



UNIVERSITY OF LEEDS

This is a repository copy of *Some deterministic and stochastic mathematical models of naive T-cell homeostasis*.

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/133020/>

Version: Accepted Version

Article:

Lythe, G and Molina-París, C (2018) Some deterministic and stochastic mathematical models of naive T-cell homeostasis. *Immunological Reviews*, 285 (1). pp. 206-217. ISSN 0105-2896

<https://doi.org/10.1111/imr.12696>

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

Some deterministic and stochastic mathematical models of naïve T-cell homeostasis

Grant Lythe and Carmen Molina-París
School of Mathematics, University of Leeds, Leeds LS2 9JT, UK

4th July 2018

Contents

1	Introduction	3
1.1	Simple deterministic models	3
1.2	A simple stochastic model	5
2	Models of clonal survival	5
2.1	Death process	6
2.2	Neutral models	6
2.3	Global competition and thymic pressure	7
2.3.1	Diffusion approximation and clonal size distribution	9
2.4	Specific competition and thymic pressure	9
2.5	Computational models	10
3	How many TCR clonotypes are maintained?	11
4	Discussion	16

Running title Mathematical models of T-cell homeostasis

Abstract

Humans live for decades, mice live for months. Over these long timescales, naïve T cells die or divide infrequently enough that it makes sense to approximate death and division as instantaneous events. The population of T cells in the body is naturally divided into clonotypes; a clonotype is the set of cells that have identical T cell receptors. While total numbers of cells, such as naïve $CD4^+$ T cells, are large enough that ordinary differential equations are an appropriate starting point for mathematical models, the numbers of cells per clonotype are not. Here we review a number of basic mathematical models of the maintenance of clonal diversity. As well as deterministic models, we discuss stochastic models that explicitly track the integer number of naïve T cells in many competing clonotypes over the lifetime of a mouse or human, including the effect of waning thymic production. Experimental evaluation of clonal diversity by bulk high-throughput sequencing has many difficulties, but the use of single-cell sequencing is restricted to numbers of cells many orders of magnitude smaller than the total number of T cells in the body. Mathematical questions associated with extrapolating from small samples are therefore key to advances in understanding the diversity of the repertoire of T cells. We conclude with some mathematical models on how to advance in this area.

Keywords stochastic, competition, ageing, extinction, computational models, single-cell sequencing, sampling

1 Introduction

A T cell is characterised by the T-cell receptors (TCRs) on its surface. The TCR is formed by random gene recombination processes in the thymus and determines the set of peptide-major histocompatibility (pMHC) complexes that the T cell is capable of reacting to. Every T cell found in the rest of the body either originated in the thymus, or is descended from a cell that did. Daughter cells inherit their TCR from their mother. More than 90 percent of putative T cells die in the thymus, either because they are insufficiently reactive, or because they react too strongly to pMHC ordinarily found in the body (37, 102, 69, 46, 70, 43, 87, 104). This thymic “selection” produces a repertoire of T cells, with many different specificities (31, 54). An adaptive immune response begins with a small number of binding events between TCR and pMHC on the surfaces of an antigen-presenting cell (APC) that has processed peptides not ordinarily found in the body (10). T cells with the correct specificity, initially rare, then enter a programme of multiple rounds of cell division (55), activation (107) and production of chemical signals. The initial binding events typically occur in one of the body’s lymph nodes (45); T cells usually circulate from node to node (53), scanning APCs for pMHC, but rarely encountering strong stimulation (92). Here, we review some models of naïve T-cell repertoire homeostasis: the maintenance by the body of a diverse repertoire of T cells capable of responding to new threats.

The field of immunology owes its name to the observation that the immune system responds more rapidly and more strongly to an infection that it has previously fought. Thus vaccination, deliberate introduction of well-chosen foreign peptides into a body, prevents the spread of diseases. This phenomenon of immunity, or property of memory, is due to a rebalancing of the repertoire of immune cells. The temporary army of T cells that springs into existence during an episode of infection does not demobilise back to its small number of precursors, but retains an enhanced presence that includes types of “memory” cells, with different patterns of circulation through organs and tissues (90, 77, 21, 100, 73, 78, 4, 11, 96, 7).

We focus on naïve T cells whose TCR is made up of an α and β chain. The set of such cells is unambiguously divided into $CD4^+$ and $CD8^+$ T cells. Every T cell makes an irreversible commitment to one of the two lineages in the thymus that is, along with TCR specificity, inherited by daughters of cells that divide. $CD4$ is a surface molecule that acts as a coreceptor when a $CD4^+$ or “helper” T cell, binds to a pMHC complex containing type II MHC; $CD8$ is a surface molecule that acts as a coreceptor when a $CD8^+$ or “cytotoxic” T cell, binds to a pMHC complex containing type I MHC. In order to understand the homeostasis and function of the immune system, further classifications of T cells into subsets including, for example central and effector memory based on $CD62L$, $CCR7$ and other surface markers, regulatory cells based on $CD25$ and $FOXP3$ and others, are necessary (73, 38). Full understanding remains elusive, in part because the distinctions between subsets are seldom clearcut, their origins and possible reversibility are the subject of active research (2, 68, 22, 29, 97).

To maintain their populations over the lifetime of the host, naïve $CD4^+$ and $CD8^+$ T cells depend on cytokine resources, especially IL-7, to survive and divide (89, 80, 79, 60, 25, 83, 71, 44). They may obtain IL-7 as they circulate through lymph nodes and the spleen, where it is produced by stromal cells such as fibroblastic reticular cells (51, 64). It is possible to devise mathematical models of the binding of IL-7 molecules to receptors, and of the processes of internalisation, degradation and recycling of complexes on the surface and in the cytoplasm of individual cells (61). Then the variables of the model may be numbers of molecules of free ligand, numbers of ligand-bound complex, numbers of internalised complexes instead of numbers of cells. However, in this review we restrict ourselves to modelling at the level of cells and populations of cells.

1.1 Simple deterministic models

We can model a large T-cell population, for example all naïve $CD4^+$ T cells in a individual, about 10^7 in an adult mouse or about 10^{11} in an adult human (20). In a single-equation model, the number of cells is represented not as an integer but by a real function of age, $N(t)$, that obeys an ordinary differential equation (ODE) of the form

$$\frac{d}{dt}N = \text{rate of thymic export} - \text{rate of cell death} + \text{rate of cell division} .$$

The simplest model, which may be obtained by assuming that cells circulate independently throughout the body and every cell is equally likely to undergo one round of division, or suffer death, in any small time interval, is

$$\frac{d}{dt}N = \theta - \beta N, \quad (1.1)$$

where θ is the rate of export of cells from the thymus and β is the death rate per cell minus the division rate per cell. If θ does not depend on age and $N(0) = 0$, then the solution of (1.1) is

$$N(t) = (1 - e^{-\beta t}) \frac{\theta}{\beta},$$

with steady state value $N_{ss} = \frac{\theta}{\beta}$. The transformation of some naïve cells into memory cells can also be included as another net loss rate contributing to the value of β .

To assess the usefulness of the simple model (1.1), let us consider reasonable values of the parameters θ and β . If a mouse's thymus produces 10^6 CD4⁺ cells per day, one peripheral naïve cell in 30 dies per day and one in 300 divides per day, then $\beta = 0.03 \text{ day}^{-1}$ and, using (1.1), we conclude that there are 33 million naïve CD4⁺ T cells at steady state. Over the timescale $1/\beta$ of approximately one month in mice (longer in humans), it is reasonable to treat cell division and death as events that may occur in a "short" time interval. On the same timescale, however, there is an important decrease in the average thymic production of a mouse. We therefore replace the constant θ by a function of age, $\theta(t)$. For mice older than six weeks, we may use the function (36)

$$\theta(t) = Ae^{-\nu t} \quad \text{cells per day,}$$

where $\nu = 0.005 \text{ day}^{-1}$ and the value of the constant A varies from mouse to mouse. If, at six weeks old, a mouse has 30 million naïve CD4⁺ T cells and its thymus is producing one million cells per day, we find

$$N(t) = \frac{\theta(t)}{\beta - \nu} = 10^7 e^{-\beta(t-42)}. \quad (1.2)$$

In this description, because $e^{-\beta t}$ decreases more rapidly as a function of age than $e^{-\nu t}$, the long-term decrease of the average number of naïve T cells in a mouse has the same timescale as the waning of thymic export to the periphery. It is also possible to model the increase in thymic output from infancy to adulthood (6). For comparison, the rate of export of naïve CD4⁺ T cells from the thymus of a young adult human is about 2×10^8 cells per day; the rate declines proportional to $e^{-\nu t}$ where $\nu = 0.0001 \text{ day}^{-1}$ (5).

This type of mathematical abstraction is valuable because its parameters have clear interpretation and their values can be measured. Division rates per cell can be estimated by Ki67 staining (36). Suppose that, after cell division, the two daughter cells remain Ki67⁺ for four days. Then, ignoring the time spent cycling before division, the division rate λ , which is the probability per day that a cell in the population divides, is equal to one eighth of the fraction of cells in that population that are Ki67⁺. Death and division rates per cell are estimated using, for example, deuterium which is incorporated into the DNA of dividing cells (9, 17, 18, 1).

Simple formulae such as (1.2) that result from simple models such as (1.1) can be fit to experimental data, using statistical techniques to understand which parameters vary from individual to individual, and which are common. To consider T-cell homeostasis as a whole in this framework, separate models of the form (1.1) can be considered for the different thymocyte populations, for CD4⁺ and CD8⁺ naïve T cells and for CD4⁺ and CD8⁺ memory T cells (9, 84, 33). Memory cells have higher rates of division than naïve cells; their input terms are not due to direct export from the thymus, but to differentiation of naïve cells. The memory compartments are typically classified into "central memory", "effector memory", and perhaps also "tissue-resident memory" or "stem-cell memory" compartments. If recent thymic emigrants behave differently from mature naïve T cells (24, 98), then they warrant a separate equation.

Models based on (1.2) cannot describe extreme scenarios such as recovery from lymphopenia (49, 30), but they are useful in describing T-cell homeostasis. It could be said that they are most useful when, as in two recent cases, they have been shown to fail. In the first case, closer examination of deuterium-labelling data reveal that the assumption that all T cells in a large population share common values of death and division rates to be inadequate. Instead, models of "kinetic heterogeneity" are constructed in which the total cell population consists of sub-populations with different rates (27). Secondly, elegant recent experiments

where endogenous T cells are gradually replaced with donor T cells after administration of the bone-marrow-depleting drug Busulfan, revealed evidence that the naïve and memory T cell subsets in mice should be further subdivided into fast and slow, or resident and incumbent, subsets (36, 33, 75). Thus, careful analysis of data in the light of simple mathematical models is shedding light on the underlying complexity of the T-cell repertoire.

1.2 A simple stochastic model

Stochastic models can respect the property that the number of cells is always an integer. Fluctuations in cell numbers, and possible extinction of a population, are included in a natural way. To illustrate this, let us consider the stochastic version of the simplest model with constant θ . Let $N(t)$ be the total number of cells alive at age t . As $\Delta t \rightarrow 0$, the mean number of new cells produced in the time interval $(t, t + \Delta t)$ is $\theta \Delta t$ and the mean number of cell deaths in the interval is $\mu N(t) \Delta t$. Let

$$P_k(t) = \text{Prob}[N(t) = k],$$

for each integer k . Let the mean total number of cells be $x(t) = \mathbb{E}(N(t))$. Then

$$x(t) = \sum_{k=1}^{\infty} k P_k(t),$$

and

$$\frac{d}{dt} x = -\mu x + \theta(1 - P_0).$$

To estimate the size of the fluctuations about the mean number of cells, we define the mean of the square of $N(t)$:

$$y(t) = \mathbb{E}(N(t)^2) = \sum_{k=1}^{\infty} k^2 P_k(t).$$

Then we can write

$$\frac{d}{dt} y(t) = \sum_{k=1}^{\infty} k^2 \frac{d}{dt} P_k(t) = -2\mu y + \theta + x(2\theta + \mu).$$

As $\mu t \rightarrow \infty$,

$$y(t) \rightarrow \left(\frac{\theta}{\mu}\right)^2 + \frac{\theta}{\mu}.$$

In other words,

$$\mathbb{E}(N(t)) \rightarrow \frac{\theta}{\mu} \quad \text{and} \quad \text{var}(N(t)) \rightarrow \frac{\theta}{\mu}.$$

The fluctuations in the total number of cells, as measured by the standard deviation, are equal to the square root of the mean. While this is useful, these types of fluctuations may be far smaller than the variation from mouse to mouse, or experiment to experiment.

2 Models of clonal survival

The most important property of the T-cell repertoire is its diversity. Every T cell in a mammal's body can be said to belong to the family of cells with which it shares a TCR, even if they never again find themselves close to each other after they, or their ancestors, leave the thymus. The number of cells in a family is an integer that increases whenever one of its cell divides and decreases whenever one of its cells dies.

In homeostatic conditions, death and division are infrequent enough that it makes sense to treat them as instantaneous and independent events, neglecting the time taken in the processes of death or division (13). However, the numbers of cells per clonotype are no longer large enough to justify the use of a ODE approximation. In particular, ODE models do not capture the possible extinction of clonotypes in a natural way. Thus, we will consider models, in which $n_i(t)$, the number of naïve T cells of a clonotype i at time t , is an integer, with cell death and division treated as the basic events in a multi-dimensional “birth-and-death” stochastic process.

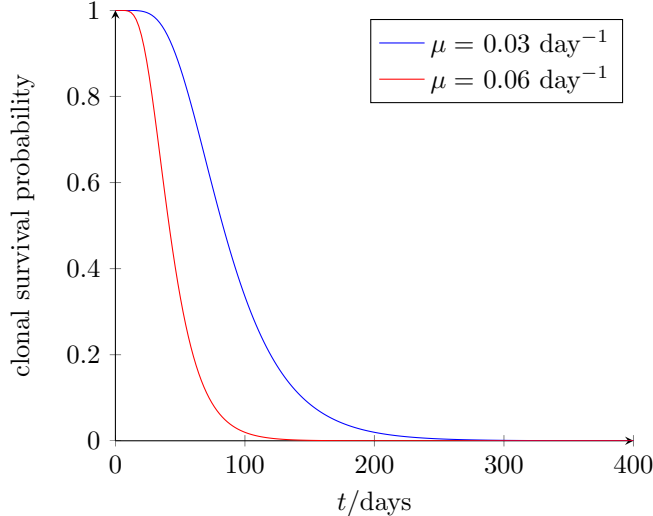


Figure 1: Survival of clonotypes in the periphery, according to (2.1), which assumes no peripheral division. The values of μ are similar to estimated death rates of naïve cells in mice and the initial number of cells per clonotype is $n_\theta = 8$.

2.1 Death process

In an adult mouse, where division of naïve T cells in the periphery is either absent or rare, the simplest model for a single clonotype can be summarised as follows: n_θ cells with the same TCR leave the thymus and begin to circulate independently through the body. They die, one by one, until the clonotype is extinct. At this end of the spectrum of scenarios, the birth-and-death process is simply a death process with rate μ , because the birth rate is zero. Assuming that each cell in each clonotype, independently, has a constant death rate μ , the probability that a clonotype that emerges from the thymus as a family of n_θ cells, survives in the periphery longer than t is given by

$$S(t) = 1 - (1 - e^{-\mu t})^{n_\theta}, \quad (2.1)$$

and is plotted in *Fig. 1*, with $n_\theta = 8$.

2.2 Neutral models

In an adult human, in contrast, the frequency of division of naïve T cells in the periphery (at least 10^8 new cells per day) outweighs thymic export (about 10^7 new cells per day, declining with age). In this case, the opposite end of our spectrum of scenarios, the simplest approximation is that each clonotype, independently, follows a “population birth-and-death process” with equal birth and death rates. That is, in a small time interval Δt the probability that any cell dies, $\mu\Delta t$, is equal to the probability that it divides. Extinction of clonotypes occurs in this type of “neutral” dynamics (39, 28) even if the total number of cells is stable.

To find the survival function of a clonotype under the assumption of neutral dynamics, we define

$$p_k(t) = \text{Prob}[n_i(t) = k | n_i(0) = n],$$

and

$$G_n(z, t) = \sum_{k=0}^{\infty} p_k(t) z^k. \quad (2.2)$$

Because cells behave independently, the descendants of each of the n_0 initial cells can be thought of as n_0 independent families and the “probability generating function” satisfies (41),

$$G_n(z, t) = G_1(z, t)^n. \quad (2.3)$$

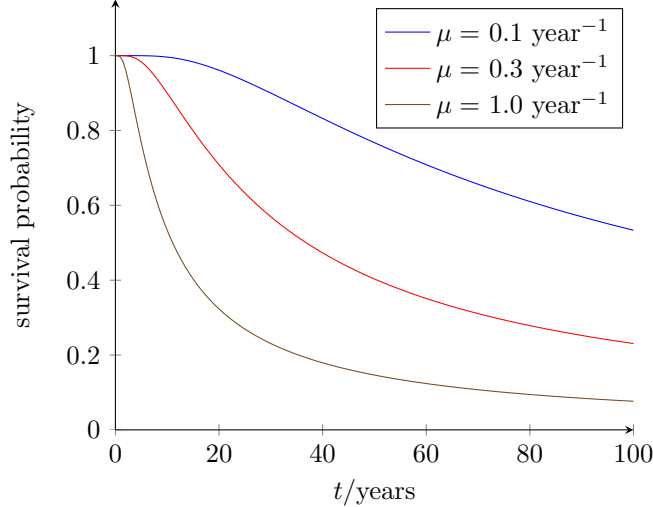


Figure 2: Survival of clonotypes in the periphery under the approximation (2.5) of neutral dynamics. Clonotypes leave the thymus as a family of eight cells.

Let us consider $G_1(z, t)$. Using the definition (2.2) we know that $G_1(z, 0) = z$. Next, consider a time Δt , short enough that the possibility of two events occurring in the time interval $(0, \Delta t)$ can be neglected. One cell, present at the beginning of the interval, survives with probability $1 - 2\mu\Delta t$, dies with probability $\mu\Delta t$, or divides with probability $\mu\Delta t$. Thus

$$G_1(z, \Delta t) = (1 - 2\mu\Delta t)z + \mu\Delta t(z^0 + z^2),$$

and

$$G_1(z, t + \Delta t) = (1 - 2\mu\Delta t)G_1(z, t) + \mu\Delta t(1 + G_1(z, t)^2).$$

We conclude that $G_1(z, t)$ satisfies

$$\frac{\partial}{\partial t}G_1(z, t) = \mu - 2\mu G_1(z, t) + \mu G_1^2(z, t). \quad (2.4)$$

The solution of (2.4) is

$$G_1(z, t) = \frac{\mu t + z(1 - \mu t)}{1 + \mu t - z\mu t}.$$

Thus, using (2.3),

$$p_0(t) = G_n(0, t) = \left(\frac{\mu t}{1 + \mu t}\right)^n. \quad (2.5)$$

The survival function, $S(t) = 1 - p_0(t)$, shown in *Fig. 2*, is the probability that $n_i(t) > 0$; in other words, the fraction of clonotypes that survive until t . Notice that extinction is certain because $p_0(t) \rightarrow 0$ as $\mu t \rightarrow \infty$ (86). The values of μ shown in *Fig. 2* are in the range of estimates for naïve and memory T cell subsets in an adult human.

While the approximation of the survival function (2.5) is useful, the neutral model has an important drawback: the mean lifetime of a clonotype is infinite. Thus, the assumption cannot be used to calculate a steady state mean number of clonotypes in the periphery, which would be equal to the product of the rate of export of new clonotypes from the thymus and their mean lifetime.

2.3 Global competition and thymic pressure

In order to consider the situation in which both thymic export and peripheral division contribute to maintaining homeostasis, we consider a population of cells, each with a label corresponding to its TCR clonotype,

competing on an equal footing for a public or “global” resource. The dynamics is governed by a combination of thymic production, division in the periphery and cell death. In this section, let us suppose that that cell division depends on access to a non-specific resource, such as IL-7, that is produced in the whole animal at a constant average rate. Thus, there is a well-defined mean number of T cell divisions per unit time, denoted γ . The T cells circulate through the body, ready for their dose of resources. We suppose that, in any small time interval, each living cell is equally likely to receive a signal causing one round of division. We continue with the simplest hypotheses that each T cell, independently, has a death rate μ .

The total number of T cells has a stationary mean, maintained by a balance between death, on one hand, and thymic production and division, on the other. The total number of new cells produced in a short time interval of duration Δt is $(\gamma + \theta n_\theta)\Delta t$. At steady state, this is balanced by the mean number of deaths in Δt , $\mu n(t)\Delta t$. Thus, the steady state mean total number of cells is

$$N_{\text{ss}} = \frac{\gamma + \theta n_\theta}{\mu} . \quad (2.6)$$

To give a mathematical description of the fate of a clonotype in the periphery, we need to consider $n_i(t)$, the number of cells of clonotype i at time t . The probability that one cell of the clonotype dies between t and $t + \Delta t$ is $\mu n_i(t)\Delta t$. The probability that one cell of type i divides between t and $t + \Delta t$ is $\lambda_i(t)\Delta t$ where

$$\lambda_i(t) = \gamma \frac{n_i(t)}{N(t)} . \quad (2.7)$$

In other words, the fraction of the global resource received by clonotype i at time t is equal to its weight in the repertoire. As the total number of cells, $N(t)$, is large and fluctuates about the mean value N_{ss} , we may use the approximation $N(t) \simeq N_{\text{ss}}$. In this approximation, the integer $n_i(t)$ follows a birth-and-death process (40) with death rate μn_i and birth rate

$$\lambda_i(t) \simeq \frac{n_i(t)}{N_{\text{ss}}} = \frac{\mu}{1 + \alpha} n_i(t) ,$$

where

$$\alpha = \frac{\theta n_\theta}{\gamma} , \quad (2.8)$$

is the non-dimensional parameter that measures the strength of thymic production versus peripheral division. The effect of the thymus on the periphery is that of a competitive “pressure” that results in the mean per-cell division rate being less than the per-cell death rate. We do not call this model “neutral” because, although all cells in the periphery follow the same rules, the division rate is a function of the status of the whole repertoire, not simply equal to the death rate.

Under these assumptions, of global competition and thymic pressure, we can calculate the distribution of clonal (family) lifetimes. We say that T_i is the clonal lifetime if n_θ cells of type i exit the thymus at age t_0 and the last cell of type i dies at age $t_0 + T_i$. The probability that the clonal lifetime is greater than t ,

$$S(t) = 1 - \left(\frac{e^{(\mu_i - \lambda_i)t} - 1}{e^{(\mu_i - \lambda_i)t} - \frac{\lambda_i}{\mu_i}} \right)^{n_\theta} , \quad (2.9)$$

is plotted in *Fig. 3*.

With thymic output and peripheral division constant then, as well as a mean steady state number of cells, there is a well-defined mean steady state number of clonotypes, N^* , equal to the product of θ and the mean lifetime of a clonotype in the periphery. In the limit $n_\theta \alpha \rightarrow 0$,

$$N^* \rightarrow N_{\text{ss}} \alpha [1 - \gamma_E - \log(n_\theta \alpha)] , \quad (2.10)$$

where $\gamma_E = 0.577\dots$ is the Euler-Mascheroni constant (52).

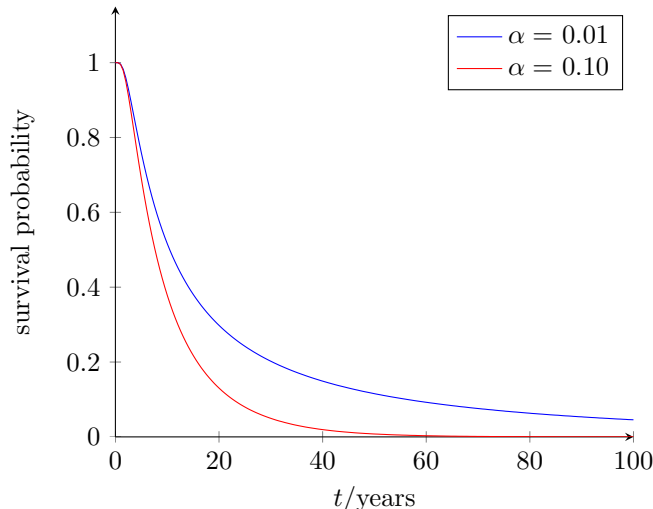


Figure 3: Survival of clonotypes in the periphery with competition for a non-specific resource, using (2.9). The parameter α measures the strength of the thymus relative to peripheral cell division. We have chosen $\mu = 1 \text{ year}^{-1}$ and $n_\theta = 8$.

2.3.1 Diffusion approximation and clonal size distribution

Under a diffusion process approximation of the birth-and-death process, the integer $n_i(t)$ is approximated by the real-valued stochastic process \mathbf{X}_t that satisfies the following stochastic differential equation (52)

$$d\mathbf{X}_t = -\alpha\mu\mathbf{X}_t dt + \sqrt{2\mu\mathbf{X}_t} d\mathbf{W}_t. \quad (2.11)$$

When $\alpha = 0$, the survival probability of this process is $S(t) = 1 - \exp(-\frac{n_0}{\mu t})$ (52), which is similar to (2.5). With the diffusion approximation, (2.11), it is possible to calculate the distribution of clonal sizes which, assuming a homeostatic steady state, is equivalent to the distribution averaged over the lifetime of one clone. The corresponding quantity for \mathbf{X}_t is the occupation density of the process \mathbf{X}_t (91). The function

$L(y, n_0)$ = occupation density at y given that the process starts at n_0 and is absorbed at 0,

satisfies

$$\mu x \frac{d^2}{dx^2} L(y, x) - \alpha \mu x \frac{d}{dx} L(y, x) = -\delta(x - y) \quad L(y, 0) = 0. \quad (2.12)$$

(The mean lifetime of a clonotype is the integral of $L(y, n_0)$ with respect to y .) The solution of (2.12) is

$$L(y, n_0) = \begin{cases} \frac{1}{\alpha\mu} \frac{1 - e^{-\alpha y}}{y} & y < n_0 \\ \frac{e^{\alpha n_0} - 1}{\alpha\mu} \frac{e^{-\alpha y}}{y} & y \geq n_0. \end{cases} \quad (2.13)$$

That is, the distribution of family sizes in the repertoire is predicted to be a function that decreases in a faster-than-exponential way.

2.4 Specific competition and thymic pressure

In our final example, we consider competition for resources that are TCR-specific. An enormous variety of self-pMHC epitopes is available to circulating T cells. As a result of the random recombination and selection processes in the thymus, any one T cell will only recognise a small fraction of those available. Let us assume that the number of different pMHC in the body is a very large, but fixed, number M (52) and that each pMHC

subset acts as a resource that is equally likely to be consumed by each of the cells capable of recognising it. Let us also assume that each pMHC has probability p of being recognised by a randomly-chosen TCR. The value of p is estimated to be 10^{-6} (66, 63, 105). The model can be thought of as a massive ecosystem of species (57, 34, 85, 94), competing for niches, subject to pressure from new species continuously exported from the thymus. If thymic export is constant, then a homeostatic steady-state is attained in which extinction of existing species in the periphery balances the production of new species in the thymus. Remarkably, the single-clonotype dynamics in such a system of large-scale competition is similar to that of global competition, in that the division rate is scaled by the factor $(1 + \alpha)^{-1}$ and the steady-state number of clonotypes satisfies (2.10). This simplicity is due to the cross-reactivity being “unfocussed” in the sense of Mason (56).

In the model just described, the total number of cells in the system has a well-defined mean. De Greef *et al.* devised a models where the total number of cells is fixed (28, 19). To implement this computationally, each time the thymus produces n_θ new cells, the same number of cells, chosen at random, are removed from the repertoire. In both types of model, the following illuminating equality holds (28, 19):

$$\theta = \mu F_1. \quad (2.14)$$

The quantity on the left-hand-side is the rate of export of new clonotypes from the thymus to the periphery. The first factor in the product on the right-hand-side is μ , the death rate per cell (equal to the inverse of the mean lifetime of a cell). Increasingly accurate estimates of μ have become available in the past few years (9, 17, 18, 1). The second factor is the number of unicellular clonotypes in the repertoire. That is, the number of TCRs that are present on only one cell in the periphery. The maximum possible value of F_1 is simply S , the number of cells in the repertoire. The equality $F_1 = S$ holds in the extreme scenario that all peripheral clonotypes consist of one cell only, otherwise $F_1 < S$.

To estimate θ , we need to divide the rate of thymic production, in cells per unit time, for which good estimates exist (20, 6), by n_θ the number of cells per clonotype at the moment of release from the thymus, which is not well established (32). Putting aside the age-dependence of the thymus, we may use the estimates $\theta = 2.5 \times 10^6 \text{ month}^{-1}$ and $\mu = 1 \text{ month}^{-1}$ for naïve CD4^+ T cells in an adult mouse, in which case $F_1 = 2.5 \times 10^6$. This suggests that at least one quarter of naïve CD4^+ T-cell clonotypes in the periphery of a mouse consist of only one cell. For comparison, we may estimate that $\theta = 10^9 \text{ year}^{-1}$ and $\mu = 0.5 \text{ year}^{-1}$ for naïve CD4^+ T cells in an adult human, and deduce that $F_1 = 2 \times 10^9$. On the one hand, this value is comfortably below the total number of naïve CD4^+ T cells in an adult human, so that it is not true that most clonotypes consist of only one cell. On the other hand, it suggests that the number of distinct TCR in the repertoire is greater than 10^{10} (52).

In *Fig. 4*, we illustrate typical theoretical life histories of a clonotype in a mouse (left) and human (right). In the former case, the simplest assumption of no peripheral division is used. In the latter case, the slight imbalance between birth and death rates ensures that the mean clonal lifetime is finite, but the clonal size may grow from its initial value before, eventually, extinction occurs.

2.5 Computational models

How should a computational model of the T cell repertoire over the lifetime of a human or mouse be constructed? Computational models provide a testing ground for provisional classification and lineage schemes for T cells. If the number of subsets is not too large, models with one ODE per subset are easily solved. Then, Bayesian statistical frameworks can be used to compare models and estimate parameter values. Should, for example, recent thymic emigrants be treated as a separate subset from naïve cells (95)? The fact that the $\text{CD4}^+:\text{CD8}^+$ ratio in the periphery of an adult mouse is close to 1, although the $\text{SP4}:\text{SP8}$ ratio in the thymus is typically 4:1 (83, 12) is consistent with the hypothesis that CD4^+ differ from CD8^+ T cells in this respect (98). Are death and division rates constant over the lifetime of a mouse or human, or do they change as part of the ageing process (103, 98, 33, 75)?

In *Fig. 5* we show a type of stochastic model of an ageing T cell repertoire with multiple competing clonotypes, with mean number of cells 10^{-5} times smaller than the number of CD4^+ T cells in an adult human. For the Figure, we have modified the stochastic dynamics, developed in (52) and described in 2.4, to include an age-dependent thymic export rate $\theta(t) = \theta_0 \sqrt{\nu t} e^{-\nu t}$. (The per-cell death and division rates, however, do not decrease with age.) The number of surviving clonotypes (those with a least one living cell) at age t peaks later than the peak of thymic production. In later decades, the total number of cells remains

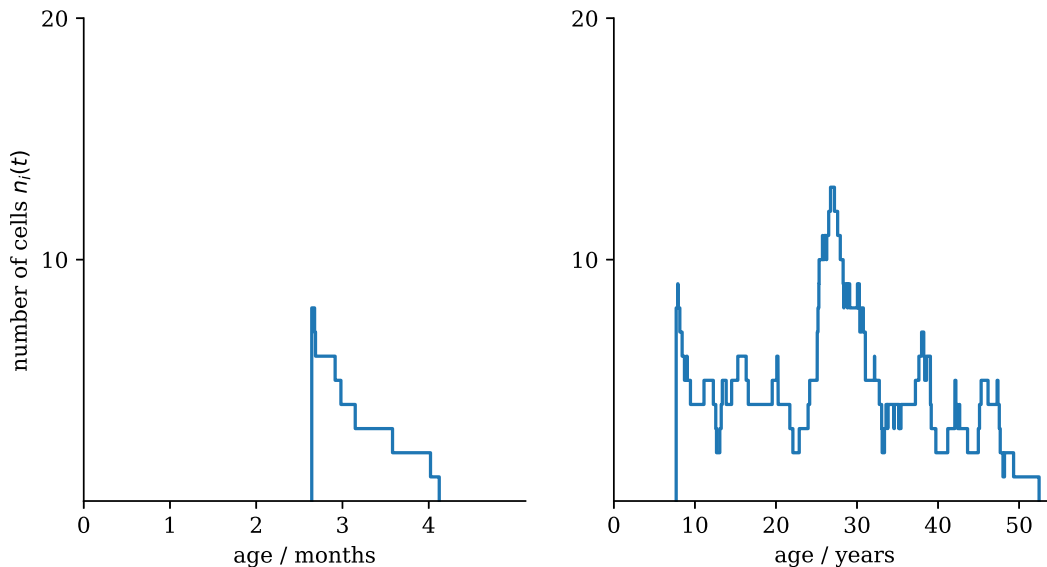


Figure 4: Realisations of stochastic models for the number of cells of one clonotype in the periphery, as a function of age of the host. The life history of a clonotype starts with exit from the thymus, and ends with extinction. In both examples, the clonotype enters the periphery as a family of eight cells. On the left, using the simplest assumption to model a mouse host: peripheral division is zero, per-cell death rate $\mu = 1$ per month. On the right, using the global competition model of adult human homeostasis, with ratio of thymic output to peripheral division, $\alpha = 0.04$ and per-cell death rate $\mu = 0.25$ per year.

relatively constant (at about 4×10^6) but the number of surviving clonotypes declines with age, about ten-fold from 20 to 100 years old.

Whether a model is deterministic or stochastic, treating all cells within a population as identical or statistically identical comes at the cost of ignoring cell-to-cell heterogeneity and particular cell-cell interactions. In agent-based models, on the other hand, we can choose the fundamental objects to be individual cells rather than populations of cells based on an *a priori* classification. Some important aspects of T cell homeostasis, such as quorum sensing mechanisms by which individual cells respond to the IL-2 in their vicinity (3), and the altruistic hypothesis, where IL-7 receptor expression is reduced on T cells that have received IL-7 signals (72), can only be modelled if differences from cell to cell within a clonotype are included (15). It is possible to introduce such cell-to-cell heterogeneity in computational models where individual cells are agents and receptor-expression levels are attributes (12).

3 How many TCR clonotypes are maintained?

How many T cells are there in a mouse and in a human? It is unusual to obtain more than blood samples from healthy humans, so the typical estimates of $10^{11} - 10^{12}$ T cells are based on extrapolation. On the other hand, it is possible to obtain all major organs from mice of various ages, albeit only once per mouse. Accurate cell counts are limited by the difficulty of extracting and separating T cells whilst avoiding, or at least keeping a strict account of, losses at various stages of purification (26, 32).

Once the total number of T cells, or of a subset defined by naïve or memory surface markers, is established, we may ask how many TCR clonotypes there are, in each subset, in humans, mice and other mammals (47, 8, 14, 93). Estimates of the number of different TCRs that could, in principle, be produced by VDJ gene rearrangement in the thymus (50, 82, 106, 65) far exceed the total number of T cells in a mouse or human body (56, 58). With the latest developments in sequencing techniques, is a direct count of the number of clonotypes in a mammal feasible? Given that it is possible to obtain a list of TCR α or β chain sequences

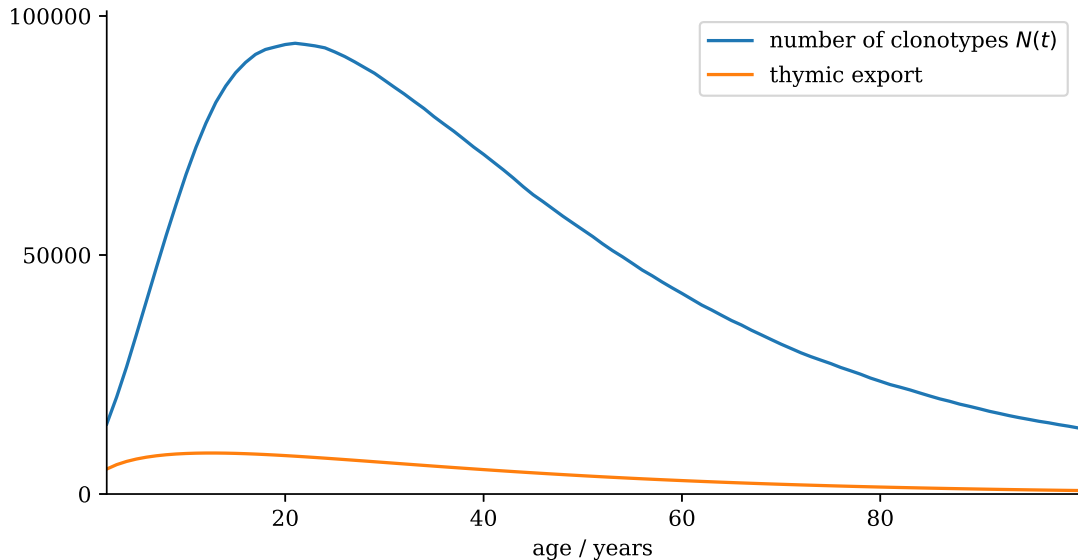


Figure 5: Number of surviving clonotypes as a function of time in a computational model of the naïve CD4⁺ repertoire with specific competition and ageing thymus. Each clonotype follows the stochastic dynamics described in 2.4, but with an age-dependent thymic export rate that peaks at 10 years. The parameters are $\nu = 0.04 \text{ year}^{-1}$, $n_\theta = 8$, $\mu = 1 \text{ year}^{-1}$, $M = 10^4$ and $p = 0.01$.

from a pool of more than a million T cells from one individual (74), a good census of the tens of millions of cells in the T-cell repertoire of a mouse may be possible in the near future. However, important practical limitations are evident from the observation that the majority of sequences typically obtained are only present as a single copy. Can such genuine instances be distinguished from errors? Similarly, if a sequence is present as two copies, is it possible to establish whether there were two cells in the pool of cells, or two RNA copies of the same TCR chain in one cell? Even at the large end of the spectrum, the distribution of family sizes is difficult to evaluate reliably because of the possibility of inhomogeneities in the PCR processes used (67).

Single-cell measurements, where PCR and sequencing are performed on one cell at a time, avoid these problems and biases, and can identify paired α and β chains (16, 88). However, their expense means that only hundreds of cells are usually sequenced from a single mammal, and estimates of diversity must therefore rely on mathematical extrapolation from small samples (99, 76, 81, 48, 62). Assuming that single-cell techniques can overcome the practical problems associated with high-throughput sequencing, the remaining problem is the mathematical one of extrapolation from small samples.

Data available to date suggest that apart, perhaps, from a few large clonotypes, the repertoire of naïve T cells in a mouse or human consists of many clonotypes with small clonal sizes. Thus, given current sample sizes, the majority of TCRs found in typical samples will be found on only one cell per sample (67). We must conclude that most TCR clonotypes are unobserved. These observations, no doubt a source of frustration to the experimentalist, are the basis of convenient approximations for mathematicians. The mathematical problem is illustrated in *Fig. 6*. We can imagine that we can take a sample of balls from a large box filled with balls of many colours. The total number of balls is known, but the number of different colours is not. What can be deduced from the colours of the balls in the sample?

Suppose that an unbiased random sample of m cells is taken, and sequenced, from a total repertoire of S cells. Let us start with the probability that one particular cell in the repertoire is part of the sample (32, 23):

$$q = \frac{m}{S}.$$

In naïve repertoire single-cell sequencing $q \ll 1$, because m , approximately 10^2 - 10^3 , is much smaller than S , approximately 10^7 (mice) or 10^{11} (humans). We seek a relationship between n_i , the number of cells of

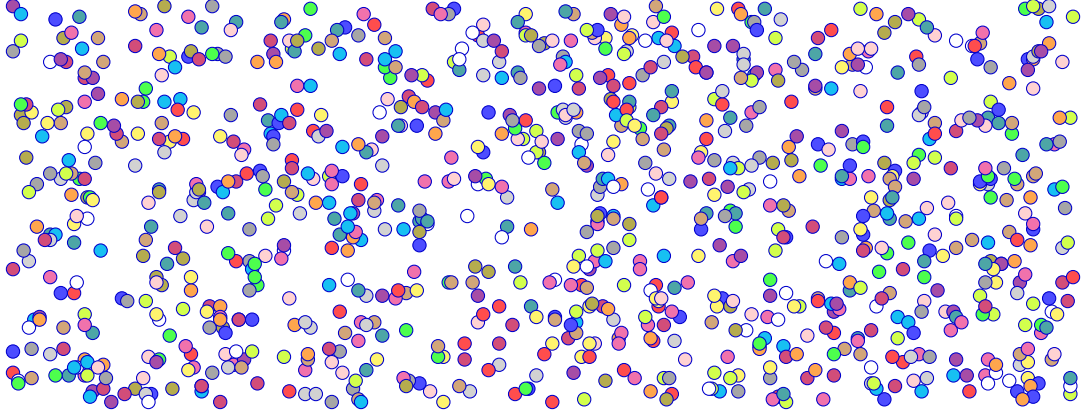


Figure 6: One circle represents one cell. Suppose that the total number of cells is known. The mathematical problem is to deduce the total number of different colours in the entire set, and distribution of number of cells per colour, based on the colours found in small sample of cells.

type i in the repertoire, and s_i , the number of cells of type i in the sample. The statement that most TCRs are unobserved and that most TCRs present in the sample are found on only one cell in the sample, is that $qn_i \ll 1$ for typical values of n_i . This is convenient to a mathematician because it is a good approximation to say that each of the n_i cells, independently, has probability q of being part of the sample. Consequently,

$$\text{Prob}[s_i = k | n_i = n] \simeq \frac{n!}{(n-k)!k!} q^k \quad k = 0, \dots, n.$$

The probability that there is one cell of type i in the sample is $\text{Prob}[s_i = 1 | n_i = n] \simeq nq$; the probability that there are two is $\text{Prob}[s_i = 2 | n_i = n] \simeq \frac{1}{2}n(n-1)q^2$. That is, we expect to find about $2/q$ singletons (only one cell with a given TCR present in the sample) for every repeat (two or more cells with the same TCR in the sample).

Of course, n_i is not known in advance. We may instead think of n_i as being drawn from the distribution of clonal family sizes in the repertoire with mean \bar{n} . In terms of the coloured balls, the number of different clonotypes in the repertoire (number of different colours) is equal to S/\bar{n} . We therefore require a relationship between the distribution of values of n_i in the repertoire and the distribution of values of s_i in the sample. The relationship is most elegantly expressed as a relationship between probability generating functions. For the distribution of clonal sizes, we define $\phi(z) = \sum_{n=1}^{\infty} \text{Prob}[n_i = n] z^n$ for $0 \leq z \leq 1$. For the distribution of

values of s_i , $\Phi(z) = \sum_{k=0}^{\infty} \text{Prob}[s_i = k] z^k$. Assuming that $qn_i \ll 1$, the relationship between $\phi(z)$ and $\Phi(z)$ is (32, 23):

$$\Phi(z) = \phi(1 - q + qz).$$

The probability that sequencing m single cells from a sample yields m different sequences is

$$\text{Prob}[\text{no repeats}] = e^{-r}, \quad \text{where } r = q^2 \sum_i n_i(n_i - 1).$$

If the n_i are chosen from a distribution of clonal family sizes with mean \bar{n} and mean square $\mathbb{E}(n^2)$ then

$$r = \frac{m^2}{2S} \left(\frac{\mathbb{E}(n^2)}{\bar{n}} - 1 \right).$$

The exponent r is a product of the factor $m^2/2S$, that does not depend on the distribution of clonal family sizes, and the factor $\mathbb{E}(n^2)/\bar{n} - 1$, that does. In *Fig. 7* we consider a power-law repertoire. We say that the

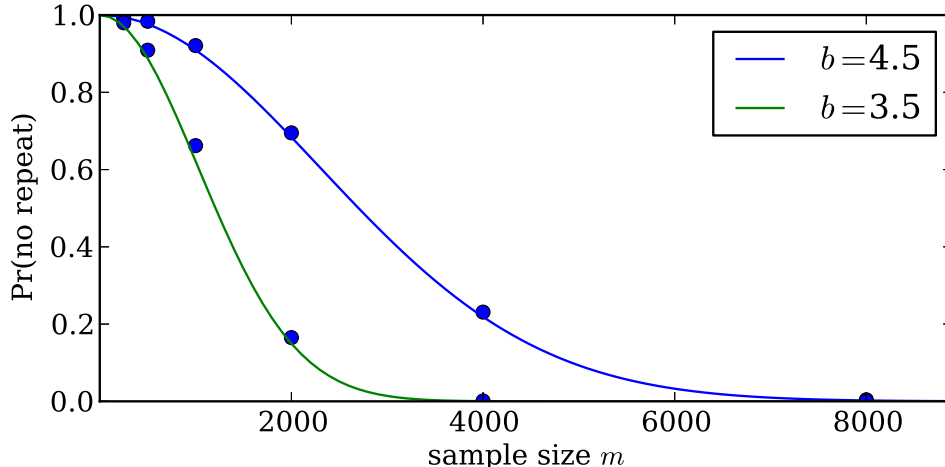


Figure 7: The probability that all the TCRs in a sample are unique. The dots are obtained by constructing repertoires of a million cells each ($S = 10^6$) and samples of size m *in silico*, using a power-law distribution of clonal family sizes with exponent b . The lines are the formula $\text{Prob}[\text{no repeat}] = e^{-r}$ with r given by (3.1).

distribution is “power-law” if the number of clonotypes consisting of n cells is proportional to n^b . If $b > 3$ then

$$r = \frac{m^2}{2S} \left(\frac{\zeta(b-2)}{\zeta(b-1)} - 1 \right), \quad (3.1)$$

where ζ is the Riemann zeta function.

Let us end with an example of what can be expected from single-cell sequencing of samples of T cells. Suppose that all the $\text{CD8}^+ \text{CD44}^-$ T cells are purified from the spleen of a mouse, about 2.9×10^6 cells (32). Suppose that the TCRs of each of 10000 individual cells are sequenced. What would we expect to see? An *in silico* simulation is shown in Fig. 8. Two hypotheses for a repertoire, with $S = 2.9 \times 10^6$ cells, are shown in the left column:

1. A power-law distribution of clonal sizes (67). In the Figure, we use the exponent $b = 2.5$, which gives a mean clonal family size of $\bar{n} = 1.95$ cells.
2. A geometric distribution of clonal sizes. That is, the ratio of the number of clonotypes made up of $k+1$ cells to the number of clonotypes made up of k cells is equal to $(1 - \frac{1}{\bar{n}})$ (32).

The sequences in a sample of m cells can be classified by the number of occurrences of each TCR. We write

$$m = m_1 + 2m_2 + 3m_3 + \dots,$$

where m_k is the number of sequences that are present k times. For example, if 20 TCR are found on three cells, 70 on two cells and all other TCRs only found on one cell, then $m_1 = 9800$, $m_2 = 70$ and $m_3 = 20$. The mean values of m_k are shown as a function of k in the right column of Fig. 8.

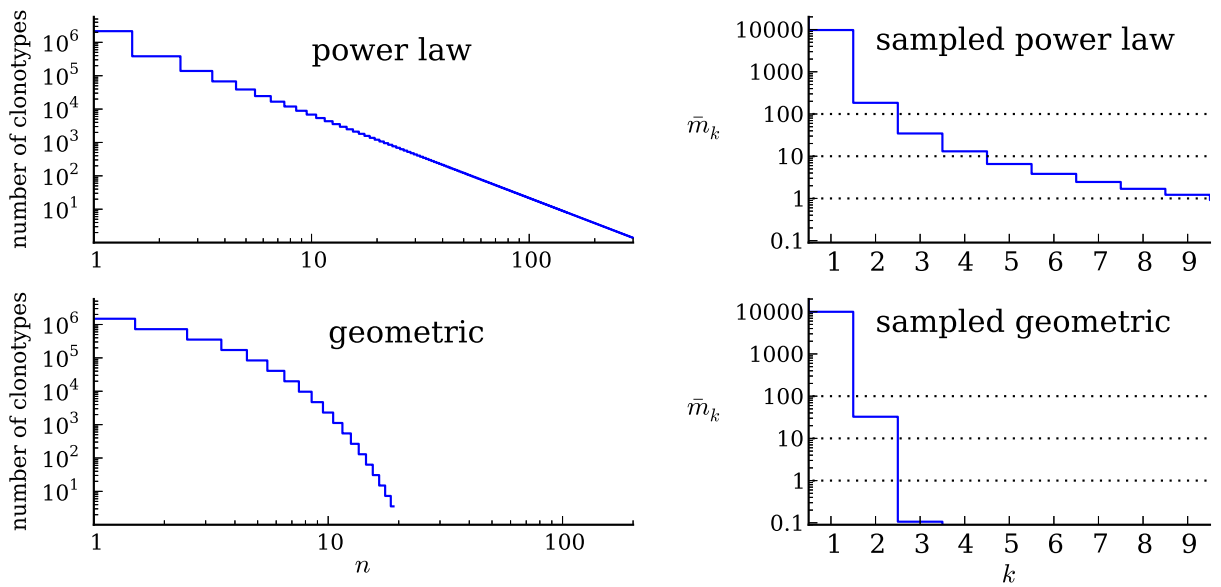


Figure 8: Simulated sampling. The number of cells in the repertoire, S , is chosen to be equal to 2900000, similar to the mean number of $CD8^+ CD44^-$ T cells in the spleen of a mouse. Two scenarios for the distribution of numbers of cells per clonotype n are shown on the left. The size of the hypothesised single-cell sequencing sample is taken to be $m = 10000$ cells, somewhat larger than current capabilities. On the right, the calculated mean numbers of singletons (TCRs only present on one cell in the sample), doublets (two cells), k -lets is shown as a function of k .

4 Discussion

Our understanding of the dynamics of an immune system repertoire of more than 10^{11} interacting cells is, not surprisingly, incomplete. Individual human T cells may live for years; following their entire histories *in vivo* as they undergo migration, activation and differentiation processes is not possible. Most knowledge comes instead from *in vitro* experiments where samples of cells are confined and stimulated, and from adoptive transfer and imaging experiments involving genetically-modified mice (101, 35, 59, 42). Mathematics provides a systematic language to express dogma, theory and conjecture, so that assumptions can be tested and improved, parameters estimated and predictions compared with data. Progress is also being made in constructing *in silico* models which, by emulating *in vitro* models and *in vivo* processes, could reduce and replace animal testing.

Acknowledgements We will always be grateful for the privilege of Robin Callard’s friendship and scientific insight.

References

- [1] R. Ahmed, L. Westera, J. Drylewicz, M. Elemans, Y. Zhang, E. Kelly, R. Reljic, K. Tesselaar, R. J. De Boer, D. C. Macallan, et al. Reconciling estimates of cell proliferation from stable isotope labeling experiments. *PLoS Computational Biology*, 11(10):e1004355, 2015.
- [2] A. Almeida, N. Legrand, M. Papiernik, and A. Freitas. Homeostasis of peripheral CD4⁺ T cells: IL-2r *alpha* and IL-2 shape a population of regulatory cells that controls CD4⁺ T cell numbers. *Journal of Immunology*, 169(9):4850, 2002.
- [3] A. Almeida, I. Amado, J. Reynolds, J. Berges, G. Lythe, C. Molina-París, and A. Freitas. Quorum-sensing in CD4⁺ T cell homeostasis: a hypothesis and a model. *Frontiers in Immunology*, 3, 2012.
- [4] S. Ariotti, J. B. Beltman, G. Chodaczek, M. E. Hoekstra, A. E. van Beek, R. Gomez-Eerland, L. Ritsma, J. van Rheenen, A. F. Marée, T. Zal, et al. Tissue-resident memory CD8⁺ T cells continuously patrol skin epithelia to quickly recognize local antigen. *Proceedings of the National Academy of Sciences*, 109(48):19739–19744, 2012.
- [5] I. Bains, R. Antia, R. Callard, and A. Yates. Quantifying the development of the peripheral naive CD4⁺ T-cell pool in humans. *Blood*, 113(22):5480, 2009.
- [6] I. Bains, R. Thiebaut, A. Yates, and R. Callard. Quantifying thymic export: Combining models of naive T cell proliferation and TCR excision circle dynamics gives an explicit measure of thymic output. *Journal of Immunology*, 183(7):4329, 2009.
- [7] L. K. Beura, S. E. Hamilton, K. Bi, J. M. Schenkel, O. A. Odumade, K. A. Casey, E. A. Thompson, K. A. Fraser, P. C. Rosato, A. Filali-Mouhim, et al. Normalizing the environment recapitulates adult human immune traits in laboratory mice. *Nature*, 532(7600):512, 2016.
- [8] J. N. Blattman, R. Antia, D. J. Sourdive, X. Wang, S. M. Kaech, K. Murali-Krishna, J. D. Altman, and R. Ahmed. Estimating the precursor frequency of naive antigen-specific CD8 T cells. *Journal of Experimental Medicine*, 195(5):657–664, 2002.
- [9] J. Borghans and R. de Boer. Quantification of T-cell dynamics: from telomeres to DNA labeling. *Immunological Reviews*, 216:35–47, 2007.
- [10] P. Bousso and E. Robey. Dynamics of CD8⁺ T cell priming by dendritic cells in intact lymph nodes. *Nature Immunology*, 4(6):579, 2003.
- [11] V. R. Buchholz, M. Flossdorf, I. Hensel, L. Kretschmer, B. Weissbrich, P. Gräf, A. Verschoor, M. Schiemann, T. Höfer, and D. H. Busch. Disparate individual fates compose robust CD8⁺ T cell immunity. *Science*, 340(6132):630–635, 2013.
- [12] M. Castro, G. Lythe, and C. Molina-París. The T cells in an ageing virtual mouse. In *Stochastic Processes, Multiscale Modeling, and Numerical Methods for Computational Cellular Biology*, pages 127–140. Springer, 2017.

- [13] D. K. Choo, K. Murali-Krishna, R. Anita, and R. Ahmed. Homeostatic turnover of virus-specific memory CD8 T cells occurs stochastically and is independent of CD4 T cell help. *Journal of Immunology*, 185(6):3436–3444, 2010.
- [14] S. M. Ciupe, B. H. Devlin, M. L. Markert, and T. B. Kepler. Quantification of total T-cell receptor diversity by flow cytometry and spectratyping. *BMC Immunology*, 14(1):1–12, 2013.
- [15] J. W. Cotari, G. Voisinne, O. E. Dar, V. Karabacak, and G. Altan-Bonnet. Cell-to-cell variability analysis dissects the plasticity of signaling of common γ chain cytokines in t cells. *Sci. Signal.*, 6(266):ra17–ra17, 2013.
- [16] T. Cukalac, W.-T. Kan, P. Dash, J. Guan, K. M. Quinn, S. Gras, P. G. Thomas, and N. L. La Gruta. Paired TCR $\alpha\beta$ analysis of virus-specific CD8⁺ T cells exposes diversity in a previously defined “narrow” repertoire. *Immunology and Cell Biology*, 2015.
- [17] R. De Boer, A. Perelson, and R. Ribeiro. Modelling deuterium labelling of lymphocytes with temporal and/or kinetic heterogeneity. *Journal of The Royal Society Interface*, 9(74):2191–2200, 2012.
- [18] R. J. De Boer and A. S. Perelson. Quantifying T lymphocyte turnover. *Journal of Theoretical Biology*, 327:45–87, 2013.
- [19] P. C. de Greef, T. Oakes, B. Gerritsen, M. Ismailb, J. M. Heather, R. Hermsen, B. Chain, and R. J. de Boer. VDJ recombination shapes the frequency distribution of the naive T-cell receptor repertoire. 2018.
- [20] I. den Braber, T. Mugwagwa, N. Vriskoop, L. Westera, R. Mögling, A. Bregje de Boer, N. Willems, E. H. Schrijver, G. Spierenburg, K. 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.
- [21] P. Doherty, J. Riberdy, and G. Belz. Quantitative analysis of the CD8⁺ T-cell response to readily eliminated and persistent viruses. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 355(1400):1093–1101, 2000.
- [22] O. Feinerman, G. Jentsch, K. Tkach, J. Coward, M. Hathorn, M. Sneddon, T. Emonet, K. Smith, and G. Altan-Bonnet. Single-cell quantification of IL-2 response by effector and regulatory T cells reveals critical plasticity in immune response. *Molecular Systems Biology*, 6(1), 2010.
- [23] M. Ferrarini, C. Molina-París, and G. Lythe. Sampling from T cell receptor repertoires. In *Modeling Cellular Systems*, pages 67–79. Springer, 2017.
- [24] P. J. Fink. The biology of recent thymic emigrants. *Annual Review of Immunology*, 31:31–50, 2013.
- [25] T. J. Fry and C. L. Mackall. The many faces of IL-7: from lymphopoiesis to peripheral T cell maintenance. *Journal of Immunology*, 174(11):6571–6576, 2005.
- [26] V. V. Ganusov and R. J. De Boer. Do most lymphocytes in humans really reside in the gut? *Trends in Immunology*, 28(12):514–518, 2007.
- [27] V. V. Ganusov, J. A. Borghans, and R. J. De Boer. Explicit kinetic heterogeneity: mathematical models for interpretation of deuterium labeling of heterogeneous cell populations. *PLoS Computational Biology*, 6(2):e1000666, 2010.
- [28] B. Gerritsen. *Sequencing, analyzing, and modeling small samples from large T cell repertoires*. PhD thesis, Utrecht University, 2018.
- [29] B. Gerritsen and A. Pandit. The memory of a killer T cell: models of CD8⁺ T cell differentiation. *Immunology & Cell Biology*, 94(3):236–241, 2016.
- [30] A. W. Goldrath and M. J. Bevan. Low-affinity ligands for the TCR drive proliferation of mature CD8⁺ T cells in lymphopenic hosts. *Immunity*, 11:183–190, 1999.
- [31] A. W. Goldrath and M. J. Bevan. Selecting and maintaining a diverse T-cell repertoire. *Nature*, 402(6759):255–262, 1999.
- [32] P. Gonçalves, M. Ferrarini, C. Molina-Paris, G. Lythe, F. Vasseur, A. Lim, B. Rocha, and O. Azogui. A new mechanism shapes the naïve CD8⁺ T cell repertoire: The selection for full diversity. *Molecular Immunology*, 85:66–80, 2017.

- [33] G. Gossel, T. Hogan, D. Cownden, B. Seddon, and A. J. Yates. Memory CD4 T cell subsets are kinetically heterogeneous and replenished from naive T cells at high levels. *Elife*, 6:e23013, 2017.
- [34] Y. Hao, N. Legrand, and A. Freitas. The clone size of peripheral CD8 T cells is regulated by TCR promiscuity. *Journal of Experimental Medicine*, 203(7):1643, 2006. ISSN 0022-1007.
- [35] E. Hawkins, J. Markham, L. McGuinness, and P. Hodgkin. A single-cell pedigree analysis of alternative stochastic lymphocyte fates. *Proceedings of the National Academy of Sciences*, 106(32):13457–13462, 2009.
- [36] T. Hogan, G. Gossel, A. J. Yates, and B. 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.
- [37] S. Jameson, K. Hogquist, and M. Bevan. Positive selection of thymocytes. *Annual Review of Immunology*, 13(1):93–126, 1995.
- [38] M. Jenkins, H. Chu, J. McLachlan, and J. Moon. On the composition of the preimmune repertoire of T cells specific for peptide-major histocompatibility complex ligands. *Annual Review of Immunology*, 28:275–294, 2009.
- [39] P. L. Johnson, A. J. Yates, J. J. Goronzy, and R. Antia. Peripheral selection rather than thymic involution explains sudden contraction in naive CD4 T-cell diversity with age. *Proceedings of the National Academy of Sciences*, 109(52):21432–21437, 2012.
- [40] S. Karlin and H. M. Taylor. *A second course in stochastic processes*. Academic, 1981.
- [41] M. Kimmel and D. E. Axelrod. *Branching Processes in Biology*. Springer, New York, 2002.
- [42] I. Kinjyo, J. Qin, S.-Y. Tan, C. J. Wellard, P. Mrass, W. Ritchie, A. Doi, L. L. Cavanagh, M. Tomura, A. Sakaue-Sawano, L. L. Cavanagh, M. Tomura, A. Sakaue-Sawano, O. Kanagawa, A. Miyawaki, P. D. Hodgkin, and W. Weninger. Real-time tracking of cell cycle progression during CD8⁺ effector and memory T-cell differentiation. *Nature Communications*, 6, 2015.
- [43] L. Klein, M. Hinterberger, G. Wirnsberger, and B. Kyewski. Antigen presentation in the thymus for positive selection and central tolerance induction. *Nature Reviews Immunology*, 9(12):833–844, 2009.
- [44] P. Koenen, S. Heinzel, E. M. Carrington, L. Happo, W. S. Alexander, J.-G. Zhang, M. J. Herold, C. L. Scott, A. M. Lew, A. Strasser, et al. Mutually exclusive regulation of T cell survival by IL-7R and antigen receptor-induced signals. *Nature Communications*, 4:1735, 2013.
- [45] M. F. Krummel, F. Bartumeus, and A. Gérard. T cell migration, search strategies and mechanisms. *Nature Reviews Immunology*, 16(3):193, 2016.
- [46] E. Ladi, X. Yin, T. Chtanova, and E. Robey. Thymic microenvironments for T cell differentiation and selection. *Nature Immunology*, 7(4):338–343, 2006.
- [47] R. Langman and M. Cohn. The ET (elephant-tadpole) paradox necessitates the concept of a unit of B-cell function: the protecton. *Molecular Immunology*, 24(7):675–697, 1987.
- [48] D. Laydon, C. Bangham, and B. Asquith. Estimating T-cell repertoire diversity: limitations of classical estimators and a new approach. *Philosophical Transactions of the Royal Society B*, 370, 2015.
- [49] J. Lewis, A. Walker, H. Castro, A. De Rossi, D. Gibb, C. Giaquinto, N. Klein, and R. Callard. Age and CD4 count at initiation of antiretroviral therapy in HIV-infected children: effects on long-term T-cell reconstitution. *Journal of Infectious Diseases*, 205(4):548–556, 2012.
- [50] M. R. Lieber. Site-specific recombination in the immune system. *FASEB journal*, 5(14):2934–2944, 1991.
- [51] A. Link, T. Vogt, S. Favre, M. Britschgi, H. Acha-Orbea, B. Hinz, J. Cyster, and S. Luther. Fibroblastic reticular cells in lymph nodes regulate the homeostasis of naive T cells. *Nature Immunology*, 8(11):1255–1265, 2007.
- [52] G. Lythe, R. E. Callard, R. L. Hoare, and C. Molina-París. How many TCR clonotypes does a body maintain? *Journal of Theoretical Biology*, 389:214–224, 2016.
- [53] J. N. Mandl, R. Liou, F. Klauschen, N. Vrisekoop, J. P. Monteiro, A. J. Yates, A. Y. Huang, and R. N. Germain. Quantification of lymph node transit times reveals differences in antigen surveillance strategies of naive CD4⁺ and CD8⁺ T cells. *Proceedings of the National Academy of Sciences*, 109(44):18036–18041, 2012.

- [54] J. N. Mandl, J. P. Monteiro, N. Vrisekoop, and R. N. Germain. T cell-positive selection uses self-ligand binding strength to optimize repertoire recognition of foreign antigens. *Immunity*, 2013.
- [55] J. M. Marchingo, A. Kan, R. M. Sutherland, K. R. Duffy, C. J. Wellard, G. T. Belz, A. M. Lew, M. R. Dowling, S. Heinzl, and P. D. Hodgkin. Antigen affinity, costimulation, and cytokine inputs sum linearly to amplify T cell expansion. *Science*, 346(6213):1123–1127, 2014.
- [56] D. Mason. A very high level of crossreactivity is an essential feature of the T-cell receptor. *Immunology Today*, 19(9):395–404, 1998. ISSN 0167-5699.
- [57] S. Merrill, R. De Boer, and A. S. Perelson. Development of the T-cell repertoire: clone size distribution. *Rocky mountain journal of mathematics*, 24(1):214–231, 1994.
- [58] J. Miles, D. Douek, and D. Price. Bias in the $\alpha\beta$ T-cell repertoire: implications for disease pathogenesis and vaccination. *Immunology and Cell Biology*, 89(3):375–387, 2011.
- [59] M. Miller, O. Safrina, I. Parker, and M. Cahalan. Imaging the single cell dynamics of CD4⁺ T cell activation by dendritic cells in lymph nodes. *Journal of Experimental Medicine*, 200(7):847–856, 2004.
- [60] B. Min, G. Foucras, M. Meier-Schellersheim, and W. E. Paul. Spontaneous proliferation, a response of naive CD4 T cells determined by the diversity of the memory cell repertoire. *Proceedings of the National Academy of Sciences*, 101(11):3874–3879, 2004.
- [61] C. Molina-Paris, J. Reynolds, G. Lythe, and M. C. Coles. Mathematical model of naive T cell division and survival IL-7 thresholds. *Frontiers in Immunology*, 4:434, 2013.
- [62] C. Molina-Paris, M. Castro, G. Lythe, and R. Ribeiro. Mathematics in modern immunology. *Interface Focus*, 6:20150093, 2016.
- [63] J. Moon, H. Chu, J. Hataye, A. Pagán, M. Pepper, J. McLachlan, T. Zell, and M. Jenkins. Tracking epitope-specific T cells. *Nature Protocols*, 4(4):565–581, 2009.
- [64] S. N. Mueller and R. N. Germain. Stromal cell contributions to the homeostasis and functionality of the immune system. *Nature Reviews Immunology*, 9(9):618–629, 2009.
- [65] A. Murugan, T. Mora, A. M. Walczak, and C. G. Callan. Statistical inference of the generation probability of T-cell receptors from sequence repertoires. *Proceedings of the National Academy of Sciences*, 109(40):16161–16166, 2012.
- [66] J. Nikolich-Zugich, M. K. Slifka, and I. Messaoudi. The many important facets of T-cell repertoire diversity. *Nature Reviews Immunology*, 4(2):123–132, 2004.
- [67] T. Oakes, J. M. Heather, K. Best, R. Byng-Maddick, C. Husovsky, M. Ismail, K. Joshi, G. Maxwell, M. Noursadeghi, N. Riddell, et al. Quantitative characterization of the T cell receptor repertoire of naive and memory subsets using an integrated experimental and computational pipeline which is robust, economical, and versatile. *Frontiers in Immunology*, 8:1267, 2017.
- [68] J. J. OShea and W. E. Paul. Mechanisms underlying lineage commitment and plasticity of helper CD4⁺ T cells. *Science*, 327(5969):1098–1102, 2010.
- [69] E. Palmer. Negative selection: clearing out the bad apples from the T-cell repertoire. *Nature Reviews Immunology*, 3(5):383–391, 2003.
- [70] E. Palmer and D. Naeher. Affinity threshold for thymic selection through a T-cell receptor–co-receptor zipper. *Nature Reviews Immunology*, 2009.
- [71] M. Palmer, V. Mahajan, J. Chen, D. Irvine, and D. Lauffenburger. Signaling thresholds govern heterogeneity in IL-7-receptor-mediated responses of naive CD8⁺ T cells. *Immunology and Cell Biology*, 89(5):581–594, 2011.
- [72] J. Park, Q. Yu, B. Erman, J. Appelbaum, D. Montoya-Durango, H. Grimes, and A. Singer. Suppression of IL7R α transcription by IL-7 and other prosurvival cytokines: a novel mechanism for maximizing IL-7-dependent T cell survival. *Immunity*, 21(2):289–302, 2004.
- [73] M. Pepper and M. Jenkins. Origins of CD4⁺ effector and central memory T cells. *Nature Immunology*, 131(6):467–471, 2011.

- [74] Q. Qi, Y. Liu, Y. Cheng, J. Glanville, D. Zhang, J.-Y. Lee, R. A. Olshen, C. M. Weyand, S. D. Boyd, and J. J. Goronzy. Diversity and clonal selection in the human T-cell repertoire. *Proceedings of the National Academy of Sciences*, page 201409155, 2014.
- [75] S. Rane, T. Hogan, B. Seddon, and A. 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.
- [76] H. S. Robins, P. V. Campregher, S. K. Srivastava, A. Wacher, C. J. Turtle, O. Kahsai, S. R. Riddell, E. H. Warren, and C. S. Carlson. Comprehensive assessment of T-cell receptor β -chain diversity in $\alpha\beta$ T cells. *Blood*, 114(19):4099–4107, 2009.
- [77] F. Sallusto, D. Lenig, R. Förster, M. Lipp, and A. Lanzavecchia. Two subsets of memory T lymphocytes with distinct homing potentials and effector functions. *Nature*, 402:34–38, 1999.
- [78] T. Sathaliyawala, M. Kubota, N. Yudanin, D. Turner, P. Camp, J. J. Thome, K. L. Bickham, H. Lerner, M. Goldstein, M. Sykes, et al. Distribution and compartmentalization of human circulating and tissue-resident memory T cell subsets. *Immunity*, 2012.
- [79] B. Seddon and R. Zamoyska. Regulation of peripheral T-cell homeostasis by receptor signalling. *Current Opinion in Immunology*, 15(3):321–324, 2003.
- [80] B. Seddon, P. Tomlinson, and R. Zamoyska. Interleukin 7 and T cell receptor signals regulate homeostasis of CD4 memory cells. *Nature Immunology*, 4(7):680–686, 2003.
- [81] N. Sepúlveda, C. Paulino, and J. Carneiro. Estimation of T-cell repertoire diversity and clonal size distribution by Poisson abundance models. *Journal of Immunological Methods*, 353(1):124–137, 2010.
- [82] A. K. Sewell. Why must T cells be cross-reactive? *Nature Reviews Immunology*, 12(9):669–677, 2012.
- [83] C. Sinclair, M. Saini, I. Schim van der Loeff, S. Sakaguchi, and B. Seddon. The long-term survival potential of mature T lymphocytes is programmed during development in the thymus. *Science Signaling*, 4(199):ra77, 2011.
- [84] C. Sinclair, I. Bains, A. J. Yates, and B. Seddon. Asymmetric thymocyte death underlies the CD4: CD8 T-cell ratio in the adaptive immune system. *Proceedings of the National Academy of Sciences*, 2013.
- [85] N. J. Singh, J. K. Bando, and R. H. Schwartz. Subsets of nonclonal neighboring CD4⁺ T cells specifically regulate the frequency of individual antigen-reactive T cells. *Immunity*, 37(4):735–746, 2012.
- [86] E. Stirk, C. Molina-Paris, and H. van den Berg. Stochastic niche structure and diversity maintenance in the T cell repertoire. *Journal of Theoretical Biology*, 255(2):237–249, 2008.
- [87] G. Stritesky, S. Jameson, and K. Hogquist. Selection of self-reactive T cells in the thymus. *Annual Review of Immunology*, 2011.
- [88] M. J. Stubbington, T. Lönnberg, V. Proserpio, S. Clare, A. O. Speak, G. Dougan, and S. A. Teichmann. T cell fate and clonality inference from single-cell transcriptomes. *Nature Methods*, 13(4):329, 2016.
- [89] J. T. Tan, E. Dudl, E. LeRoy, R. Murray, J. Sprent, K. I. Weinberg, and C. D. Surh. IL-7 is critical for homeostatic proliferation and survival of naive T cells. *Proceedings of the National Academy of Sciences*, 98(15):8732–8737, 2001.
- [90] C. Tanchot, F. A. Lemonnier, B. Pérarnau, A. A. Freitas, and B. Rocha. Differential requirements for survival and proliferation of CD8 naive or memory T cells. *Science*, 276(5321):2057–2062, 1997.
- [91] H. Taylor and S. Karlin. *An introduction to stochastic modeling*. Academic Press, 1998.
- [92] J. Textor, S. E. Henrickson, J. N. Mandl, U. H. Von Andrian, J. Westermann, R. J. De Boer, and J. B. Beltman. Random migration and signal integration promote rapid and robust T cell recruitment. *PLoS computational biology*, 10(8):e1003752, 2014.
- [93] N. Thomas, K. Best, M. Cinelli, S. Reich-Zeliger, H. Gal, E. Shifrut, A. Madi, N. F. riedman, J. Shawe-Taylor, and B. Chain. Tracking global changes induced in the CD4 T cell receptor repertoire by immunisation with a complex antigen using local sequence features of CDR3 protein sequence. [biorxiv.org](https://doi.org/10.1101/003752), 2014.

- [94] P. G. Thomas, A. Handel, P. C. Doherty, and N. L. La Gruta. Ecological analysis of antigen-specific CTL repertoires defines the relationship between naive and immune T-cell populations. *Proceedings of the National Academy of Sciences*, 110(5):1839–1844, 2013.
- [95] V. Thomas-Vaslin, H. Altes, R. de Boer, and D. Klatzmann. Comprehensive assessment and mathematical modeling of T cell population dynamics and homeostasis. *Journal of Immunology*, 180(4):2240, 2008.
- [96] J. J. Thome, K. L. Bickham, Y. Ohmura, M. Kubota, N. Matsuoka, C. Gordon, T. Granot, A. Griesemer, H. Lerner, T. Kato, et al. Early-life compartmentalization of human T cell differentiation and regulatory function in mucosal and lymphoid tissues. *Nature Medicine*, 22(1):72–77, 2016.
- [97] T. van den Broek, J. A. Borghans, and F. van Wijk. The full spectrum of human naive T cells. *Nature Reviews Immunology*, page 1, 2018.
- [98] V. van Hoven, J. Drylewicz, L. Westera, I. den Braber, T. Mugwagwa, K. Tesselaar, J. A. Borghans, and R. J. de Boer. Dynamics of recent thymic emigrants in young adult mice. *Frontiers in Immunology*, 8:933, 2017.
- [99] V. Venturi, K. Kedzierska, S. J. Turner, P. C. Doherty, and M. P. Davenport. Methods for comparing the diversity of samples of the T cell receptor repertoire. *Journal of Immunological Methods*, 321(1):182–195, 2007.
- [100] V. Vezys, A. J. Yates, K. A. Casey, G. Lanier, R. Ahmed, R. Antia, and D. Masopust. Memory CD8 T-cell compartment grows in size with immunological experience. *Nature*, 457(7226):196–199, 2008.
- [101] U. von Andrian, C. Mackay, et al. T-cell function and migration. two sides of the same coin. *The New England Journal of Medicine*, 343(14):1020, 2000.
- [102] H. Von Boehmer, I. Aifantis, F. Gounari, O. Azogui, L. Haughn, I. Apostolou, E. Jaeckel, F. Grassi, and L. Klein. Thymic selection revisited: how essential is it? *Immunological Reviews*, 191(1):62–78, 2003.
- [103] L. Westera, V. Hoven, J. Drylewicz, G. Spierenburg, J. F. Velzen, R. J. Boer, K. Tesselaar, and J. A. Borghans. Lymphocyte maintenance during healthy aging requires no substantial alterations in cellular turnover. *Aging Cell*, 14(2):219–227, 2015.
- [104] A. J. Yates. Theories and quantification of thymic selection. *Frontiers in Immunology*, 5:13, 2014.
- [105] W. Yu, N. Jiang, P. J. Ebert, B. A. Kidd, S. Müller, P. J. Lund, J. Juang, K. Adachi, T. Tse, M. E. Birnbaum, et al. Clonal deletion prunes but does not eliminate self-specific $\alpha\beta$ CD8⁺ T lymphocytes. *Immunity*, 42(5):929–941, 2015.
- [106] V. Zarnitsyna, B. Evavold, L. Schoettle, J. Blattman, and R. Antia. Estimating the diversity, completeness, and cross-reactivity of the T cell repertoire. *Frontiers in Immunology*, 4:485, 2013.
- [107] D. Zehn, C. King, M. Bevan, and E. Palmer. Tcr signaling requirements for activating T cells and for generating memory. *Cellular and Molecular Life Sciences*, 69(10):1565–1575, 2012.