INTRODUCTION

A T cell is characterized 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% 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. This thymic “selection” produces a repertoire of T cells, with many different specificities. An adaptive immune response begins with a small number of binding events between TCR and pMHC on the surfaces of antigen-presenting cells (APCs) that have processed peptides not ordinarily found in the body. T cells with the correct specificity, initially rare, then enter a program of multiple rounds of cell division and production of chemical signals. The initial binding events typically occur in one of the body’s lymph nodes; T cells usually circulate from node to node, scanning APCs for pMHC, but rarely encountering strong stimulation. 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

aging, competition, computational models, extinction, sampling, single-cell sequencing, stochastic
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 demobilize 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.17–26

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 I 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.21,27 Full understanding remains elusive, in part because the distinctions between subsets are seldom clear-cut, their origins and possible reversibility are the subject of active research.28–32

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.33–40 They may obtain IL-7 as they circulate, represented not as an integer but by a real function of age, \( t \). Then, the processes of internalization, degradation, and recycling of complexes mathematical models of the binding of IL-7 molecules to receptors, and of the average thymic production of a mouse. We therefore replace the constant \( \beta \) by a function of age, \( \beta(t) \). For mice older than 6 weeks, we may use the function45

\[
\beta(t) = Ae^{-\nu t} \text{ cells per day},
\]

where \( \nu = 0.005 \text{ per day} \) and the value of the constant A varies from mouse to mouse. If, at 6 weeks old, a mouse has 30 million naïve CD4+ T cells and its thymus is producing one million cells per day, we find

\[
N(t) = \frac{\theta(t)}{\beta - \nu} 10^7 e^{\nu(t-42)}. \tag{2}
\]

In this description, because \( e^{\nu t} \) decreases more rapidly as a function of age than \( e^{-t} \), 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.36 For comparison, the rate of export of naïve CD4+ T cells from the thymus of a young adult human is about \( 2 \times 10^8 \text{ cells per day} \); the rate declines proportional to \( e^{-t} \) where \( \nu = 0.0001 \text{ per day} \).47

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.45 Suppose that, after cell division, the two daughter cells remain Ki67+ for 4 days. Then, ignoring the time spent cycling before division, the division rate \( \lambda \), 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.48–51

Simple formulae such as Eq. (2) that result from simple models such as Eq. (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 Eq. (1) can be considered for the different thymocyte populations, for CD4+ and CD8+ naïve T cells and for CD4+ and CD8+ memory T cells.48,52,53 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,54,55 then they warrant a separate equation.

Models based on Eq. (2) cannot describe extreme scenarios such as recovery from lymphopenia,56,57 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-labeling 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 subpopulations with different rates.58 Second, 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.45,53,59 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 \( \theta \). Let \( N(t) \) be the total number of cells alive at time \( t \). As \( \Delta t \to 0 \), 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

\[
\frac{d}{dt} x = -\mu x + \theta.
\]

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 \( \mu t \to \infty \),

\[
y(t) \to \left( \frac{\theta}{\mu} \right)^2 + \frac{\theta}{\mu}.
\]

In other words,

\[
\mathbb{E}(N(t)) \to \frac{\theta}{\mu} \quad \text{and} \quad \text{var}(N(t)) \to \frac{\theta}{\mu}.
\]

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 cells 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.60 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 Eq. (3), which assumes no peripheral division. The values of \( \mu \) are similar to estimated death rates of naïve cells in mice and the initial number of cells per clonotype is \( n_0 = 8 \).
FIGURE 2  Survival of clonotypes in the periphery under the approximation (7) of neutral dynamics. Clonotypes leave the thymus as a family of eight cells.

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 summarized as follows: $n_p$ 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 $\mu$, because the birth rate is zero. Assuming that each cell in each clonotype, independently, has a constant death rate $\mu$, the probability that a clonotype that emerges from the thymus as a family of $n_p$ cells, survives in the periphery longer than $t$ is given by

$$S(t) = 1 - (1 - e^{-\mu t})^{n_p},$$

and is plotted in Fig. 1, with $n_p = 8$.

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 $\Delta t$ 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 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

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

and

$$G_n(z,t) = \sum_{k=0}^n p_k(t) z^k.$$  \hspace{1cm} (4)

Because cells behave independently, the descendants of each of the $n$ initial cells can be thought of as $n$ independent families and the “probability generating function” satisfies

$$G_n(z,t) = G_1(z,t)^n.$$ \hspace{1cm} (5)

Let us consider $G_1(z,t)$. Using the definition (4) we know that $G_1(z,0) = z$. Next, consider a time $\Delta t$, 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,

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

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

$$\frac{d}{dt}G_1(z,t) = \mu - 2\mu G_1(z,t) + \mu G_1^2(z,t).$$ \hspace{1cm} (6)

The solution of (6) is

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

Thus, using (5),

$$p_n(t) = G_n(0,t) = \left(\frac{\mu t}{1 + \mu t}\right)^n.$$ \hspace{1cm} (7)

The survival function, $S(t) = 1 - p_n(t)$, shown in Fig. 2, is the probability that $n(t) > 0$; in other words, the fraction of clonotypes that survive until $t$. Notice that extinction is certain because $p_0(t) \rightarrow 1$ as $\mu t \rightarrow \infty$. The values of $\mu$ 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 (7) 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 cell division depends on access to a nonspecific 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 $\gamma$. 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 hypothesis that each T cell, independently, has a death rate $\mu$.

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

$$N_{ss} = \frac{\gamma + \theta \eta}{\mu}.$$  

(8)

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

$$\lambda_i(t) = \frac{n_i(t)}{N_{ss}}.$$  

(9)

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 $N(t) \approx N_{ss}$. In this approximation, the integer $n_i(t)$ follows a birth-and-death process\(^{65}\) with death rate $\mu n_i$ and birth rate

$$\lambda_i(t) = n_i(t)$$

and the mean lifetime of a clonotype in the periphery. In the limit $n_i \Delta t \rightarrow 0$,

$$N^* \rightarrow N_{ss} \left[ 1 - \gamma \eta - \log (n_i \alpha) \right].$$  

(12)

where $\gamma \eta = 0.577\ldots$ is the Euler-Mascheroni constant.\(^{66}\)

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 $X_i$, that satisfies the following stochastic differential equation\(^{66}\)

$$dX_i = -\alpha \mu X_i dt + \sqrt{2\alpha X_i} dW_i.$$  

(13)

When $\alpha = 0$, the survival probability of this process is

$$S(t) = 1 - \exp \left( -\frac{\eta}{\mu} t \right),$$  

which is similar to (7). With the diffusion approximation, (13), 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 $X_i$ is the occupation density of the process $X_i$.\(^{67}\) The function $L(y, n_0)$ = occupation density at $y$ given that the process starts at $n_0$ and is absorbed at 0, satisfies

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

(14)

(The mean lifetime of a clonotype is the integral of $L(y, n_0)$ with respect to $y$.) The solution of (14) is

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

(15)

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 recognize 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 \( M \) and that each pMHC subset acts as a resource that is equally likely to be consumed by each of the cells capable of recognizing it. Let us also assume that each pMHC has probability \( p \) of being recognized by a randomly chosen TCR. The value of \( p \) is estimated to be \( 10^{-6} \).\(^{68-70} \) The model can be thought of as a massive ecosystem of species,\(^ {71-74} \) 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 (12). This simplicity is due to the cross-reactivity being "unfocussed" in the sense of Mason.\(^ {75} \)

In the model just described, the total number of cells in the system has a well-defined mean. De Greef et al. devised a model where the total number of cells is fixed.\(^ {62,76} \) To implement this computationally, each time the thymus produces \( n \) 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:\(^ {62,76} \)

\[
\theta = \mu F_1.
\]  

(16)

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 \( \mu \), the death rate per cell (equal to the inverse of the mean lifetime of a cell). Increasingly accurate estimates of \( \mu \) have become available in the past few years.\(^ {48-51} \) 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 \( \theta \), we need to divide the rate of thymic production, in cells per unit time, for which good estimates exist,\(^ {44,46} \) by \( n \), the number of cells per clonotype at the moment of release from the thymus, which is not well established.\(^ {77} \) Putting aside the age-dependence of the thymus, we may use the estimates \( \theta = 2.5 \times 10^9 \) per month and \( \mu = 1 \) per month 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 \( \theta = 10^9 \) per year and \( \mu = 0.5 \) per year 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 \( 10^{10} \).\(^ {66} \)

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? The fact that the CD4⁺:CD8⁺ ratio in the periphery of an adult mouse is close to 1, although the SP4:SP8 ratio in the thymus is typically 4:1, consistent with the hypothesis that CD4⁺ differ from CD8⁺ T cells in this respect. Are death and division rates constant over the lifetime of a mouse or human, or do they change as part of the aging process?

In Fig. 5, we show a type of stochastic model of an aging T-cell repertoire with multiple competing clonotypes, with mean number of cells times smaller than the number of CD4⁺ T cells in an adult human. For the Figure, we have modified the stochastic dynamics, developed in Ref. [66] and described in Section 2.4, to include an age-dependent thymic export rate \( \theta(t) = \theta_0 V(t) e^{-\mu t} \). (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 relatively constant (at about \( 4 \times 10^9 \)) 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, and the altruistic hypothesis, where IL-7 receptor expression is reduced on T cells that have received IL-7 signals, can only be modeled if differences from cell to cell within a clonotype are included. It is possible to introduce such cell-to-cell heterogeneity in computational models where individual cells are agents and receptor-expression levels are attributes.

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 \( 10^{11} \) 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.

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. Estimates of the number of different TCRs that could, in principle, be produced by VDJ gene rearrangement in the thymus far exceed the total number of T cells in a mouse or human body. 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 \( \alpha \) or \( \beta \) chain sequences from a pool of more than a million T cells from one individual, 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.

Single-cell measurements, where PCR and sequencing are performed on one cell at a time, avoid these problems and biases, and can identify paired \( \alpha \) and \( \beta \) chains. 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. 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. 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 colors. The total number of balls is known, but the number of different colors is not. What can be deduced from the colors 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:

\[
q = \frac{m}{S}
\]

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 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,

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

The probability that there is one cell of type \( i \) in the sample is \( \text{Prob}[s_i = 1 | n_i = n] \approx nq \); the probability that there are two is \( \text{Prob}[s_i = 2 | n_i = n] \approx \frac{1}{2} n(n-1)q^2 \). 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 \( n \). In terms of the colored balls, the number of different clonotypes in the repertoire (number of different colors) is equal to \( S/n \). 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 \( \phi(z) = \sum_k \text{Prob}[s_i = k] z^k \) for \( 0 \le z \le 1 \). For the distribution of values of \( s_i \), \( \Phi(z) = \sum_k \text{Prob}[s_i = k] z^k \). Assuming that \( qn_i \ll 1 \), the relationship between \( \phi(z) \) and \( \Phi(z) \) is

\[
\Phi(z) = \phi(1 - q + az).
\]

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

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

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

\[
r = \frac{m^2}{25} \left( \frac{\text{E}(n^2)}{\bar{n}} - 1 \right).
\]

**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 colors in the entire set, and distribution of number of cells per color, based on the colors found in small sample of cells (mammals).}

**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 (17).
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 \( E(n^2)/n-1 \), that does. In Fig. 7, we consider a power-law repertoire. We say that the distribution is "power-law" if the number of clonotypes consisting of \( n \) cells is proportional to \( n^b \). If \( b > 3 \) then

\[
r = \frac{m^2}{2S} \left( \frac{\zeta(b-2)}{\zeta(b-1)} - 1 \right),
\]

where \( \zeta \) 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 CD8+ CD44− T cells are purified from the spleen of a mouse, about \( 2.9 \times 10^6 \) cells.\(^{77} \) 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.\(^{95} \) In the Figure, we use the exponent \( b = 2.5 \), which gives a mean clonal family size of \( n = 1.95 \) 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 \( (1 - \frac{1}{n})^{77} \).

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

\[
m = m_1 + 2m_2 + 3m_3 + \ldots ,
\]

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.

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.\(^{104-107} \) 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.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

REFERENCES

1. Jameson S, Hogquist K, Bevan M. Positive selection of thymocytes. Ann Rev Immunol. 1995;13:93-126.
2. Von Boehmer H, Aifantis I, Gounari F, et al. Thymic selection revisited: how essential is it? Immunol Rev. 2003;191:62-78.
3. Palmer E. Negative selection: clearing out the bad apples from the T-cell repertoire. Nat Rev Immunol. 2003;3:383–391.
4. Ladi E, Yin X, Chtanova T, Robey E. Thymic microenvironments for T cell differentiation and selection. Nat Immunol. 2006;7:338–343.

5. Palmer E, Naehr D. Affinity threshold for thymic selection through a T-cell receptor co-receptor zipper. Nat Rev Immunol. 2009;9:207–213.

6. Klein L, Hinterberger M, Wirsberger G, Kiyewski B. Antigen presentation in the thymus for positive selection and central tolerance induction. Nat Rev Immunol. 2009;9:833–844.

7. Stritesky G, Jameson S, Hougquist K. Selection of self-reactive T cells in the thymus. Ann Rev Immunol. 2012;30:95–114.

8. Yates AJ. Theories and quantification of thymic selection. Front Immunol. 2014;5:13.

9. Goldrath AW, Bevan MJ. Selecting and maintaining a diverse T-cell repertoire. Nature. 1999;402:255–262.

10. Mandl JN, Monteiro JP, Vrisekoop N, Germain RN. T cell-positive selection uses self-ligand binding strength to optimize repertoire recognition of foreign antigens. Immunity. 2013;38:263–274.

11. Bousso P, Robey E. Dynamics of CD8+ T cell priming by dendritic cells in intact lymph nodes. Nat Immunol. 2003;4:579.

12. Marchingo JM, Kan A, Sutherland RM, et al. Antigen affinity, co-stimulation, and cytokine inputs sum linearly to amplify T cell expansion. Science. 2014;346:1123–1127.

13. Zehn D, King C, Bevan M, Palmer E. TCR signaling requirements for activating T cells and for generating memory. Cell Mol Life Sci. 2012;69:1565–1575.

14. Krummel MF, Buxtermus F, Gérard A. T cell migration, search strategies and mechanisms. Nat Rev Immunol. 2016;16:193.

15. Mandl JN, Liou R, Klauschen, F, et al. Quantification of lymph node transit times reveals differences in antigen surveillance strategies of naive CD4+ and CD8+ T cells. Proc Natl Acad Sci USA. 2012;109:18036–18041.

16. Textor J, Henrickson SE, Mandl JN, et al. Random migration and signal integration promote rapid and robust T cell recruitment. PLoS Comput Biol. 2014;10:e1003752.

17. Tanchot C, Lemonnier FA, Péarnau B, Freitas AA, Rocha B. Differential requirements for survival and proliferation of CD8 naive or memory T cells. Science. 1997;276:2057–2062.

18. Gallust F, Leng D, Förster R, Lipp M, Lanzavecchia A. Two subsets of memory T lymphocytes with distinct homing potentials and effector functions. Nature. 1999;402:34–38.

19. Doherty P, Ribery D, Belz G. Quantitative analysis of the CD8+ T-cell response to readily eliminated and persistent viruses. Philos Trans R Soc Lond B Biol Sci. 2000;355:1093–1101.

20. Vezys V, Yates AJ, Casey KA, et al. Memory CD8+ T-cell compartment grows in size with immunological experience. Nature. 2008;457:196–199.

21. Pepper M, Jenkins M. Origins of CD4+ effector and central memory T cells. Nat Immunol. 2011;13:467–471.

22. Sathaliyawala T, Kubota M, Yudanin N, et al. Distribution and compartmentalization of human circulating and tissue-resident memory T cell subsets. Immunity. 2013;38:187–191.

23. Ariotti S, Beltman JB, Chodaczek G, et al. Tissue-resident memory CD8+ T cells continuously patrol skin epithelia to quickly recognize local antigen. Proc Natl Acad Sci USA. 2012;109:19739–19744.

24. Buchholz VR, Flossdorf M, Hensel I, et al. Disparate individual fates compose robust CD8+ T cell immunity. Science. 2013;340:630–635.

25. Thome J, Bickham KL, Ohmura Y, et al. Early-life compartmentalization of human T cell differentiation and regulatory function in mucosal and lymphoid tissues. Nat Med. 2016;22:72–77.

26. Beura LK, Hamilton SE, Bi K, et al. Normalizing the environment re-captulates adult human immune traits in laboratory mice. Nature. 2016;532:512.

27. Jenkins M, Chu H, McLachlan J, Moon J. On the composition of the preimmune repertoire of T cells specific for peptide-major histocompatibility complex ligands. Ann Rev Immunol. 2009;28:275–294.

28. Almeida A, Legrand N, Papiernik M, Freitas A. Homeostasis of peripheral CD4+ T cells: IL-2R alpha and IL-2 shape a population of regulatory cells that controls CD4+ T cell numbers. J Immunol. 2002;169:4850.

29. OShea JJ, Paul WE. Mechanisms underlying lineage commitment and plasticity of helper CD4+ T cells. Science. 2010;327:1098–1102.

30. Feinerman O, Jentsch G, Tkach, K, et al. Single-cell quantification of IL-2 response by effector and regulatory T cells reveals critical plasticity in immune response. Mol Syst Biol. 2010;6:437.

31. Gerritsen B, Pandit A. The memory of a killer T cell: models of CD8+ T cell differentiation. Immunity. 2016;94:236–241.

32. van den Broek T, Borghans JA, van Wijk F. The full spectrum of human naive T cells. Nat Rev Immunol. 2018;18:363–373.

33. Tan JT, Dudl E, LeRoy E, et al. IL-7 is critical for homeostatic proliferation and survival of naive T cells. Proc Natl Acad Sci USA. 1998;95:8732–8737.

34. Seddon B, Tomlinson P, Zamoytsa R, Interleukin 7 and T cell receptor signals regulate homeostasis of CD4 memory cells. Nat. 2003;4:680–686.

35. Seddon B, Zamoytsa R. Regulation of peripheral T cell homeostasis by receptor signalling. Curr Opin Immunol. 2003;15:321–324.

36. Min B, Foucras G, Meier-Schellersheim M, Paul WE. Spontaneous proliferation, a response of naive CD4 T cells determined by the diversity of the memory cell repertoire. Proc Natl Acad Sci USA. 2004;101:3874–3879.

37. Fry TJ, Mackall CL. The many faces of IL-7: from lymphopoiesis to peripheral T cell maintenance. J Immunol. 2005;174:6571–6576.

38. Sinclair C, Saini M, de Loeff I, Sakaguchi S, Seddon B. Long-term survival potential of mature T lymphocytes is programmed during development in the thymus. Sci. 2011;335:477.

39. Palmier M, Mahajan V, Chen J, Irvine D, Lauffenburger D. Signaling thresholds govern heterogeneity in IL-7-receptor-mediated responses of naive CD8+ T cells. Immunol Cell Biol. 2011;89:581–594.

40. Koenen P, Heinz R, Carrington EM, et al. Mutually exclusive regulation of T cell survival by IL-7R and antigen receptor-induced signals. Nat Commun. 2013;4:1735.

41. Link A, Vogt T, Favre S, et al. Fibroblastic reticular cells in lymph nodes regulate the homeostasis of naive T cells. Nat Immunol. 2007;8:1255–1265.

42. Mueller SN, Germain RN. Stromal cell contributions to the homeostasis and functionality of the immune system. Nat Rev Immunol. 2009;9:618–629.

43. Molina-Paris C, Reynolds J, Lythe G, Coles MC. Mathematical model of naive T cell division and survival IL-7 thresholds. Front Immunol. 2013;4:434.

44. den Braber I, Mugwagwa T, Vrisekoop N, et al. Maintenance of peripheral naive T cells is sustained by thymus output in mice but not humans. Immunity. 2012;36:288–297.

45. Hogan T, Gossel G, Yates AJ, Seddon B. Temporal fate mapping reveals age-linked heterogeneity in naive T lymphocytes in mice. Proc Natl Acad Sci USA. 2015;112:E6917–E6926.

46. Bains I, Thiebaut R, Yates A, Callard R. Quantifying thymic export: combining models of naive T cell proliferation and TCR excision circle dynamics gives an explicit measure of thymic output. J Immunol. 2009;183:4329.

47. Bains I, Antia R, Callard R, Yates A. Quantifying the development of the peripheral naive CD4+ T-cell pool in humans. Blood. 2009;113:5480.

48. Borghans, J, de Boer R. Quantification of T-cell dynamics: from telomeres to DNA labeling. Immunol Rev. 2007;216:35–47.
93. Miles J, Douek D, Price D. Bias in the αβ T-cell repertoire: implications for disease pathogenesis and vaccination. *Immunol Cell Biol.* 2011;89:375–387.

94. Qi Q, Liu Y, Cheng Y, et al. Diversity and clonal selection in the human T-cell repertoire. *Proc Natl Acad Sci USA.* 2014;111:13139–13144.

95. Oakes T, Heather JM, Best K, 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. *Front Immunol.* 2017;8:1267.

96. Cukalac T, Kan W-T, Dash P, et al. Paired TCRαβ analysis of virus-specific CD8⁺ T cells exposes diversity in a previously defined “narrow” repertoire. *Immunol Cell Biol.* 2015;93:804–814.

97. Stubbington MJ, Lönnberg T, Proserpio V, et al. T cell fate and clonality inference from single-cell transcriptomes. *Nat Methods.* 2016;13:329.

98. Venturi V, Kedzierska K, Turner SJ, Doherty PC, Davenport MP. Methods for comparing the diversity of samples of the T cell receptor repertoire. *J Immunol Methods.* 2007;321:182–195.

99. Robins HS, Campregher PV, Srivastava SK, et al. Comprehensive assessment of T-cell receptor β-chain diversity in αβ T cells. *Blood.* 2009;114:4099–4107.

100. Sepúlveda N, Paulino C, Carneiro J. Estimation of T-cell repertoire diversity and clonal size distribution by Poisson abundance models. *J Immunol Methods.* 2010;353:124–137.

101. Laydon D, Bangham C, Asquith B. Estimating T-cell repertoire diversity: limitations of classical estimators and a new approach. *Phil Trans R Soc B.* 2015;370:20140291.

102. Molina-Paris C, Castro M, Lythe G, Ribeiro R. Mathematics in modern immunology. *Interface Focus.* 2016;6:20150093.

103. Ferrarini M, Molina-Paris C, Lythe G. Sampling from T cell receptor repertoires. In: *Modeling Cellular Systems.* Cham: Springer; 2017:67–79.

104. von Andrian U, Mackay C, et al. T-cell function and migration. Two sides of the same coin. *N Engl J Med.* 2000;343:1020.

105. Hawkins E, Markham J, McGuinness L, Hodgkin P. A single-cell pedigree analysis of alternative stochastic lymphocyte fates. *Proc Natl Acad Sci USA.* 2009;106:13457–13462.

106. Miller M, Safrina O, Parker I, Cahalan M. Imaging the single cell dynamics of CD4⁺ T cell activation by dendritic cells in lymph nodes. *J Exp Med.* 2004;200:847–856.

107. Kinjo I, Qin J, Tan S-Y, et al. Real-time tracking of cell cycle progression during CD8⁺ effector and memory T-cell differentiation. *Nat Commun.* 2015;6:6301.

---

**How to cite this article:** Lythe G, Molina-Paris C. Some deterministic and stochastic mathematical models of naive T-cell homeostasis. *Immunol Rev.* 2018;285:206–217. [https://doi.org/10.1111/imr.12696](https://doi.org/10.1111/imr.12696)