio ($Z(t)$) in Figure 7). Residuals were normally distributed for data for CD8 only (Shapiro-Wilk p-value=0.267 vs. 0.019 for CD4). Parameters and confidence intervals are comparable to results obtained from separate fitting procedures (see Table A.1 and Figure A.4).

While we allow for more freedom for the parameter estimation and avoid imposing fitting inaccuracies in the thymus onto the periphery, resulting residence times remain comparable to separate fitting procedures and values reported in literature (Figure 4). The same process was performed for the frequency of YFP in the peripheral cell population (Equation 2) and the cell count of fluorescent mSPs (Equation 3a). Again, residuals for CD4 data were not normally distributed (Shapiro-Wilk p-value=0.027), while they are for CD8 data (p-value=0.870). We find parameters for the thymic output (Θ) and for the cells' decay (μ) to be slightly higher when fitting both datasets simultaneously (see Figure A.4). Residence times are estimated within the same range as mentioned earlier (Figure 4).

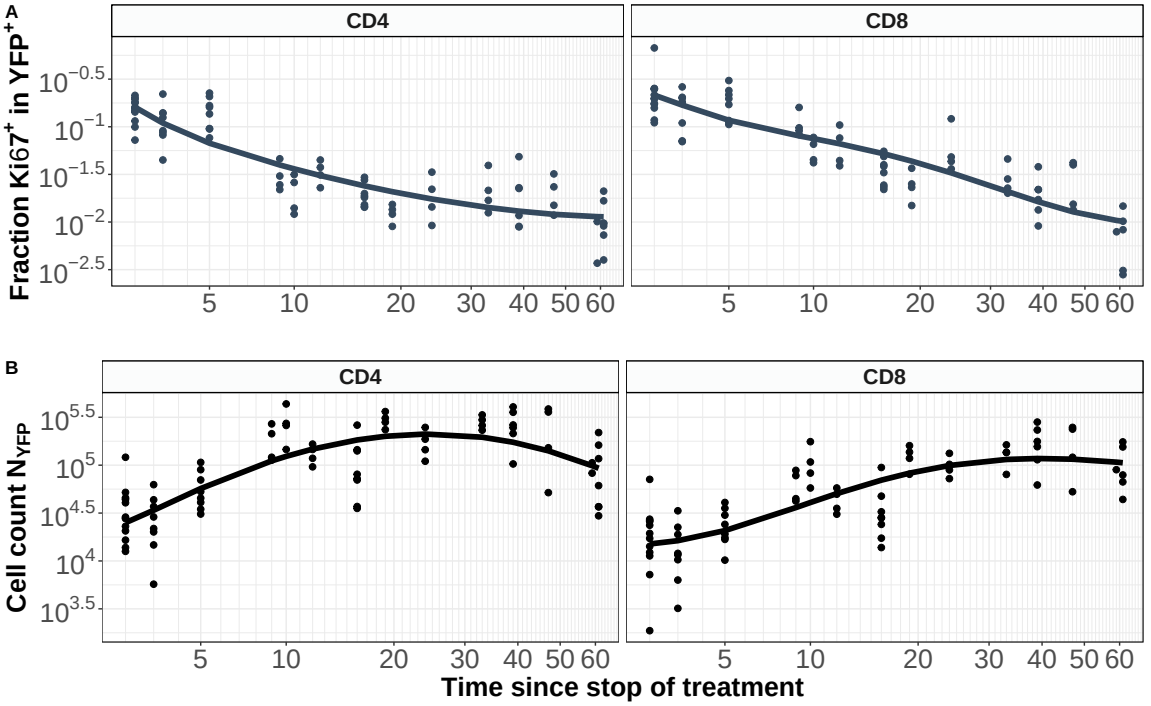
Differences in parameters when fitting separately and simultaneously are more prominent for the model incorporating Ki67 expression within the YFP compartment (Figure A.4B, fits in Figure A.3). When modeling absolute cell counts, Ki67 fluorescent abundance in mSPs was fitted simultaneously (Figure A.1). When fitting the frequency of $Ki67^{high}YFP^+$ cells in the periphery, one also needs to fit the size of the fluorescent naive cell pool (N_{YFP}) changing over time, the number of $Ki67^{high}YFP^+$ mSPs ($K^+(t)$) as well as the abundance of fluorescent mSPs ($Y(t)$) (Figure A.2). We note that neither fitting the frequencies nor the cell count of Ki67 expression within the YFP compartment with Equation 4 and Equation 5 respectively result in normally distributed residuals (Shapiro-Wilk p-values < 0.05). With the large number of parameters to be estimated, we conclude that the parameter search space is too large when fitting simultaneously. This is reflected in high uncertainty in the estimation of the residence time of peripheral cells (Figure 4).

Extrapolating to the general population

Using kinetics estimated from fluorescent data, we project predictions on the cell count and frequency of Ki67 in the general population. We redefine L^+ and L^- in Equation 4 to be Ki67 expressing (+) and lacking (−) non-fluorescent cells. Accordingly, $K^{+/-}$ reflects the cell count of Ki67 expressing mSPs in the thymus. We model the slight decay over time thereof using $K^+ = N_{mSP}ae^{-bt}$ and $K^- = N_{mSP}(1 - ae^{-bt})$ (Figure A.5). N_{mSP} reflects the number of mSPs and is assumed constant. Their values are estimated to be 695,676 cells and 264,050 cells for CD4 and CD8 respectively. On a biological level, K^+ and K^- serve



(a) Cell numbers



(b) Frequencies

Figure 6: Models of naive T cell dynamics that explicitly track Ki67 expression. (**a**) Absolute numbers of Ki67$^{+/-}$ cells modeled with Equation 4, and fitted using a log transformation. (**b**) Fitted Ki67^{+} frequencies within naive YFP^{+} populations (**A**) and numbers of YFP^{+} naive T cells (**B**) with Equation 5 using log-transformation. Parameters shown in Table A.2.

as cell counts of RTEs leaving the thymus at rate Θ. Using geometric means of the data as initial values and parameters obtained from the different methods performed we projected predictions of Ki67 expression in the naive compartment on the data (see Figure A.6B and C). Similarly, we project onto the frequency of Ki67 in the periphery. Note, that Equation 5 can be altered, since the reference population changed. While the size of the fluorescent population changed over time, the general naive T cell population (N_{nai}) is assumed to be at steady state. We estimated N_{nai} to be 7,392,357 cells and 4,773,591 cells for CD4 and CD8 respectively.

This allows for the estimation of Ki67 frequencies (now l^+) using

$$\frac{dl^+}{dt} = \frac{\Theta K^+(t)}{N_{nai}} - l^+(\mu + \beta) + p(2 - l^+).$$

We conclude that while the data may suggest various models, we catch the general trend of decay (Figure A.6C).

Considering intrapopulational variation in the number of mSPs, naive T cells or in the fraction of Ki67, we also predicted the frequency of Ki67 expression in the periphery per mouse i ($k^+_{nai,i}$).

$$k^+_{nai,i} = \frac{\Theta K^+_{mSP,i}}{\beta N_{nai,i}} - \frac{2p}{\beta}$$

with $K^+_{mSP,i}$ representing individual cell counts of Ki67 expressing mSPs and $N_{nai,i}$ naive cell counts. We find that residuals accumulate around zero by visual inspection, supporting our estimated cell kinetics (Figure A.7).

Discussion

We here introduce a method of cell-cycle dependent labeling that is inherited to a cells offspring. This work serves as a proof-of-concept that this system can be used to model lymphocyte dynamics. Label is obtained only if a cell undergoes mitosis at the time of tamoxifen administration. Similar to deuterium and BrdU labeling, the tool is not known to disturb cellular processes or kinetics. It is to be noted that the tool of YFP label here introduced is rather inefficient. This is reflected by small fractions of YFP label in the cell population (below 20%, not shown). Beyond, we are aware of the disadvantage brought about by the necessity of modeling source-labeling, here in the thymus, possibly causing uncertainty.

This study facilitated YFP labeling data along with Ki67 expression data from 61 mice of different ages reported as both cell counts and frequencies. We explain the experimental data using two mathematical models and two methods of fitting. While thymic output is commonly modeled to decrease (exponentially) with age (for instance [9], first mentioned by [6]), it is here regarded to occur at a constant rate, while the actual number of cells eventually migrating to the periphery is determined by the population pool of mSPs at time t. This approach is backed up by previous research stating an influx of 0.7% from a possibly-shrinking thymus [16].
Our first model incorporates YFP abundance only (Equation 1 and Equation 2) and assumes no self-renewal of peripheral naive T cells [2–4, 7, 10]. The model is based on empirical description of label availability in the thymus ($Y(t)$ and $Z(t)$). We find that residence times obtained using best fits of this model are comparable to previously reported values (see YFP model in Figure 4). This is for either fitting YFP availability in the thymus prior, as well as for simultaneous fitting of thymic- and peripheral data. However, confidence intervals are fairly large when allowing for simultaneous fitting of thymic- and peripheral data (see Figure 4). We conclude that using the division reporter YFP system in the form of label abundance only, delivers cell kinetics consistently stable between different fitting methodologies. Parameters also line up with ones reported in literature (Figure 4), supporting the statement of reliable deduction of cell kinetics.

Next, we regarded the subpopulation of Ki67 expression within fluorescent populations (Equation 4 and Equation 5). This allowed for the estimation of proliferation as well as loss of Ki67 expression. We find that estimated residence times and Ki67 expression are comparable to values in literature. Self-renewal is present at a very low level only, even less so for the CD8 subgroup (see Table A.2). This agrees with previous observations of low to no peripheral cell division [2–4, 7, 10]. Similarly, estimates for the rate of Ki67 expression loss (β) in the CD8 population agree with previously reported approximations [19]. Values for β for best fits for data of CD4 T cells are slightly lower (Table A.2). This discrepancy in parameter values is unexpected.
Fitting thymic and peripheral Ki67$^{\text{high/low}}$YFP$^+$ data simultaneously introduced unnormally distributed residuals and very large confidence intervals (Figure 4, fits in Figure A.3). Greater differences in parameter estimates depending on the method of fitting in the model incorporating the Ki67 expression in the YFP population rather than YFP expression only can be explained by the dependence of involved

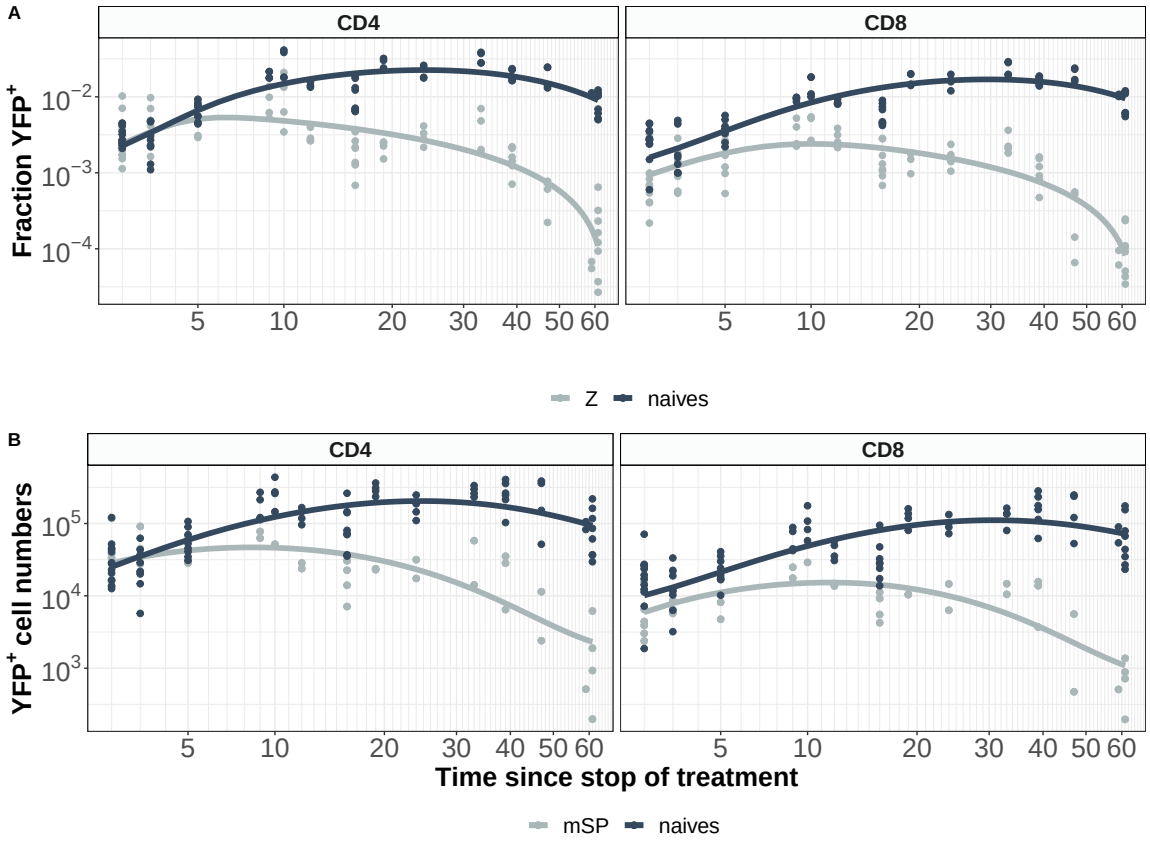


Figure 7: Adjusted fitting procedure to allow for uncertainty in label availability in the thymus. Simultaneous fitting of fluorescent data in the thymus and periphery. (**A**) Frequencies of YFP naive T cells (Equation 2) and $Z(t)$ (Equation 3b fitted with logit transformation. (**B**) Cell numbers of YFP$^+$ naive T cells (Equation 1) and YFP$^+$ mSPs (Equation 3a) fitted with log transformation.

datasets. While data may be assumed independent for YFP data only, this may not be the case for Ki67 data (Figure S.5).

Projecting parameters obtained from fitting Ki67 abundance within the peripheral population on the general population allows for careful conclusions that YFP-expressing cells do not show significantly different parameters from unlabeled cells (Figure A.6). Similarly, we can predict Ki67 expression fractions for individual mice using parameters obtained from YFP labeling data reliably (see Figure A.7).

We conclude suitability of the fluorescent division reporter system for T cell dynamic approximations. While empirical description of label availability in the thymus may introduce uncertainty, approximations of residence- and interdivision times here derived are comparable to literature. With its characteristic to be inherited to offspring rather than getting lost upon cell death, YFP labeling of dividing thymocytes serves as a useful tool in kinetic assessment and may find use in future research. Beyond its application in the current study, YFP labeling may also be facilitated as a division counter system.

Methods

Experimental system and naive T cell biology
Data was collected from *Ki67-mCherry-CreERT* strain mice. Dividing cells express both a Ki67-mCherry fusion protein and CreERT inducible *Cre* construct from the *Mki67* locus (Figure 1). Active cell division can be monitored by mCherry fluorescence, which yields more consistent readouts compared to antibody assays. Permanent labeling can be induced by feeding tamoxifen, to indelibly mark dividing cells and their progeny with YFP (Figure 1). Tamoxifen has a lifetime of a few hours *in vivo* [20], allowing to determine a precise window in which cell proliferation is to be monitored.

Model fitting
Any fit in this study was performed using the FME wrapper Grind in R version 4.1.2. Our algorithm ran 1000 bootstraps for any parameter search using the implemented pseudorandom-search algorithm. To allow for normal distribution of residuals, data and model were normalized using log-, logit- or exponential transformations.

References

[1] Abul K Abbas, Andrew H Lichtman, and Shiv Pillai. Cellular and molecular immunology E-book. Elsevier Health Sciences, 7 edition, 2012.

[2] Yves Modigliani, Graça Coutinho, Odile Burlen-Defranoux, Antonio Coutinho, and Antonio Bandeira. Differential contribution of thymic outputs and peripheral expansion in the development of peripheral t cell pools. European journal of immunology, 24(5):1223–1227, 1994.

[3] Ineke den Braber, Tendai Mugwagwa, Nienke Vrisekoop, Liset Westera, Ramona Mögling, Anne Bregje de Boer, Neeltje Willems, Elise HR Schrijver, Gerrit Spierenburg, Koos Gaiser, et al. Maintenance of peripheral naive t cells is sustained by thymus output in mice but not humans. Immunity, 36(2):288–297, 2012.

[4] Thea Hogan, Graeme Gossel, Andrew J Yates, and Benedict Seddon. Temporal fate mapping reveals age-linked heterogeneity in naive t lymphocytes in mice. Proceedings of the National Academy of Sciences, 112(50):E6917–E6926, 2015.

[5] Sanket Rane, Thea Hogan, Edward Lee, Benedict Seddon, and Andrew J Yates. Towards a unified model of naive T cell dynamics across the lifespan. Elife, 11, 2022.

[6] GG Steinmann, B Klaus, and H-K Müller-Hermelink. The involution of the ageing human thymic epithelium is independent of puberty: a morphometric study. Scandinavian journal of immunology, 22(5):563–575, 1985.

[7] Liset Westera, Vera van Hoeven, Julia Drylewicz, Gerrit Spierenburg, Jeroen F van Velzen, Rob J de Boer, Kiki Tesselaar, and José AM Borghans. Lymphocyte maintenance during healthy aging requires no substantial alterations in cellular turnover. Aging cell, 14(2):219–227, 2015.

[8] Ananda W Goldrath and Michael J Bevan. Selecting and maintaining a diverse t-cell repertoire. Nature, 402(6759):255–262, 1999.

[9] Bas E Dutilh and Rob J De Boer. Decline in excision circles requires homeostatic renewal or homeostatic death of naive t cells. Journal of theoretical biology, 224(3):351–358, 2003.

[10] Sanket Rane, Thea Hogan, Benedict Seddon, and Andrew J Yates. Age is not just a number: Naive t cells increase their ability to persist in the circulation over time. PLoS biology, 16(4):e2003949, 2018.

[11] Arnold Reynaldi, Norah L Smith, Timothy E Schlub, Cybelle Tabilas, Vanessa Venturi, Brian D Rudd, and Miles P Davenport. Fate mapping reveals the age structure of the peripheral t cell compartment. Proceedings of the National Academy of Sciences, 116(10):3974–3981, 2019.

[12] Jeff E Mold, Pedro Réu, Axel Olin, Samuel Bernard, Jakob Michaëlsson, Sanket Rane, Andrew Yates, Azadeh Khosravi, Mehran Salehpour, Göran Possnert, Petter Brodin, and Jonas Frisén. Cell generation dynamics underlying naive T-cell homeostasis in adult humans. PLoS Biol, 17(10):e3000383, 2019.

[13] Jose AM Borghans and Rob J De Boer. Quantification of t-cell dynamics: from telomeres to dna labeling. Immunological reviews, 216(1):35–47, 2007.

[14] Kohsuke Sasaki, Tomoyuki Murakami, Masahiro Kawasaki, and Manabu Takahashi. The cell cycle associated change of the ki-67 reactive nuclear antigen expression. Journal of cellular physiology, 133(3):579–584, 1987.

[15] Iain Miller, Mingwei Min, Chen Yang, Chengzhe Tian, Sara Gookin, Dylan Carter, and Sabrina L Spencer. Ki67 is a graded rather than a binary marker of proliferation versus quiescence. Cell reports, 24(5):1105–1112, 2018.

[16] Stuart P Berzins, Richard L Boyd, and Jacques FAP Miller. The role of the thymus and recent thymic migrants in the maintenance of the adult peripheral lymphocyte pool. The Journal of experimental medicine, 187(11):1839–1848, 1998.

[17] Hiroshi Mohri, Alan S Perelson, Keith Tung, Ruy M Ribeiro, Bharat Ramratnam, Martin Markowitz, Rhonda Kost, Leor Weinberger, Denise Cesar, Marc K Hellerstein, et al. Increased turnover of t lymphocytes in hiv-1 infection and its reduction by antiretroviral therapy. The Journal of experimental medicine, 194(9):1277–1288, 2001.

[18] Ruy M Ribeiro, Hiroshi Mohri, David D Ho, and Alan S Perelson. In vivo dynamics of t cell activation, proliferation, and death in hiv-1 infection: why are cd4+ but not cd8+ t cells depleted? Proceedings of the National Academy of Sciences, 99(24):15572–15577, 2002.

[19] Graeme Gossel, Thea Hogan, Daniel Cownden, Benedict Seddon, and Andrew J Yates. Memory cd4 t cell subsets are kinetically heterogeneous and replenished from naive t cells at high levels. elife, 6:e23013, 2017.

[20] Joel M Reid, Matthew P Goetz, Sarah A Buhrow, Chad Walden, Stephanie L Safgren, Mary J Kuffel, Kathryn E Reinicke, Vera Suman, Paul Haluska, Xiaonan Hou, et al. Pharmacokinetics of endoxifen and tamoxifen in female mice: implications for comparative in vivo activity studies. Cancer chemotherapy and pharmacology, 74(6):1271–1278, 2014.

A Appendix

Table A.1: Parameters of best fits for Equation 1 and Equation 2. Initial conditions were also estimated but are not shown here. Values are reported along with 95% Confidence Intervals for either fitting data from the thymus and periphery separately (Separate fitting) or in one step (Simultaneous fitting).

	Thymic output (Θ) [1/days]		Residence time (1/μ) [days]	
Separate fit				
CD4	0.357	[0.281, 0.437]	25.3	[20.7, 34.0]
	0.352	[0.272, 0.443]	26.7	[19.3, 43.3]
CD8	0.424	[0.311, 0.556]	34.9	[25.3, 59.4]
	0.459	[0.310, 0.653]	34.7	[21.5, 70.0]
Simultaenous fit				
CD4	0.427	[0.357, 0.501]	19.7	[16.0, 25.3]
	0.356	[0.280, 0.441]	26.0	[18.7, 40.5]
CD8	0.540	[0.396, 0.685]	25.9	[18.0, 43.2]
	0.477	[0.327, 0.631]	35.3	[22.1, 87.2]

Frequency Cell count

Table A.2: Parameters of best fits of Equation 4 and Equation 5. Initial parameters have also been estimated but are not reported here. Values are reported along with 95% Confidence Intervals for either fitting data from the thymus and periphery separately (Separate fitting) or in one step (Simultaneous fitting).

	Thymic output (Θ) [1/days]		Residence time (1/μ) [days]		Ki67 expression time (1/β) [days]		Interdivision time (1/p) [days]	
Separate fit								
CD4	0.385	[0.253, 0.526]	22.8	[15.8, 42.8]	1.94	[1.37, 2.67]	438	[249, 792]
	0.459	[0.328, 0.626]	24.9	[17.8, 42.8]	1.62	[1.03, 2.26]	399	[179, 1333]
CD8	0.331	[0.249, 0.532]	83.0	[31.1, 100.0]	3.74	[2.51, 5.37]	897	[349, 3413]
	0.508	[0.320, 0.764]	39.9	[22.4, 206.2]	3.43	[2.10, 4.77]	NA	NA
Simultaneous fit								
CD4	0.477	[0.193, 0.480]	20.5	[18.6, 53.4]	1.69	[1.00, 2.31]	316	[140, 506]
	0.922	[0.189, 1.000]	7.2	[2.9, 100.0]	1.31	[1.00, 3.62]	344	[42, 10000]
CD8	0.597	[0.205, 0.600]	24.7	[21.6, 330.4]	1.76	[1.12, 8.84]	243	[146, 100000]
	0.891	[0.339, 1.000]	1.5	[1.5, 100.0]	1.00	[1.00, 6.42]	15	[15, 10000]

Frequency Cell count

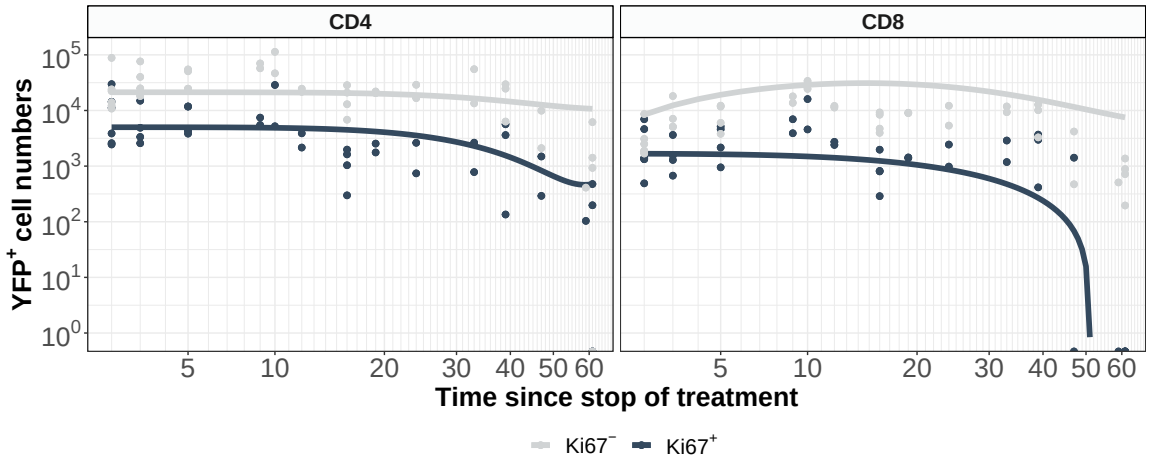


Figure A.1: Cell count of Ki67$^{high/low}$ fluorescent mSPs fitted simultaneously with peripheral cell counts of Ki67$^{high/low}$YFP^{+} naive T cells using Equation 4 and the derivative of Equation 3a and a steady Ki67 expression in the thymus.

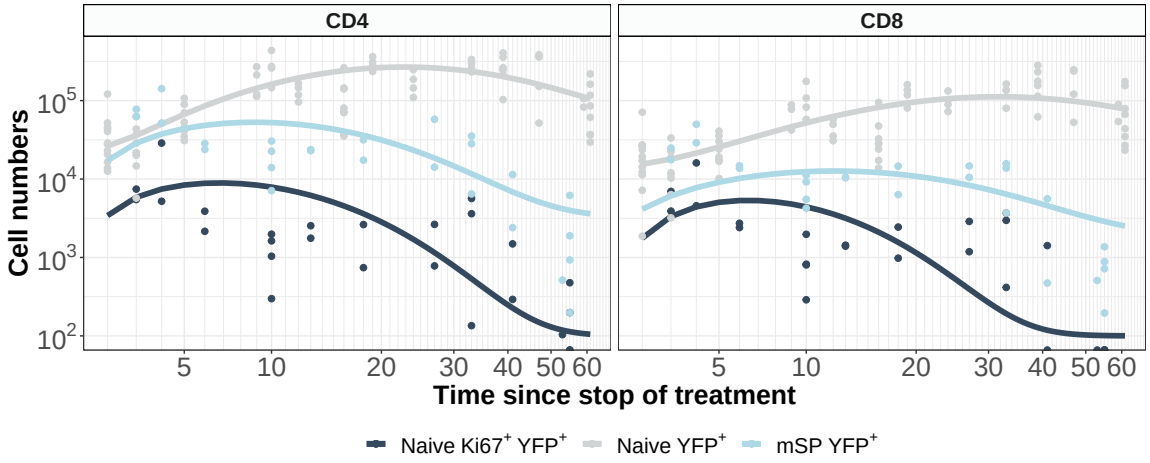
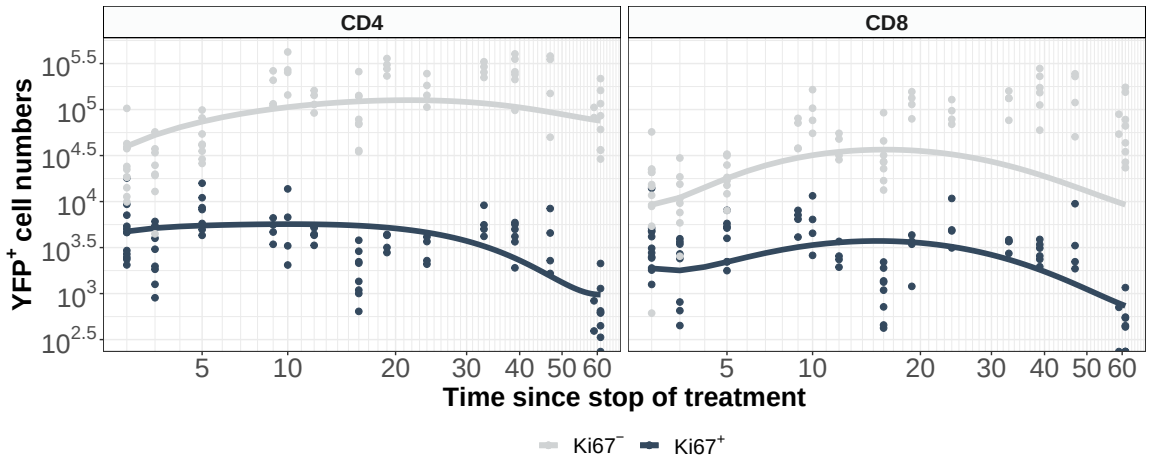
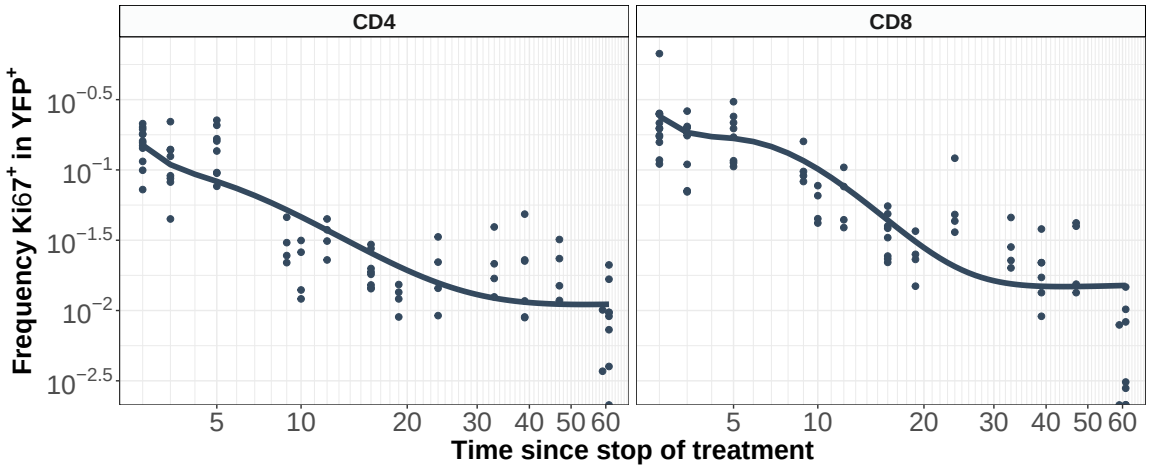


Figure A.2: Cell count of Ki67^{high} fluorescent mSPs, fluorescent naive T cells and mSPs fitted simultaneously with the peripheral frequency of Ki67^{high}YFP^{+} naive T cells of the simultaneous fit of Equation 5, the derivative of Equation 3a and Ki67 expression.



(a) Cell counts



(b) Frequencies

Figure A.3: Simultaneous fitting of fluorescent and Ki67 expression data in the thymus and periphery. **(a)** Best fit for cell counts of $Ki67^{high/low}YFP^{+}$ as described in Equation 4, fitted with log transformation. **(b)** Frequencies of Ki67 expression in YFP^{+} populations fitted with logit transformation.

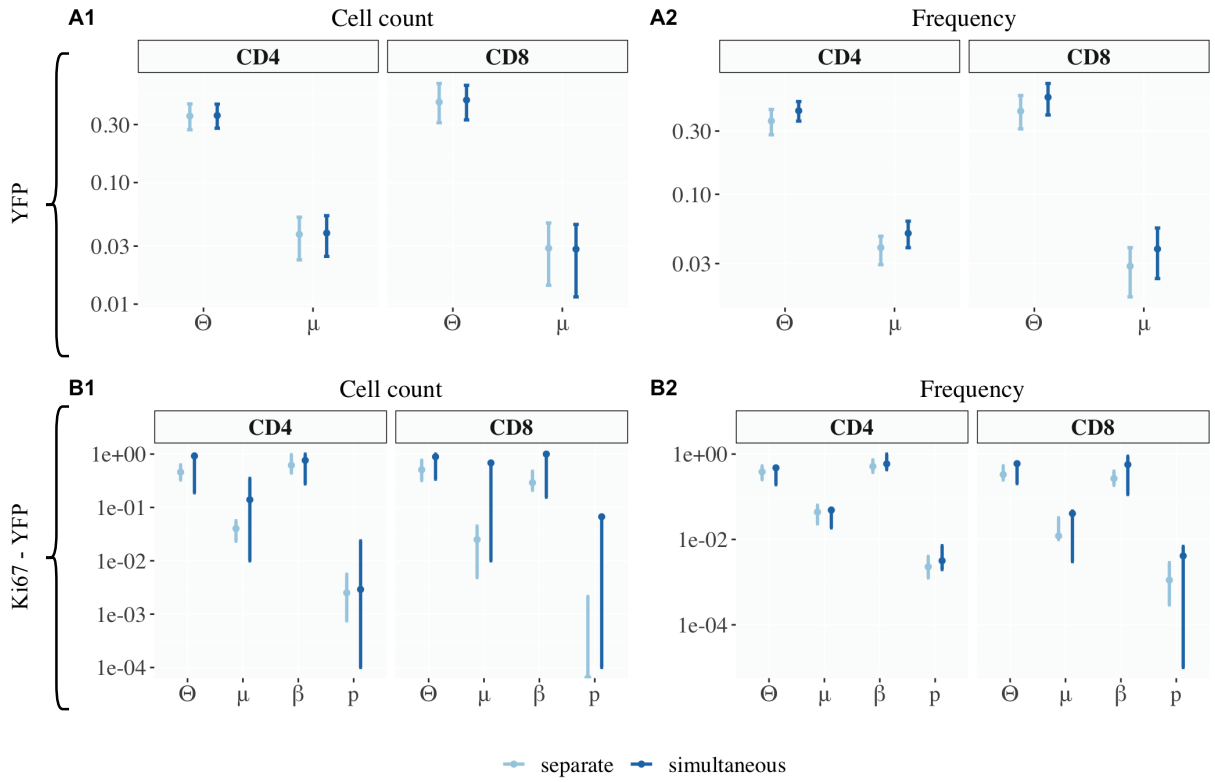


Figure A.4: Comparison of parameters obtained from either fitting data in the thymus and periphery to their models separately, or together. **A1** corresponds to parameters of the best fits for models for Equation 3b and Equation 1. **A2** shows results for Equation 3b and Equation 1. The lower panel shows fit fot models inclusing both YFP- and Ki67 data. Equation 4 as estimated along with $K^{+/-}$ is shown in **B1**. The frequencies of Ki67 within the YFP population were approximated with Equation 5.

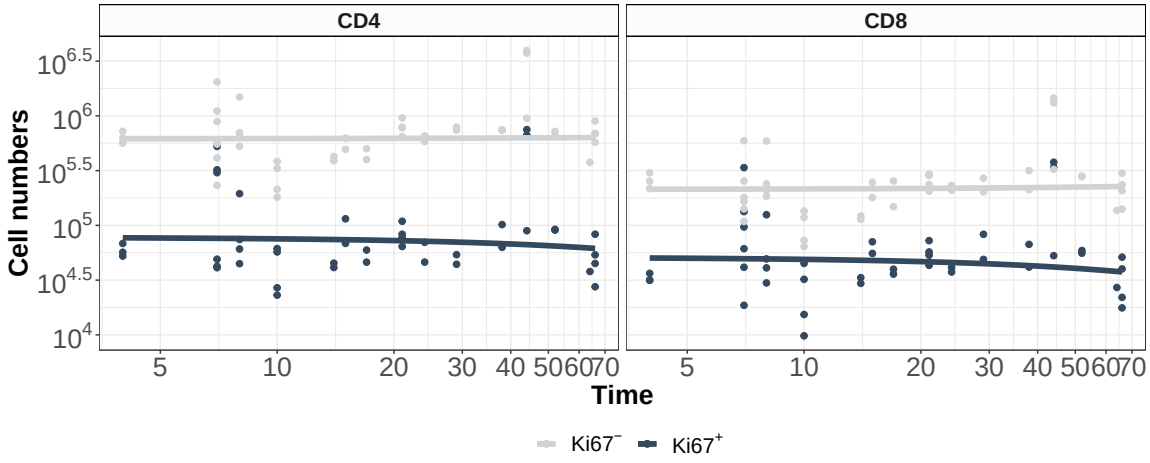


Figure A.5: Approximation of cell counts of Ki67 expressing mSPs using $K^{+} = Nae^{-bt}$ and $K^{-} = N(1 - ae^{-bt})$. N represent the total count of mSPs in the thymus and is estimated to be 695676 and 264050 for CD4 and CD8 respectively.

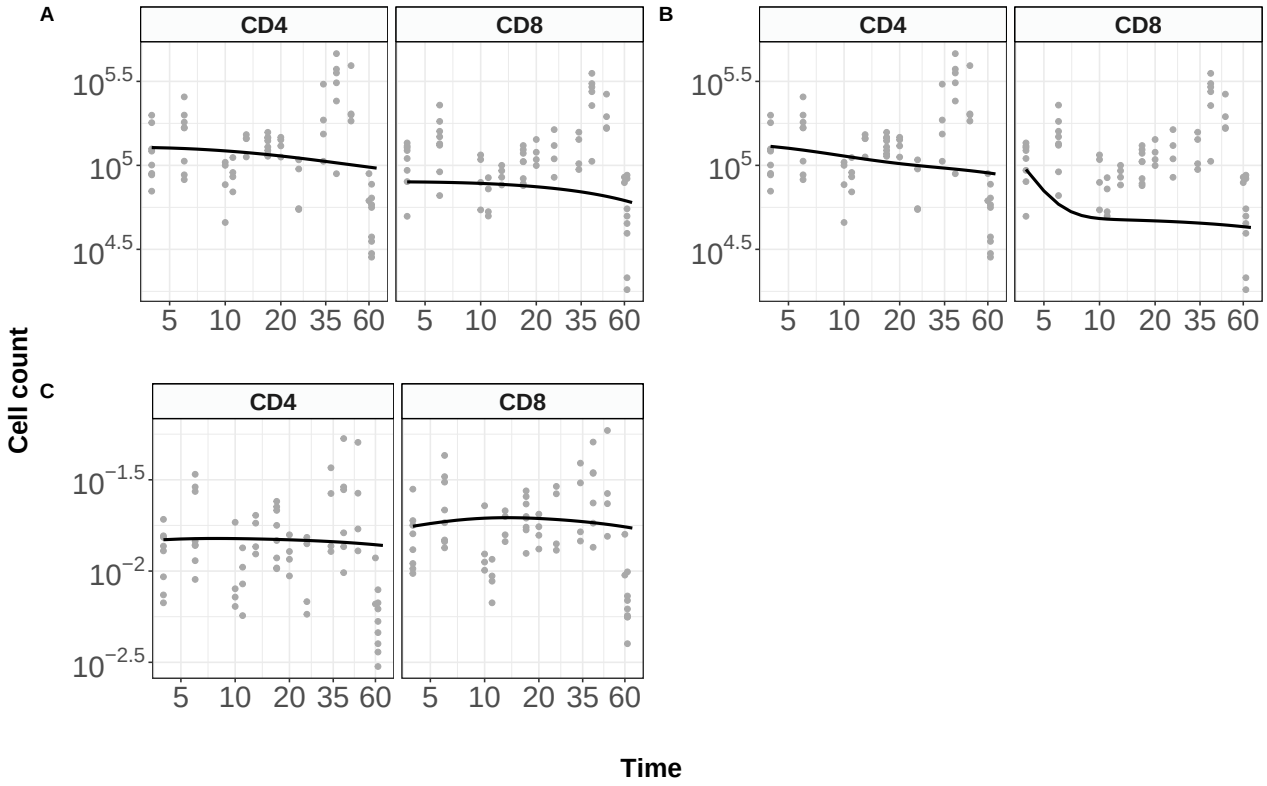


Figure A.6: Prediction of Ki67 expression in the general naive T cell population using three different sets of parameter. **A** uses parameters obtained from separate fitting of $Ki67^{high/low}YFP^{+}$ cell counts in the thymus and periphery, while **B** uses parameters obtained from simultaneous fitting. **C** predicts frequencies of Ki67 in the peripheral naive cell population using estimates of separate fits in thymus and periphery.

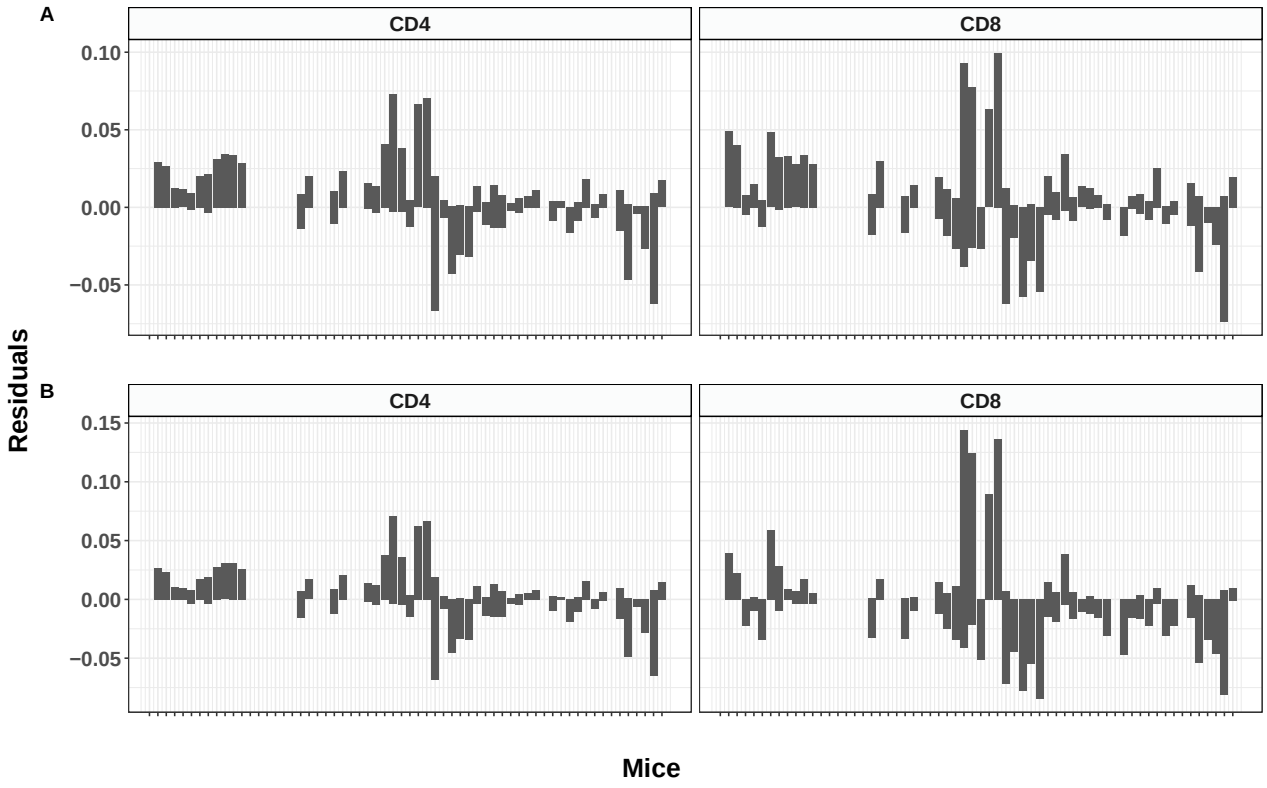


Figure A.7: Predicted the frequency of Ki67 expression in the periphery per mouse i ($k^+_{nai,i}$) with $k^+_{nai,i} = \frac{\Theta K^+_{mSP,i}}{\beta N_{nai,i}} - \frac{2p}{\beta}$, where $K^+_{mSP,i}$ are individual cell counts of Ki67 expressing mSPs and $N_{nai,i}$ naive cell counts in mouse i. Parameters are estimates from best fits of Equation 4 and Ki67 fluorescent cell counts in separate fitting of thymus and periphery (**A**), or approximates of Equation 5 using Ki67 frequencies within fluorescent populations obtained from separate fitting (**B**).

S Supporting information

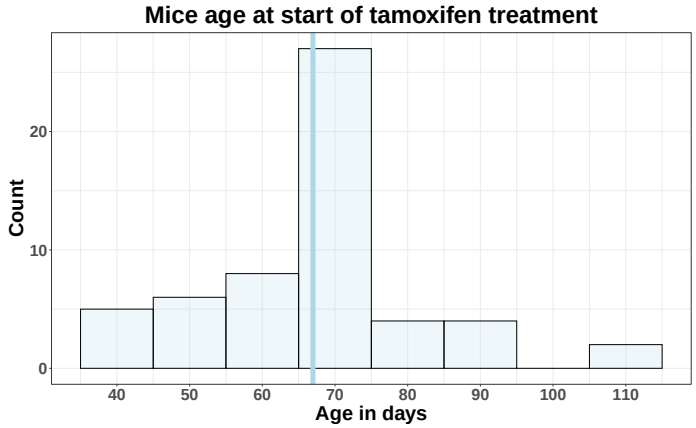


Figure S.1: Distribution of ages of mice at start of tamoxifen treatment. Vertical line marks mean age at onset.

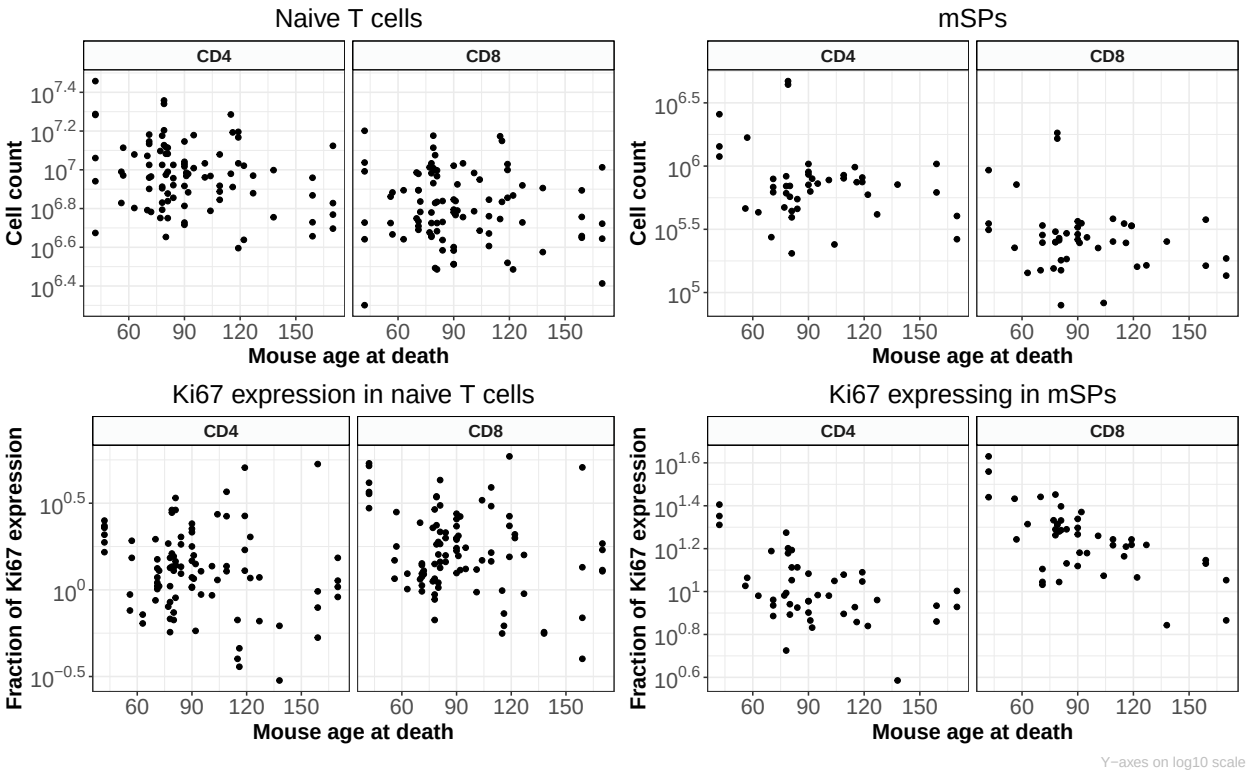
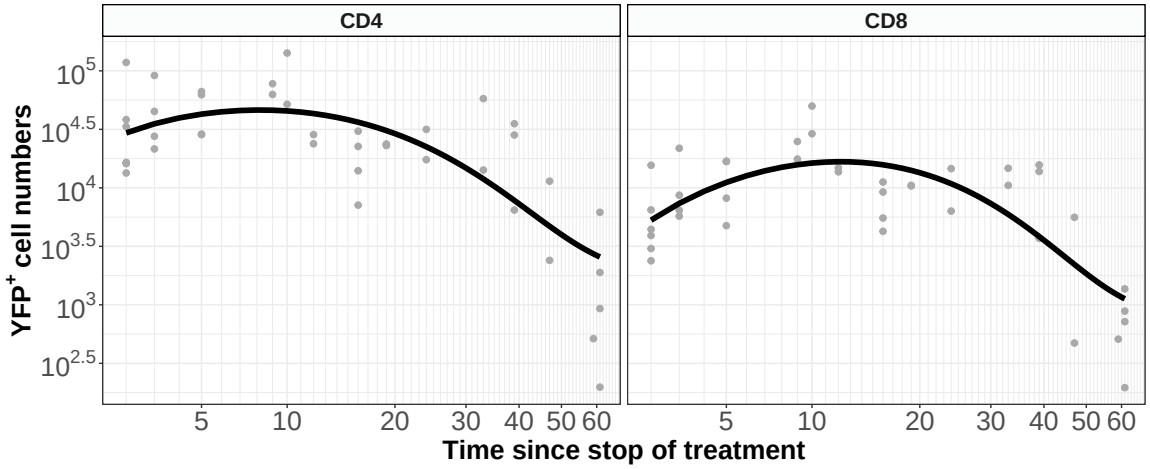
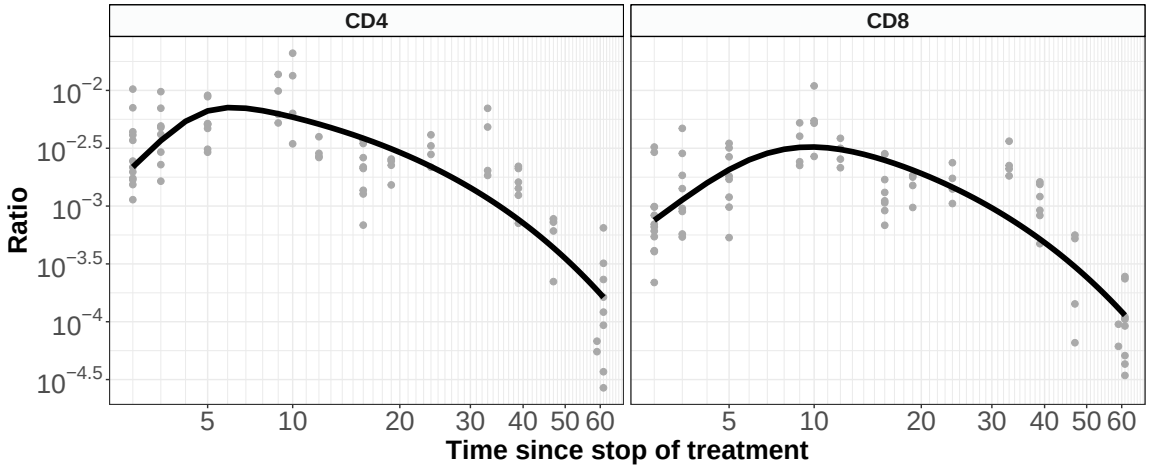


Figure S.2: Basic characteristics of each mouse, against age at death. **A** Numbers of naive CD4 and CD8 T cells, recovered from spleen and lymph nodes; **B** mature SP4 and SP8 thymocytes; **C** Ki67 expression in naive CD4 and CD8 T cells; **D** Ki67 expression among SP thymocytes.



(a) Y(t)



(b) Z(t)

Figure S.3: Phenomenological functions describing the timecourses of YFP-expression among mature SP cells in the thymus after tamoxifen treatment. (**a**) Best fits of the time-varying numbers of YFP$^+$ mSPs ($Y(t)$; Equation 3a). (**b**) The ratio of YFP$^+$ mSPs to naive T cells numbers ($Z(t)$; Equation 3b).

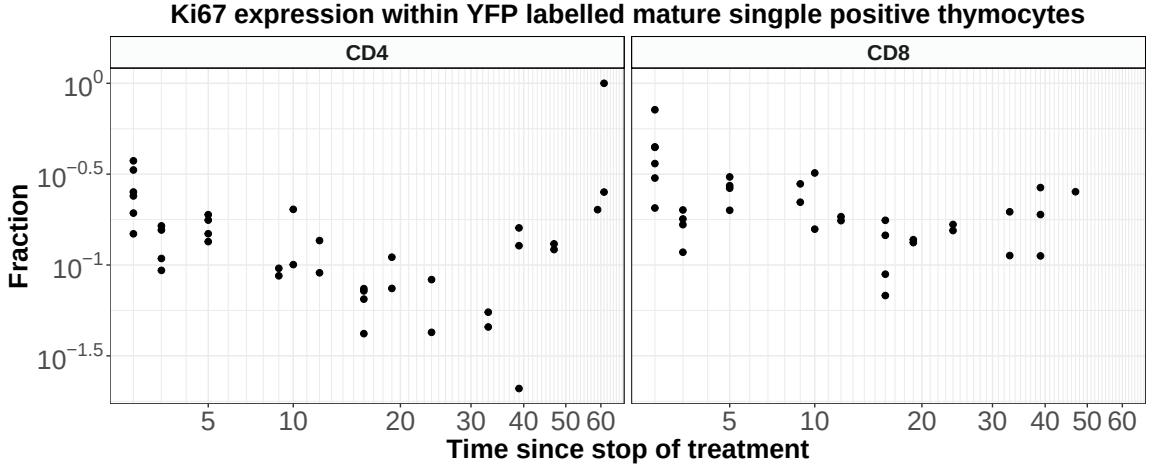


Figure S.4: The fraction of Ki67 expression within the YFP expressing mSP population is assumed rather constant as indicated by the data.

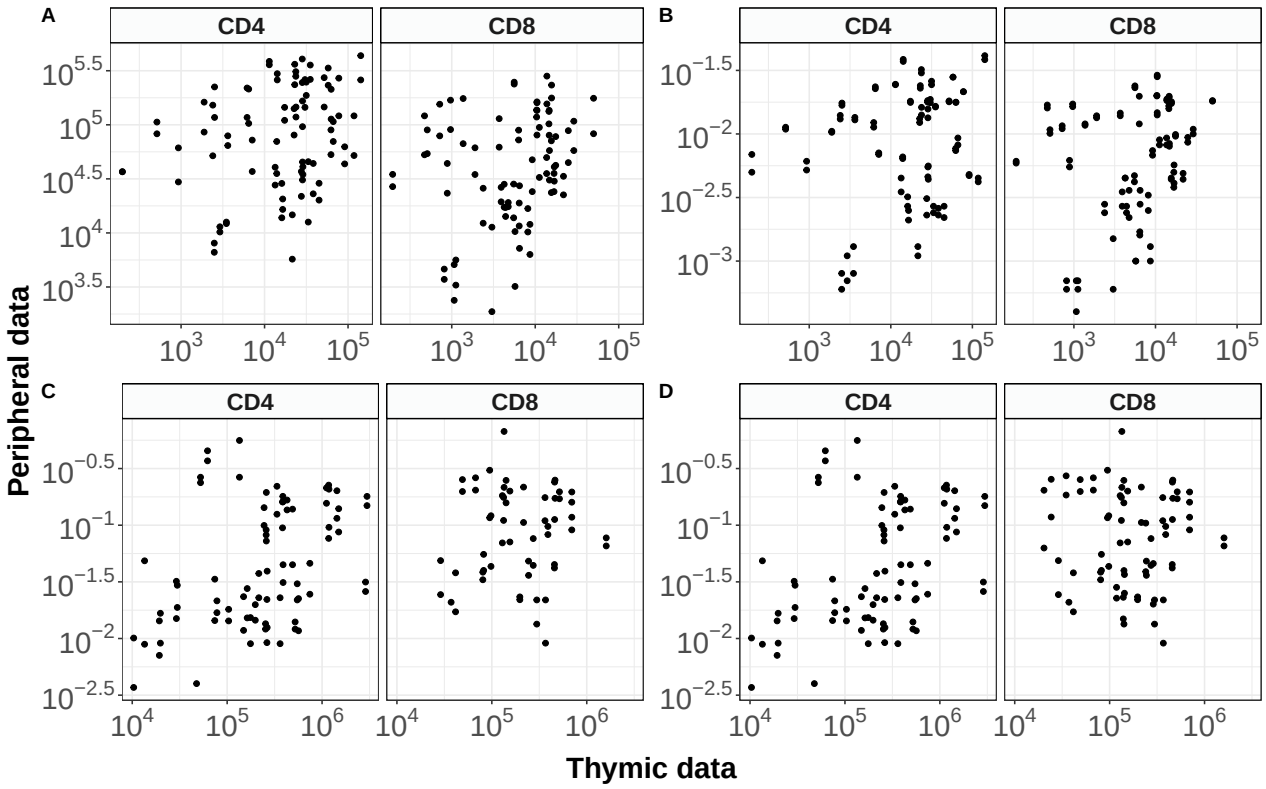


Figure S.5: Correlation of thymic and peripheral data per mouse. Pearson correlation coefficients: **A** 0.3288, **B** 0.2880, **C** 0.6779, **D** 0.1041.

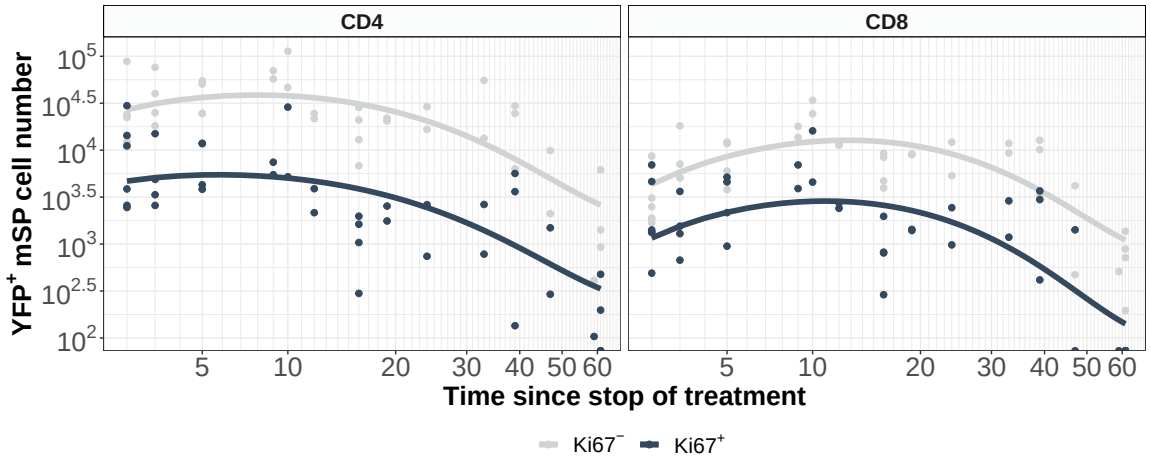


Figure S.6: Absolute number of Ki67-expressing mSPs within group of fluorescent mSPS. Cell numbers are fitted using Equation 3a and a second degree polynom for the fraction of Ki67 within and log-transformation.