Estimating the diversity, completeness, and cross-reactivity of the T cell repertoire

Veronika I. Zarnitsyna[1]*, Brian D. Evavold[2], Louis N. Schoettle[3], Joseph N. Blattman[3] and Rustom Antia[1]*

[1] Department of Biology, Emory University, Atlanta, GA, USA
[2] Department of Microbiology and Immunology, Emory University, Atlanta, GA, USA
[3] Center for Infectious Diseases and Vaccinology, School of Life Sciences, Arizona State University, Tempe, AZ, USA

In order to recognize and combat a diverse array of pathogens the immune system has a large repertoire of T cells having unique T cell receptors (TCRs) with only a few clones specific for any given antigen. We discuss how the number of different possible TCRs encoded in the genome (the potential repertoire) and the number of different TCRs present in an individual (the realized repertoire) can be measured. One puzzle is that the potential repertoire greatly exceeds the realized diversity of naïve T cells within any individual. We show that the existing hypotheses fail to explain why the immune system has the potential to generate far more diversity than is used in an individual, and propose an alternative hypothesis of "evolutionary sloppiness." Another immunological puzzle is why mice and humans have similar repertoires even though humans have over 1000-fold more T cells. We discuss how the idea of the "protecton," the smallest unit of protection, might explain this discrepancy and estimate the size of "protecton" based on available precursor frequencies data. We then consider T cell cross-reactivity – the ability of a T cell clone to respond to more than one epitope. We extend existing calculations to estimate the extent of expected cross-reactivity between the responses to different pathogens. Our results are consistent with two observations: a low probability of observing cross-reactivity between the immune responses to two randomly chosen pathogens; and the ensemble of memory cells being sufficiently diverse to generate cross-reactive responses to new pathogens.

Keywords: $\alpha\beta$ T cell, repertoire, precursor frequency, cross-reactivity, pathogen recognition

1. INTRODUCTION

The clonal selection theory of adaptive immunity requires that the immune system is able to produce a large and diverse repertoire of immune cells (clones), with each cell expressing a receptor with different antigenic specificity (1, 2). Following infection, the few clones that are specific for the antigens expressed by the pathogen proliferate and differentiate into effector cells which control the infection. Subsequently, the maintenance of an increased number of these pathogen-specific cells results in long-lasting immunological memory (3–5). Accurate quantification of changes in the numbers of antigen-specific cells during infection and vaccination, together with advances in molecular and cellular biology, has allowed us to make considerable progress toward understanding the dynamics of the generation of immune responses (3, 6, 7) and the requirements for pathogen control (8, 9). Furthermore, deep sequencing technology has provided a first quantitative snapshot of the diversity of immune cells (10, 11). These technological advances set the stage for understanding the relationship between the diversity of immune cells (the repertoire) and immune protection from an extensive array of pathogens to which we are exposed.

We begin by outlining our current understanding of T cell receptor diversity and discussing problems associated with the quantification of the T cell repertoire. Next, we explore how diverse the immune system needs to be by exploring the relationship between the diversity of the T cell repertoire and its ability to provide protection from pathogens. Finally we consider how the degree of specificity of T cells (often defined by measuring how cross-reactive they are) affects the relationship between the repertoire and host response to a given pathogen. We focus on $\alpha\beta$ T cells and the term "T cell" refers to the CD8 subpopulation of T cells unless we explicitly specify a different subpopulation.

We have intentionally used simple models and calculations because, in the absence of detailed information on the terms and parameters, simpler models frequently generate more robust qualitative results than complex models (12, 13). The focus of the paper is to highlight the limitations arising from uncertainties in current estimates of parameters, and in particular to gain maximum insight from the one key parameter – the precursor frequency of T cells specific for different epitopes – that can be accurately measured. Throughout this paper we emphasize current puzzles and problems and, where possible, suggest new approaches to solving them.

2. MEASURING THE DIVERSITY OF THE T CELL REPERTOIRE

2.1. WHAT IS THE POTENTIAL REPERTOIRE?

T cells develop from progenitor cells in the thymus where the germline T cell receptor (TCR) α and β genes undergo somatic recombination of the V-J and V-D-J gene segments, respectively (14, 15). The antigenic specificity of each T cell is determined

by the amino acid sequence of these rearranged TCR genes, and in particular by the hypervariable complementarity determining region 3 (CDR3) that mostly account for direct contacts with peptides presented on major histocompatibility complex (MHC) proteins, and is encoded by the junction of the V, (D), and J gene segments (16). The diversity of generated TCR genes is therefore due to: (1) selection of one from a number of possible V, D, and J gene segments, (2) semi-random cleavage of recombination hairpin intermediates, and (3) N-nucleotide addition and subtraction at the junction of V, D, and J genes (17, 18). Finally, the pairing of different α and β chains to generate a functional receptor results in additional diversity (19).

How many different T cell receptors can be generated? The first steps toward understanding the magnitude of the diversity of the T cell repertoire came from the pioneering studies that identified the molecular mechanisms involved in the recombination of V, (D), and J gene segments and N-region diversification described above for the generation of the α and β TCR chains (14, 15, 19). It was estimated that these processes together with pairing between different α and β chains could give rise to around 10^{15} possible $\alpha\beta$ TCR (19). The question of the potential number of TCR sequences has recently been revisited and significantly larger estimates for the diversity of the TCR β chain were obtained (20, 21). Murugan et al. (21) used statistical analysis of non-productive TCR β chain to conclude that the CDR3 (variable) region of the TCR β chain alone has a potential diversity of $\sim 10^{14}$ different sequences. They used empirical β chain data to show that N-nucleotide insertions at the V-D and D-J junctions are uncorrelated, their length distributions are nearly identical and their lengths could exceed six nucleotides which was assumed in previous estimates (19). We might expect that a similar analysis would result in upward revision for the potential diversity of the α chain (though the estimates of diversity would increase less than for the β chain because the α chain has only one V-J junction). This will result in a truly enormous potential repertoire of over 10^{20} for the $\alpha\beta$ TCR.

2.2. WHAT IS THE REALIZED REPERTOIRE IN AN INDIVIDUAL?

Only mature T cells that have passed thymic selection (naïve T cells) can be employed in immune responses for protection against pathogens. Thus, in order to understand the balance between diversity and protection, the most important measurement is the "realized" T cell diversity in an individual (i.e., the actual number of different TCR in the mature T cell compartment).

The diversity of the naïve T cell repertoire was initially estimated prior to the advent of deep sequencing technologies by the use of spectrotyping, which involved amplifying mRNA from particular V-J sequence combinations, separating the amplified products on the basis of size, and exhaustive conventional sequencing of a particular length CDR3 product. The diversity of TCR sequences in this sample was then extrapolated to the total T cell population.

2.2.1. TCR diversity and clone size in humans

Arstila et al. (22) used spectrotyping to estimate that there are 10^6 β chains in the blood each pairing on average with at least 25 different α chains, and consequently proposed a lower bound to the estimate of the T cell repertoire in humans of around 2.5×10^7

specificities. Advances in deep sequencing have confirmed that estimation of β chains is in the range of $1 - 4 \times 10^6$ (20, 23, 24).

There is however considerable uncertainty about the extent to which 2.5×10^7 specificities underestimates the diversity of T cells in humans (25, 26). A repertoire of 2.5×10^7 suggests a naïve clone size on average of over 4×10^3 cells ($>10^{11}/(2.5 \times 10^7)$). This could happen if each clone gets produced multiple times or if once produced a given clone would undergo about 12 rounds of division. The first scenario is unlikely, given the very large estimates of potential diversity (19–21). If the second scenario happens, it must occur in the periphery. Expansion of clones in the thymus would result in a much lower frequency of detectable T cell receptor excision circles (TRECs) in the naïve pool of recent thymic emigrants than is currently observed (27–29). Arstila et al. points out that naïve T cells in the periphery could divide more than 12 times during a human lifespan (26). However, as the total number of naïve T cells remains relatively stable (because division is balanced by death) changes in clone size would have to arise from stochastic drift and this seems unlikely.

2.2.2. TCR diversity and clone size in mice

Interestingly, it was estimated that TCR β chain diversity in mouse spleen is quite similar to the one measured in human blood. The β chain repertoire of $5 - 8 \times 10^5$ specificities with each variable domain of β chain sequence being shared by 30–40 T splenocytes have been reported using spectrotyping technology (30). Pairing with α chain was estimated to add a factor of 2.4 and resulted in total $\alpha\beta$ TCR diversity of 2×10^6. Taking into account that there are 2×10^7 splenocytes it was estimated that the clone size of $\alpha\beta$ TCR is equal to 10 cells (30). The bias in recombination and α-β TCR pairing will likely affect the T cell clone size. A recent study that enumerated the number of naïve T cells specific for different epitopes suggests that there are between 15 and 1500 unique cells in the mouse spleen specific for any given epitope, implying that the number of naïve cells with a given TCR α-β combination is very small, and indeed that most clonotypes have clone size of one (31, 32). This is in contrast with the earlier estimates that suggest an average clone size of 10 cells/clone in the spleen (30). Consequently, it brings the repertoire in the spleen toward the total number of naïve T cells in the spleen, and increases the lower bound of the total $\alpha\beta$ T cell repertoire in the mouse by an order of magnitude. In this case the estimate of 2×10^7 specificities becomes very close to a lower bound estimation for human T cell repertoire.

2.2.3. Limitations in estimates of realized diversity

Current estimates of the realized diversity are lower bounds. The limitations of these studies is the lack of information on the pairing of different TCR α and β chains. Bulk sequencing of a single chain, or even of both TCR α and β chains, is not sufficient to inform us of the potential diversity (33). In principle this problem could be comprehensively addressed by single cell sequencing that would obtain linked α and β chain sequences, but this remains technically and financially infeasible for the large sample sizes required to evaluate naïve repertoires with high diversity (34); the cost of single cell sequencing remains at $1 per cell, making the analysis of T cells from a single mouse more than a $10 million experiment! Oil

emulsion lysis strategies (35) combined with micro-sequencing have increased the capacity of such single-cell studies, but these still are only able to capture <1% of the total naïve T cell repertoire in a single mouse. New techniques or methods need to be developed.

In order to have an accurate and comprehensive quantitative description of diversity, it is important to define what we mean by diversity. We can describe T cell repertoire diversity in terms of summary measures of diversity borrowed from the ecological literature. This includes enumerating the number of distinct clones or computing the Simpson diversity index (36) that takes into account the number of clones and their frequencies. However these summary approaches compress all of the diversity information into a single number. A more comprehensive statistical approach retains the frequency distribution of different clone sizes. In **Figure 1** we show a plot of the frequency distribution of β chain sequences in the mouse naïve T cells using preliminary data. We find a majority of β chain sequences are present at low frequencies and fewer sequences occurring at much higher frequencies. A key problem is that we do not know the α chain sequences pairing with each of these β chains, and this restricts our ability to infer diversity of T cells from these observations.

Several clones in **Figure 1** have very high frequencies and the exact underling mechanisms are not known. The sequences which are more common (generated more frequently) are more likely to be shared between different individuals. It was reported that inbred mice and individuals with the same MHC share some T cells with identical receptors (11, 20, 37, 38). These constitute "public" T cell clones, in contrast with the majority of the T cell clones that are unique to an individual and comprise the "private" part of a repertoire response. In general, the frequency of public TCR clonotypes exceeds what is expected if T cells were chosen at random with equal probability from the total potential repertoire, and perhaps
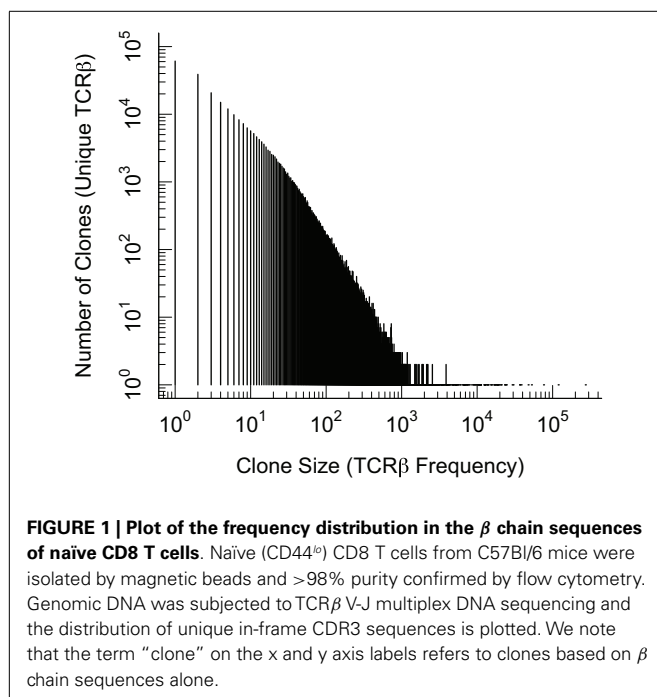
not surprisingly, the clone size of public T cells is higher than that of private T cells in the naïve T cell repertoire [(33); Blattman et al. unpublished results]. This has been suggested to arise due to MHC restriction during thymic selection, biased frequencies of recombination, as well as degeneracy in the genetic code which allows more than one nucleotide sequence to give rise to the same amino acid sequence (33, 39). The factors involved in the evolution and/or selection of public T cell clonotypes and their possible role in the control of infections remain puzzling questions.

In summary we have estimates of the potential repertoire of upward of 10^{20} TCR. The estimates of the realized repertoire suggest lower bounds of 2.5×10^7 and 2×10^6 in humans and mice. Two puzzles which we will address are: why humans and mice might have similar repertoire sizes (Section 3.2); and why the potential repertoire so greatly exceed the realized repertoire (Section 3.3).

3. UNDERSTANDING DIVERSITY, THE REPERTOIRE AND CROSS-REACTIVITY

In this section we use quantitative calculations to explore the consequences of the observations on the repertoire described in the previous section. We begin by looking at whether the diversity of the repertoire may be explained by the relationship between diversity and protection. We then address questions associated with our current understanding of repertoire diversity and cross-reactivity.
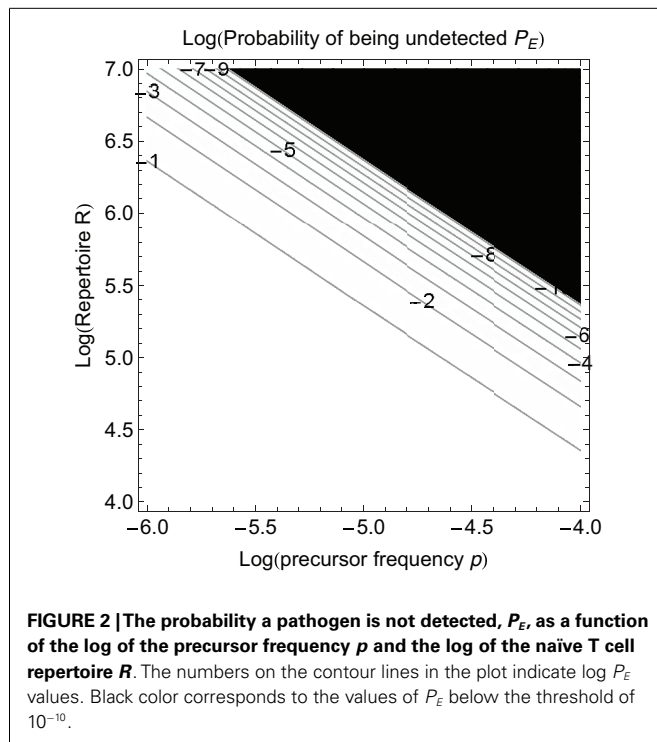
3.1. RELATIONSHIP BETWEEN DIVERSITY AND PROTECTION

Clearly a large repertoire is required to generate a T cell response to a diverse array of pathogens. However, to our knowledge, few empirical studies consider the relationship between the repertoire and protection. To some extent the paucity of experiments on this topic is because of difficulties in quantifying the repertoire (see earlier discussion). Studies on mice, expressing a single fixed transgenic TCR chain (either α or β) that measure the number of different paired endogenously recombined TCR chains, have shown that pairing is not completely random, as these mice express repertoires of reduced diversity and altered V gene usage (40–43). However, even in these mice there is still sufficient diversity to generate effective, albeit reduced, responses to control pathogen infections.

A relatively simple calculation can be made to estimate how diverse the TCR repertoire needs to be in order to provide reliable protection following infection with a pathogen. To provide protection against a pathogen there must be some number of clones present in the repertoire that are specific for that pathogen. Here, we extend the logic outlined in (44, 45). Let R be the T cell repertoire and let p_i be the probability that a randomly chosen TCR binds to i^{th} of the k epitopes derived from a given pathogen ($i = 1:k$). Note that p_i is also equal to the precursor frequency of T cells for i^{th} epitope. A pathogen is not detected if all R naïve T cell clones fail to recognize it, and this will happen with probability.

$$P_E = (1 - p_1)^R (1 - p_2)^R \ldots (1 - p_k)^R$$

$$\approx \exp(-p_1 R) \exp(-p_2 R) \ldots \exp(-p_k R) = \exp\left(-R \sum_{i=1}^{k} p_i\right)$$

$$(1)$$

FIGURE 1 | Plot of the frequency distribution in the β chain sequences of naïve CD8 T cells. Naïve (CD44^{lo}) CD8 T cells from C57Bl/6 mice were isolated by magnetic beads and >98% purity confirmed by flow cytometry. Genomic DNA was subjected to TCRβ V-J multiplex DNA sequencing and the distribution of unique in-frame CDR3 sequences is plotted. We note that the term "clone" on the x and y axis labels refers to clones based on β chain sequences alone.

FIGURE 2 | The probability a pathogen is not detected, P_E, as a function of the log of the precursor frequency p and the log of the naïve T cell repertoire R. The numbers on the contour lines in the plot indicate log P_E values. Black color corresponds to the values of P_E below the threshold of 10^{-10}.
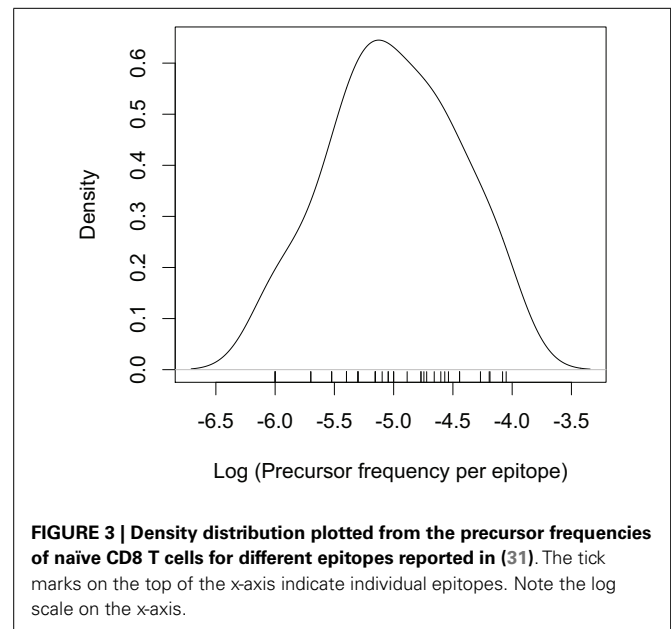
which gives

$$R = \frac{-\ln(P_E)}{\sum_{i=1}^{k} p_i} \qquad (2)$$

Equation (2) tells us how big the repertoire must be to detect at level P_E. **Figure 2** shows how the probability that a pathogen is not detected by the immune system depends on the repertoire size and the total precursor frequency $p = \sum p_i$. There is a very rapid decline in the probability of not being detected once the product of p and R becomes sufficiently large. We should note that P_E is often termed as the "probability of escape" but it should not be confused with the usage of the term "escape" that refers to the generation of escape mutants in T cell epitopes after recognition has already occurred following infections such as HIV.

If we know the precursor frequencies for pathogen epitopes and total number of epitopes we can relate the probability of being not detected to the repertoire R. We have much more accurate estimates for precursor frequencies against specific epitopes than for repertoire sizes (31, 46, 47). A recently developed method that combines pMHC tetramer staining with magnetic particle-based cell enrichment allows for the direct measurement of the frequency of naïve cells to different epitopes for both mice and humans (31, 48). **Figure 3** shows the density distribution of naïve T cell precursor frequencies for different CD8 T cell epitopes in mice determined by this cell enrichment method using MHC tetramers complexed with different class I-restricted peptides (31). The total number of responded cells per mouse (naïve precursor frequency multiplied by total CD8 T cell number) varied from 15 in response to the L-338:D^b epitope of LCMV to 1500 in response to an epitope from the murine cytomegalovirus (31). These estimates concur
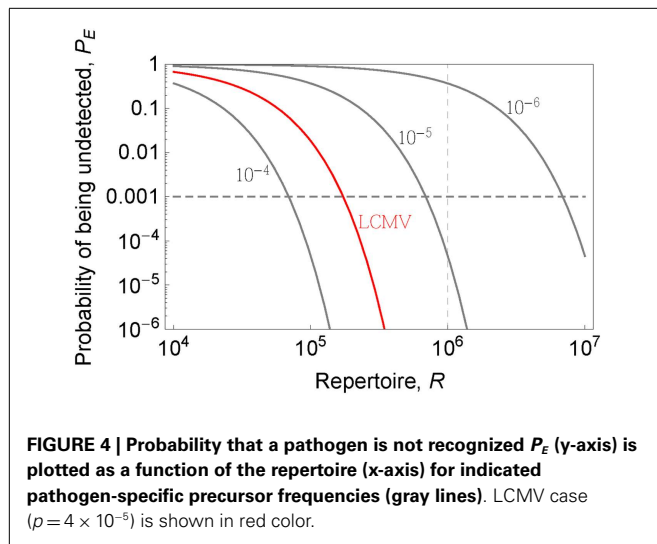

FIGURE 3 | Density distribution plotted from the precursor frequencies of naïve CD8 T cells for different epitopes reported in (31). The tick marks on the top of the x-axis indicate individual epitopes. Note the log scale on the x-axis.

with previous *in vivo* estimates of precursor frequencies. These studies transferred different numbers of naive epitope-specific T cells and measured the proportion of the response arising from expansion of host versus donor cells following virus infection (46, 47). The effect of changing precursor frequencies on the probability of been undetected, P_E, is given by equation (2) and plotted in **Figures 2** and **4**. Note, that precursor frequencies plotted in **Figure 3** are likely biased toward immunodominant epitopes. Immunodominance is a complex issue, and the major factors that affect the magnitude of the T cell response to a particular epitope include: the processing and presentation of peptide on MHC (i.e., the amount of epitope); the number of precursor cells for this epitope; their affinities for the epitope; the extent of their recruitment and competition between the T cells for this and other pathogen epitopes (31, 49–51).

3.2. SCALING AND THE CONCEPT OF A "PROTECTON"

We now consider the scaling of the repertoire with the size of the organism. A few pathogen-specific precursors in a tadpole are likely to be able to mount a faster and more effective response than the same number of cells in an elephant (52). Langman and Cohen proposed the basic functional unit, the "protecton," capable of providing robust protection. They suggested a tadpole (smallest vertebrate) has a single "protecton," and the number of "protectons" scales with the size of an organism. Localized infections are surveyed by a draining lymph node rather than the entire immune system and thus we expect this unit should contain at least one "protecton." Clearly the calculations for P_E (how diverse the immune system needs to be to recognize pathogens) pertains to the "protecton" [see equations (1) and (2)].

Lets estimate the diversity in a "protecton." **Figure 4** shows how the probability of being undetected depends on the size of the repertoire R for different total precursor frequencies. The precursor frequency of T cells specific for a pathogen is, to a first approximation, the sum of the precursor frequencies for that

FIGURE 4 | Probability that a pathogen is not recognized P_E (y-axis) is plotted as a function of the repertoire (x-axis) for indicated pathogen-specific precursor frequencies (gray lines). LCMV case ($p = 4 \times 10^{-5}$) is shown in red color.

pathogen's different epitopes presented by MHC proteins. This can be estimated for LCMV by combining reported naïve precursor frequencies for few measured epitopes (31) and measurements of the fraction of the total LCMV specific responses which is directed against these epitopes (53). This gives us a total precursor frequency for LCMV specific T cells equal to ~4×10^{-5}, and from **Figure 4**, a level of protection $P_E = 10^{-3}$ (i.e., 1 in 1000 "protectons" will fail to recognize LCMV) requires the repertoire in the "protecton" to be about 1.7×10^5. In order to know the level of protection against diverse pathogens we need to know the distribution of precursor frequencies to pathogens. The existing data gives us lower bounds (because only cells specific for a few epitopes are measured) to the precursor frequencies of viruses such as MCMV (~1.3×10^{-4}), Influenza (~4×10^{-5}), Vaccinia (~1.1×10^{-4}), RSV (~4.5×10^{-5}), HSV (~2.9×10^{-5}), and VSV (~10^{-5}) (31). The precursor frequencies for those viruses are comparable or greater than that for LCMV with the exception of VSV and HSV for which only single epitope data were reported in (31). If this trend holds (i.e., the precursor frequency per pathogen is >10^{-5}) it might suggest that having a repertoire of 7×10^{-5} is sufficient to give robust protection at the level $P_E = 10^{-3}$, and thus define the size of the "protecton." For P_E one order of magnitude lower and higher, i.e., $P_E = 10^{-4} - 10^{-2}$ will require a repertoire of ~$9 \times 10^5 - 5 \times 10^5$, suggesting that our estimate is quite robust to changes in P_E (see **Figure 4**). We can expect the area of local surveillance (a small lymph node) in mice to have at least this number of different T cells.

How much bigger should the total repertoire size be so that the area corresponding to one "protecton," randomly filled with T cells from the total circulating cells, has a relatively low number of clone repeats? We estimated that if f is a fraction of clone repeats in the "protecton" area with m cells, the total repertoire size R is bounded as $(1 - f)(m - 1)/(2f) < R < (m - 1)/(2f)$. For example, for 5 or 10% of clone repeats in m we will have a multiplication factor for m for the total diversity in the ranges ~$9.5 - 10$ or ~$4.5 - 5$, respectively. To derive this formula we used two assumptions: first, the clones are equal in size and second, the size of total repertoire multiplied by clone size is much bigger than the size m. These

calculations show that the total diversity doesn't need to be much higher than the diversity in a "protecton."

Several theoretical papers previously estimated that the repertoire of B and T cells scales as $ln(cM)$, where c is a constant and M is the mass of an organism (45, 54). It was also estimated that humans should have B cell repertoire 2–5 times larger than mice (45) and similar reasoning could be apply to T cell repertoire. The diversity of the repertoire is linked to clone size and it was estimated that the size of T cell clones should scale as M and, correspondingly, the total number of T cells should scale as $Mln(cM)$ (45). Wiegel and Perelson's estimate shows that the repertoire in a human need not be much higher than that in a mouse even though the number of naïve cells in these organisms differs by over 10^3 fold [mice have ~10^8 T cells (30, 46) and humans between 10^{11} and 10^{12} T cells (22, 55, 56)].

Another reason for why humans need a more diverse repertoire than mice pertains to the number of pathogens to which they are exposed. As humans live longer than mice, other factors being equal, they will be exposed to more pathogens and require a lower P_E.

3.3. EVOLUTIONARY CONSIDERATIONS: WHY ENCODE SUCH A DIVERSE POTENTIAL REPERTOIRE?

The calculations described in the previous section are consistent with the diversity of the repertoire that is observed in mice and humans (22, 30) (lower bound diversity in the range of 2×10^6 to 2.5×10^7 unique $\alpha\beta$ T cells), and the diversity is sufficient to generate a low probability that a given pathogen is not detected ($P_E < 10^{-4}$). What those estimations don't explain is why the immune system is able to generate a potential diversity of more than 10^{15} T cell specificities (19–21) that is vastly in excess of the realized repertoire?

Let's consider a number of potential explanations for why the potential repertoire needs to be much larger than the realized repertoire. One simplistic explanation takes into account the observation that most of the generated progenitor T cells are deleted during positive and negative selection in the thymus. If a fraction f of the T cells generated in the thymus gets selected (i.e., pass positive or negative selection) then the potential repertoire should be $(1/f)$ times the peripheral repertoire. Since only 3–5% of T cells pass thymic selection (57, 58), the potential repertoire need only be at most 33-fold higher than the realized repertoire, thus ruling out this explanation.

A second potential reason is the need to successfully recognize peptides in the context of the hundreds of MHC alleles in the population. The reported extent of thymic selection (see previous paragraph) allows us to reject this hypothesis – different cells may be selected in different MHC backgrounds but in all cases 3–5% of T cells pass thymic selection.

A third potential reason is the need to prevent pathogen escape mutations – mutations in an epitope that prevent it from being recognized by the immune system. To a first approximation having more than one epitope is the key factor that prevents escape – if the pathogen has k epitopes the probability of escaping all epitopes declines as μ^k, where μ is the probability of mutation leading to the loss of one epitope. The number of epitopes to which a response is generated involves many factors such as immunodominance,

MHC diversity, and T cell diversity. The relationship between these quantities is not understood and we do not know the contribution of T cell diversity to immunodominance due to problems in estimating TCR diversity described above. However from **Figure 4** we note that the repertoire is sufficiently large to enable robust detection of subdominant epitopes in a biologically reasonable range of precursor frequencies [**Figure 2**; (31)].

A fourth potential reason considers the temporal aspect and changes in the repertoire over the lifespan of an individual. Thymic emigration results in a constantly changing repertoire over time. The total number of different T cells present in the individual over its lifespan could be much greater than its repertoire at any given time. In humans, for example, if we assume that thymic emigration is of the order of $10^7 - 10^8$ cells per day (59, 60) then the realized repertoire over a lifespan might be as much as 10^{12} specificities which is much closer to the potential repertoire. There are two problems with this approach. First, it does not explain why mice have about the same potential repertoire as humans since a similar calculation for mice would result in a realized repertoire over the lifespan several orders of magnitude lower than humans. This is because both mice thymic output of the order of 10^6 cells per day (61–63) and lifespan are smaller than for humans. Second, protection is related to the repertoire at a given time point. Changing the particular clones in the repertoire over time does not help unless the relevant clones are present at the time of infection or generated during the infection and consequently able to help with clearance of the pathogen. The continual influx of cells of new specificities is thus unlikely to be of significance for acute infections which are relatively brief, but has been suggested to contribute to the maintenance of the response during persistent infections (64–66). In the case of persistent infections, however, an occasional new pathogen-specific clone is unlikely to clear the infection if the much larger number of clones at the onset were not able to do so – and the new clone is likely to be exhausted rapidly. Finally, temporal aspects could change the total repertoire if we consider the sum of both naïve and memory compartments. As naïve cells convert to memory cells each time we confront an infection, the replenishment of naïve compartment with the cells of new specificities would increase the total repertoire (naïve plus memory compartments). However, even if the memory compartment is as diverse as the naïve, the total diversity would increase at most by a factor of 2 in comparison to the naïve compartment alone. Taken together we don't expect temporal aspects to account for the differences in the sizes of the potential and realized repertoires.

We now describe a new evolutionary explanation that we call "evolutionary sloppiness." The process of generation of diversity by recombination and N nucleotide addition and deletion are sloppy processes. To reduce the amount of diversity that can be generated might require putting additional costly constraints on these processes. This would explain why organisms are able to generate far more TCR diversity (in excess of 10^{15} TCR) then is needed. Finally we note that not all aspects of biology result in perfectly optimized solutions (67).

Additionally, it has been suggested that the thymus is an energy- and resource-expensive organ (68) but these energetic costs have yet to be quantified. Energetic costs to cell production and thymic selection would favor expansion of clones after thymic selection

(i.e., to have an amplifier). This amplification could take place in the thymus or periphery and would scale with the size of the organism. This would result in clone sizes in men being ~1000-fold higher than in mice, which is unlikely (see discussion on TRECs in Section 2.2.1). Accurate estimates for clone sizes in humans and mice should allow us to resolve this question.

3.4. T CELL CROSS-REACTIVITY

Cross-reactivity is related to the observation that a given T cell can respond to more than one epitope, including epitopes that show strong sequence homology or completely unrelated (69–76). As might be expected the frequency of the former is higher than that of the latter. Flexible TCR-pMHC binding sites were suggested as a possible structural explanation for known high degree of $\alpha\beta$ TCRs cross-reactivity to different pMHCs (77–80). Cross-reactivity can also arise in T cell clones with incomplete allelic exclusion at the α chain loci resulting in one β chain pairing with two different α chains. An upper bound on the frequency of such clones was estimated to be 30% (81–83).

The pioneering experiments of Selin and Welsh (69) found that the CD8 T cell responses of mice to pathogens such as the Pichinde virus (PV), Vaccinia virus (VV), and Lymphocytic choriomeningitis virus (LCMV) showed high levels of cross-reactivity. They found that prior vaccination with one of these viruses expanded a specific CD8 T cell subset that could be boosted during stimulation by the other viruses and showed an unexpectedly complex relationship between the responses to different viruses with asymmetry depending on the order of viral exposure (infection A followed by B stimulated different cross-reactivity than B followed by A) (69, 71, 84). In a very recent paper (80) the cross-reactivity studies were extended to analysis of the CD4 T cell repertoire against pathogens to which individuals had never been exposed. Surprisingly, they found that a large fraction of the CD4 T cells specific for these pathogens exhibited a memory phenotype and suggested that they had been generated by cross-reactive responses to other previously encountered pathogens including heterologous infections or environmental antigens.

The extent of cross-reactivity between the immune responses to different pathogens is of practical importance. Murine studies have not only demonstrated the presence of T cells that could cross-react between different pathogens such as PV, VV, and LCMV, but also showed that this cross-reactivity affected pathogenesis during subsequent infection (84–86). If these occurrences of cross-reactive responses are rare, then the examples above are simply interesting curiosities. If, on the other hand, cross-reactivity is common then we would need to move from our current paradigm, which looks at each infection independent of other infections, to a more complex view that incorporates the terms for the interactions between the immune responses to different pathogens. Thus a key step is to quantify the extent of cross-reactivity in the immune responses to different pathogens.

How can we predict the level of cross-reactivity between two pathogens? The current approach is based on the observation that number of possible peptide-MHC complexes is much larger than the total number of T cells, suggesting that a given T cell must be able to recognize many different peptide-MHC (i.e. have a high level of cross-reactivity) (87, 88). However it is not clear how these

parameters can be measured. We propose an alternative calculation that allows estimation of the extent of cross-reactivity from the precursor frequency of T cells for pathogens – a parameter that can be reliably determined. Using slightly modified notation from (87) we first define four parameters:

R Repertoire (the number of clonotypically different naïve T cells in the repertoire)
r The number of different T cell clonotypes that will respond to the same peptide
N The total number of potentially immunogenic foreign peptides in the environment
n The number of different peptides to which a single T cell clonotype will respond

These four parameters are linked by the conservation equation (87):

$$rN = nR \qquad (3)$$

Lets suggest that a given pathogen has k epitopes to which T cells can mount a response. For a given T cell, the probability to recognize at least one epitope from a given pathogen could be written as:

$$1 - \left[1 - \frac{n}{N}\right]^k \qquad (4)$$

The probability that the same T cell will recognize at least one epitope on each of two pathogens with k epitopes (i.e., will be cross-reactive) will be a square of expression equation (4) and the probability to find at least one cross-reactive clone is equal to:

$$\left[1 - \left[1 - \frac{n}{N}\right]^k\right]^2 \times R \qquad (5)$$

Note, that in derived equation (5) we don't know the parameters k, n, and N which makes it very difficult to apply directly. Interestingly, for one epitope ($k = 1$) and with application of cross-reactivity equation (3) the equation (5) simplifies to:

$$\left[1 - \left[1 - \frac{n}{N}\right]\right]^2 \times R = \left[1 - \left[1 - \frac{r}{R}\right]\right]^2 \times R = \left[\frac{r}{R}\right]^2 \times R \quad (6)$$

Under the assumption that all naïve T cells able to respond to a given epitope are clonotypically different, which is supported by recent data (31, 80), we can think of r/R as a precursor frequency for a given epitope. The problem of estimating cross-reactivity in this case will be similar to the problem of estimating the probability of randomly choosing a two-colored ball from an urn when the frequencies of each of two colors are known. Interestingly, the measured naïve precursor frequencies for different immunogenic epitopes are similar for mice and humans and range from 1 to 100 cells per million cells (31). According to equation (3), this similarity immediately implies that cross-reactivity of each T cell receptor, n, is in the same range for mice and humans.

For the case when k > 1, we could not directly use the formula (5), due to unknown parameters, but can use a simple probabilistic calculation based on sampling multiple colored balls. We can write the frequency of cross-reactive cells between two randomly chosen pathogens (A and B) in terms of the precursor frequencies of T cells to these two pathogens, p_A and p_B, and the total number of cells T:

$$\text{Expected number of cross-reactive cells} = p_A p_B \times T \qquad (7)$$

and if we have clones of same size equal to T/R,

$$\text{Expected number of cross-reactive clones} = p_A p_B \times R \qquad (8)$$

where R equals the repertoire (the number of different T cell clones in each individual). When the average number of cross-reactive clones is less than one, equation (8) gives us the probability of observing a cross-reactive response between two pathogens in a single individual. We can use this framework together with the data on precursor frequencies (31) described in **Figure 2** to get an estimate of the extent of cross-reactivity between the responses to unrelated pathogens. As described earlier we have an approximate precursor frequency per epitope of 10^{-5}, and a precursor frequency per pathogen, with LCMV as an example, of about 4×10^{-5}. If we assume there are about 2×10^7 naïve CD8 T cells per mouse (89) then the number of cross-reactive cells between two unrelated pathogens will be ~0.032, which suggests cross-reactivity is very rare (a single cross-reactive clone will be found <4% of the time). In order to observe cross-reactivity between two random pathogens in a mouse we would need to have a precursor frequency per pathogen of at least 2.2×10^{-4}, assuming that precursor frequencies are similar for both pathogens. There are quite a few reported examples of cross-reactive T cell responses to different pathogens. In addition to the experiments of Selin and Welsh (69), cross-reactivity has been reported between influenza virus and hepatitis C virus (72), EBV (73) or HIV (74), LCMV and vaccinia virus (75), and coronavirus and human papillomavirus (76). It remains to be seen if the observed cases of cross-reactivity arise from a reporting bias (failure to observe cross-reactivity between two pathogens is unlikely to be reported) or because some of the assumptions of our model are incorrect and need to be modified. For example, we assume all T cell clones have the same level of cross-reactivity – and introducing heterogeneity may dramatically increase the chances to observe cross-reactivity.

Even if cross-reactive T cell responses to two specific pathogens are rare, the accumulation of many successive infections could result in fairly frequent cross-reactivity between a new pathogen and sum of all the pathogens the individual has previously encountered. This is what was observed when T cell precursor frequencies were measured for novel pathogens in blood from adults (80). The precursor frequency for cells recognizing a self-antigen or an unexposed viral epitope was the same as earlier estimated in mice and humans (31) and ranged from one to ten cells per million of CD4 T cells. The surprise was that over half of the precursor cells specific for novel pathogens such as HIV (to which an individual had never been exposed) were of the memory phenotype (80), suggesting that they may have arisen as a consequence of exposure to a different previously encountered pathogen(s). Alternatively, these memory cells could be pseudo-memory cells acquired via the process known as "homeostatic proliferation" and driven by

interaction with low-affinity self pMHCs that previously induced positive selection (90–94).

The Su et al. paper (80) raised an interesting question: why do memory cells invariably contribute about 50–80% of the precursors to pathogen the individual has never encountered? One possibility is that the memory repertoire is sufficiently large to be "complete." In this case if we draw the same amount of cells from either naïve or memory compartments (or mixture from both) we will have the same precursor frequency for a pathogen. Then the relative contribution of naïve and memory cells to precursors is equal to their relative frequencies, and is scaled by the stimulation threshold which is known to be lower for memory cells.

We note that our equation (7) allows an estimation of cross-reactivity for unrelated pathogens or peptides and, based on reported precursor frequencies for different epitopes, we expect cross-reactivity to be rare. Several studies allowed to estimate the rate of cross-reactivity for closely related peptides. Su et al. (80) identified potential pathogens responsible for generating T cells cross-reactive to HIV in HIV-negative individuals as follows: they generated clones from the HIV precursors and identified two epitopes to which these clones were specific. Then using a BLAST search of pathogen sequences they identified 24 sequences similar to the two HIV epitopes. About 21% of the HIV clones responded to two of the BLAST sequences corresponding to environmental pathogens. This number is comparable to result obtained in the earlier study, which showed that although the majority of 171 generated variant peptides of strongly immunodominant HLA-A2-restricted HIV gag epitope were able to bind HLA-A2, only one third were recognized by specific T cells (95). These two studies may give the rate of cross-reactivity for closely related peptides (21–33%) and could be particularly important in the context of a variable virus with an increased rate of mutations within epitopes (96).

Cross-reactive responses may be of clinical importance in the generation of pathology and autoimmunity. Several studies pointed that cross-reactivity may be the cause of increased immunopathology during successive unrelated viral infections (84–86) or as a result of application of T cell based therapy (97, 98). Expansion of cross-reactive T cell clones due to previous infections may underlie autoimmune diseases (99–101). Sometimes a pathogen epitope stimulates T cells in the context of a different MHC from the self-epitopes that react with these T cells, for example, Epstein-Barr virus EBN13-HLA-B8-specific cytotoxic T cells were shown to cross-react with a variety of self peptides presented by HLA-B35 (102). Together these observations point out that cross-reactive T cell responses might operate on different levels and much remains to be done to understand the extent of cross-reactivity and how it may differ in CD8, CD4, and regulatory populations of T cells.

4. SUMMARY

We have reviewed current estimates of the T cell repertoire and identified their key limitations. Further progress will require the development of methods to determine the pairing of TCR α and β chains and thus more accurate quantification of the T cell diversity. Current estimates raise the puzzling question of why the potential repertoire is many orders of magnitude greater than the realized repertoire. We suggest that existing hypotheses are not able to explain this puzzle and have proposed an alternative hypothesis of "evolutionary sloppiness."

One of the interesting observation that became obvious from our estimations is that precursor frequency per pathogen inherently links the TCR diversity and cross-reactivity which allows to predict the level of cross-reactivity between two random pathogens or unrelated peptides. Our estimates suggest that although cross-reactivity is a rare event for immunologically naïve individuals, probability to see the cross-reactive memory T cells becomes very high with an increase in successive infections.

Finally we note that we need to move from our current paradigm, which looks at each infection independent of other infections, to a more complex view that incorporates the terms for the interactions between the immune responses to different pathogens.

REFERENCES

1. Burnet F. A modification of Jerne's theory of antibody production using the concept of clonal selection. *CA Cancer J Clin* (1957) **26**:119–21. doi:10.3322/canjclin.26.2.119

2. Talmage DW. Allergy and immunology. *Annu Rev Med* (1957) **8**:239–56. doi:10.1146/annurev.me.08.020157.001323

3. Murali-Krishna K, Altman J, Suresh M, Sourdive D, Zajac A, Miller J, et al. Counting antigen-specific CD8+ T cells: a reevaluation of bystander activation during viral infection. *Immunity* (1998) **8**:177–87. doi:10.1016/S1074-7613(00)80470-7

4. Jamieson B, Ahmed R. T cell memory: long term persistence of virus specific cytotoxic cells. *J Exp Med* (1989) **169**:1993–2005. doi:10.1084/jem.169.6.1993

5. Miller JF, Sprent J. Cell-to-cell interaction in the immune response. VI. Contribution of thymus-derived cells and antibody-forming cell precursors to immunological memory. *J Exp Med* (1971) **134**(1):66–82. doi:10.1084/jem.134.1.66

6. Perelson AS. Modeling viral and immune system dynamics. *Nat Rev Immunol* (2002) **2**(1):28–36. doi:10.1038/nri700

7. Antia R, Ganusov VV, Ahmed R. The role of models in understanding CD8+ T-cell memory. *Nat Rev Immunol* (2005) **5**(2):101–11. doi:10.1038/nri1550

8. Regoes RR, Barber DL, Ahmed R, Antia R. Estimation of the rate of killing by cytotoxic T lymphocytes in vivo. *Proc Natl Acad Sci USA* (2007) **104**(5):1599–603. doi:10.1073/pnas.0508830104

9. Yates AJ, Van Baalen M, Antia R. Virus replication strategies and the critical CTL numbers required for the control of infection. *PLoS Comp Biol* (2011) **7**(11):e1002274. doi:10.1371/journal.pcbi.1002274

10. Sherwood AM, Desmarais C, Livingston RJ, Andriesen J, Haussler M, Carlson CS, et al. Deep sequencing of the human TCRγ and TCRβ repertoires suggests that TCRβ rearranges after $\alpha\beta$ and $\gamma\delta$ T cell commitment. *Sci Transl Med* (2011) **3**(90):90ra61.

11. Venturi V, Quigley MF, Greenaway HY, Ng PC, Ende ZS, McIntosh T, et al. A mechanism for TCR sharing between T cell subsets and individuals revealed by pyrosequencing. *J Immunol* (2011) **186**(7):4285–94. doi:10.4049/jimmunol.1003898

12. Hilborn R, Mangel M. The ecological detective: confronting models with data. In: Hilborn R, Mangel M, editors. *Monographs in Population Biology*. Princeton, NJ: Princeton University Press (1997). 28 p.

13. May RM. Uses and abuses of mathematics in biology. *Science* (2004) **303**(5659):790–3. doi:10.1126/science.1094442

14. Kurosawa Y, von Boehmer H, Haas W, Sakano H, Trauneker A, Tonegawa S. Identification of d segments of immunoglobulin heavy-chain genes and their rearrangement in T lymphocytes. *Nature* (1981) **290**(5807):565–70. doi:10.1038/290565a0

15. Hayday AC, Saito H, Gillies SD, Kranz DM, Tanigawa G, Eisen HN, et al. Structure, organization, and somatic rearrangement of T cell gamma genes. *Cell* (1985) **40**(2):259–69. doi:10.1016/0092-8674(85)90140-0

16. Garboczi DN, Ghosh P, Utz U, Fan QR, Biddison WE, Wiley DC. Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. *Nature* (1996) **384**(6605):134–41. doi:10.1038/384134a0

17. Oltz EM. Regulation of antigen receptor gene assembly in lymphocytes. *Immunol Res* (2001) **23**(2–3):121–33. doi:10.1385/IR:23:2-3:121

18. Jolly CJ, Cook AJL, Manis JP. Fixing DNA breaks during class switch recombination. *J Exp Med* (2008) **205**(3):509–13. doi:10.1084/jem.20080356

19. Davis MM, Bjorkman PJ. T-cell antigen receptor genes and T-cell recognition. *Nature* (1988) **334**(6181):395–402. doi:10.1038/334395a0

20. Robins HS, Srivastava SK, Campregher PV, Turtle CJ, Andriesen J, Riddell SR, et al. Overlap and effective size of the human CD8+ T cell receptor repertoire. *Sci Transl Med* (2010) **2**(47):47ra64. doi:10.1126/scitranslmed.3001442

21. Murugan A, Mora T, Walczak AM, Callan CG Jr. Statistical inference of the generation probability of T-cell receptors from sequence repertoires. *Proc Natl Acad Sci U S A* (2012) **109**(40):16161–6. doi:10.1073/pnas.1212755109

22. Arstila TP, Casrouge A, Baron V, Even J, Kanellopoulos J, Kourilsky P. A direct estimate of the human alphabeta T cell receptor diversity. *Science* (1999) **286**(5441):958–61. doi:10.1126/science.286.5441.958

23. Robins HS, Campregher PV, Srivastava SK, Wacher A, Turtle CJ, Kahsai O, et al. Comprehensive assessment of T-cell receptor beta-chain diversity in alphabeta T cells. *Blood* (2009) **114**(19):4099–107. doi:10.1182/blood-2009-04-217604

24. Warren RL, Freeman JD, Zeng T, Choe G, Munro S, Moore R, et al. Exhaustive T-cell repertoire sequencing of human peripheral blood samples reveals signatures of antigen selection and a directly measured repertoire size of at least 1 million clonotypes. *Genome Res* (2011) **21**(5):790–7. doi:10.1101/gr.115428.110

25. Kesmir C, Borghans JA, de Boer RJ. Diversity of human $\alpha\beta$ T cell receptors. *Science* (2000) **288**:1135. doi:10.1126/science.288.5469.1135a

26. Arstila TP, Casrouge A, Baron V, Even J, Kanellopoulos J, Kourilsky P. Diversity of human alpha beta T cell receptors. *Science* (2000) **288**(5469):1135. doi:10.1126/science.288.5469.1135a

27. Zhang L, Lewin SR, Markowitz M, Lin HH, Skulsky E, Karanicolas R, et al. Measuring recent thymic emigrants in blood of normal and HIV-1-infected individuals before and after effective therapy. *J Exp Med* (1999) **190**(5):725–32. doi:10.1084/jem.190.5.725

28. Bains I, Antia R, Callard R, Yates AJ. Quantifying the development of the peripheral naive CD4+ T-cell pool in humans. *Blood* (2009) **113**(22):5480–7. doi:10.1182/blood-2008-10-184184

29. McFarland RD, Douek DC, Koup RA, Picker LJ. Identification of a human recent thymic emigrant phenotype. *Proc Natl Acad Sci U S A* (2000) **97**(8):4215–20. doi:10.1073/pnas.070061597

30. Casrouge A, Beaudoing E, Dalle S, Pannetier C, Kanellopoulos J, Kourilsky P. Size estimate of the alpha beta TCR repertoire of naive mouse splenocytes. *J Immunol* (2000) **164**(11):5782–7.

31. Jenkins MK, Moon JJ. The role of naive T cell precursor frequency and recruitment in dictating immune response magnitude. *J Immunol* (2012) **188**(9):4135–40. doi:10.4049/jimmunol.1102661

32. Tubo NJ, Pagán AJ, Taylor JJ, Nelson RW, Linehan JL, Ertelt JM, et al. Single naive CD4(+) T cells from a diverse repertoire produce different effector cell types during infection. *Cell* (2013) **153**(4):785–96. doi:10.1016/j.cell.2013.04.007

33. Venturi V, Price DA, Douek DC, Davenport MP. The molecular basis for public T cell responses? *Nat Rev Immunol* (2008) **8**(3):231–8. doi:10.1038/nri2260

34. Dash P, McClaren JL, Oguin TH III, Rothwell W, Todd B, Morris MY, et al. Paired analysis of TCRα and TCRβ chains at the single-cell level in mice. *J Clin Invest* (2011) **121**(1):288–95. doi:10.1172/JCI44752

35. Turchaninova MA, Britanova OV, Bolotin DA, Shugay M, Putintseva EV, Staroverov DB, et al. Pairing of T-cell receptor chains via emulsion PCR. *Eur J Immunol* (2013) **43**(9):2507–15. doi:10.1002/eji.201343453

36. 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**(1–2):182–95. doi:10.1016/j.jim.2007.01.019

37. Bousso P, Casrouge A, Altman JD, Haury M, Kanellopoulos J, Abastado JP, et al. Individual variations in the murine T cell response to a specific peptide reflect variability in naive repertoires. *Immunity* (1998) **9**(2):169–78. doi:10.1016/S1074-7613(00)80599-3

38. Quigley MF, Greenaway HY, Venturi V, Lindsay R, Quinn KM, Seder RA, et al. Convergent recombination shapes the clonotypic landscape of the naive T-cell repertoire. *Proc Natl Acad Sci U S A* (2010) **107**(45):19414–9. doi:10.1073/pnas.1010586107

39. Livak F, Burtrum DB, Rowen L, Schatz DG, Petrie HT. Genetic modulation of T cell receptor gene segment usage during somatic recombination. *J Exp Med* (2000) **192**(8):1191–6. doi:10.1084/jem.192.8.1191

40. Brändle D, Bürki K, Wallace VA, Rohrer UH, Mak TW, Malissen B, et al. Involvement of both T cell receptor V alpha and V beta variable region domains and alpha chain junctional region in viral antigen recognition. *Eur J Immunol* (1991) **21**(9):2195–202. doi:10.1002/eji.1830210930

41. Perkins DL, Wang YS, Fruman D, Seidman JG, Rimm IJ. Immunodominance is altered in T cell receptor (beta-chain) transgenic mice without the generation of a hole in the repertoire. *J Immunol* (1991) **146**(9):2960–4.

42. Turner SJ, Cose SC, Carbone FR. TCR alpha-chain usage can determine antigen-selected TCR beta-chain repertoire diversity. *J Immunol* (1996) **157**(11):4979–85.

43. Burns RP Jr, Natarajan K, LoCascio NJ, O'Brien DP, Kobori JA, Shastri N, et al. Molecular analysis of skewed Tcra-V gene use in T-cell receptor beta-chain transgenic mice. *Immunogenetics* (1998) **47**(2):107–14. doi:10.1007/s002510050335

44. De Boer RJ, Perelson AS. How diverse should the immune system be? *Proc R Soc Lond B Biol Sci* (1993) **252**(1335):171–5. doi:10.1098/rspb.1993.0062

45. Wiegel FW, Perelson AS. Some scaling principles for the immune system. *Immunol Cell Biol* (2004) **82**(2):127–31. doi:10.1046/j.0818-9641.2004.01229.x

46. Blattman JN, Antia R, Sourdive DJ, Wang X, Kaech SM, Murali-Krishna K, et al. Estimating the precursor frequency of naive antigen-specific CD8 T cells. *J Exp Med* (2002) **195**(5):657–64. doi:10.1084/jem.20001021

47. Whitmire JK, Benning N, Whitton JL. Precursor frequency, nonlinear proliferation, and functional maturation of virus-specific CD4+ T cells. *J Immunol* (2006) **176**(5):3028–36.

48. Moon JJ, Chu HH, Pepper M, McSorley SJ, Jameson SC, Kedl RM, et al. Naive CD4(+) T cell frequency varies for different epitopes and predicts repertoire diversity and response magnitude. *Immunity* (2007) **27**(2):203–13. doi:10.1016/j.immuni.2007.07.007

49. Thomas PG, Handel A, Doherty PC, La Gruta NL. Ecological analysis of antigen-specific CTL repertoires defines the relationship between naive and immune T-cell populations. *Proc Natl Acad Sci U S A* (2013) **110**(5):1839–44. doi:10.1073/pnas.1222149110

50. La Gruta NL, Rothwell WT, Cukalac T, Swan NG, Valkenburg SA, Kedzierska K, et al. Primary CTL response magnitude in mice is determined by the extent of naive T cell recruitment and subsequent clonal expansion. *J Clin Invest* (2010) **120**(6):1885–94. doi:10.1172/JCI41538

51. Yewdell JW, Bennink JR. Immunodominance in major histocompatibility complex class I-restricted T lymphocyte responses. *Annu Rev Immunol* (1999) **17**:51–88. doi:10.1146/annurev.immunol.17.1.51

52. Langman R, Cohen M. The E-T (elephant-tadpole) paradox necessitates the concept of a unit of B-cell function: the protection. *Mol Immunol* (1987) **24**:675–97. doi:10.1016/0161-5890(87)90050-2

53. Masopust D, Murali-Krishna K, Ahmed R. Quantitating the magnitude of the lymphocytic choriomeningitis virus-specific CD8 T-cell response: it is even bigger than we thought. *J Virol* (2007) **81**(4):2002–11. doi:10.1128/JVI.01459-06

54. Perelson AS, Wiegel FW. Scaling aspects of lymphocyte trafficking. *J Theor Biol* (2009) **257**(1):9–16. doi:10.1016/j.jtbi.2008.11.007

55. Westermann J, Pabst R. Lymphocyte subsets in the blood: a diagnostic window on the lymphoid system? *Immunol Today* (1990) **11**(11):406–10. doi:10.1016/0167-5699(90)90160-B

56. Clark DR, de Boer RJ, Wolthers KC, Miedema F. T cell dynamics in HIV-1 infection. *Adv Immunol* (1999) **73**:301–27. doi:10.1016/S0065-2776(08) 60789-0

57. Shortman K, Vremec D, Egerton M. The kinetics of T cell antigen receptor expression by subgroups of CD4+8+ thymocytes: delineation of CD4+8+3(2+) thymocytes as post-selection intermediates leading to mature T cells. *J Exp Med* (1991) **173**(2):323–32. doi:10.1084/jem.173.2.323

58. Huesmann M, Scott B, Kisielow P, von Boehmer H. Kinetics and efficacy of positive selection in the thymus of normal and T cell receptor transgenic mice. *Cell* (1991) **66**(3):533–40. doi:10.1016/0092-8674(81)90016-7

59. Vrisekoop N, den Braber I, de Boer AB, Ruiter AFC, Ackermans MT, van der Crabben SN, et al. Sparse production but preferential incorporation of recently produced naive T cells in the human peripheral pool. *Proc Natl Acad Sci U S A* (2008) **105**(16):6115–20. doi:10.1073/pnas.0709713105

60. Hazenberg MD, Borghans JA, de Boer RJ, Miedema F. Thymic output: a bad TREC record. *Nat Immunol* (2003) **4**(2):97–9. doi:10.1038/ni0203-97

61. Tough D, Sprent J. Turnover of naive- and memory-phenotype T cells. *J Exp Med* (1994) **179**:1127–35. doi:10.1084/jem.179.4.1127

62. Scollay RG, Butcher EC, Weissman IL. Thymus cell migration quantitative aspects of cellular traffic from the thymus to the periphery in mice. *Eur J Immunol* (1980) **10**(3):210–8. doi:10.1002/eji.1830100310

63. Gabor MJ, Scollay R, Godfrey DI. Thymic T cell export is not influenced by the peripheral T cell pool. *Eur J Immunol* (1997) **27**(11):2986–93. doi:10.1002/eji. 1830271135

64. Vezys V, Masopust D, Kemball CC, Barber DL, O'Mara LA, Larsen CP, et al. Continuous recruitment of naive T cells contributes to heterogeneity of antiviral CD8 T cells during persistent infection. *J Exp Med* (2006) **203**(10):2263–9. doi:10.1084/jem.20060995

65. Yang Y, Al-Mozaini M, Buzon MJ, Beamon J, Ferrando-Martinez S, Ruiz-Mateos E, et al. CD4 T-cell regeneration in HIV-1 elite controllers. *AIDS* (2012) **26**(6):701–6. doi:10.1097/QAD.0b013e3283519b22

66. Hatzakis A, Touloumi G, Karanicolas R, Karafoulidou A, Mandalaki T, Anastassopoulou C, et al. Effect of recent thymic emigrants on progression of HIV-1 disease. *Lancet* (2000) **355**(9204):599–604. doi:10.1016/S0140-6736(99) 10311-8

67. Gould S, Lewontin R. The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme. *Proc R Soc Lond B Biol Sci* (1979) **205**:581–98. doi:10.1098/rspb.1979.0086

68. George A, Ritter M. Thymic involution with ageing: obsolescence or good housekeeping? *Immunol Today* (1996) **17**:267–72. doi:10.1016/0167-5699(96) 80543-3

69. Selin LK, Nahill SR, Welsh RM. Cross-reactivities in memory cytotoxic T lymphocyte recognition of heterologous viruses. *J Exp Med* (1994) **179**(6):1933–43. doi:10.1084/jem.179.6.1933

70. Selin LK, Brehm MA, Naumov YN, Cornberg M, Kim S-K, Clute SC, et al. Memory of mice and men: CD8+ T-cell cross-reactivity and heterologous immunity. *Immunol Rev* (2006) **211**:164–81. doi:10.1111/j.0105-2896.2006.00394.x

71. Welsh RM, Selin LK. No one is naive: the significance of heterologous T-cell immunity. *Nat Rev Immunol* (2002) **2**(6):417–26.

72. Wedemeyer H, Mizukoshi E, Davis AR, Bennink JR, Rehermann B. Cross-reactivity between hepatitis c virus and influenza a virus determinant-specific cytotoxic T cells. *J Virol* (2001) **75**(23):11392–400. doi:10.1128/JVI.75.23. 11392-11400.2001

73. Clute SC, Watkin LB, Cornberg M, Naumov YN, Sullivan JL, Luzuriaga K, et al. Cross-reactive influenza virus-specific CD8+ T cells contribute to lymphoproliferation in Epstein-Barr virus-associated infectious mononucleosis. *J Clin Invest* (2005) **115**(12):3602–12. doi:10.1172/JCI25078

74. Acierno PM, Newton DA, Brown EA, Maes LA, Baatz JE, Gattoni-Celli S. Cross-reactivity between HLA-A2-restricted FLU-M1:58-66 and HIV p17 GAG:77-85 epitopes in HIV-infected and uninfected individuals. *J Transl Med* (2003) **1**(1):3. doi:10.1186/1479-5876-1-3

75. Kim S-K, Cornberg M, Wang XZ, Chen HD, Selin LK, Welsh RM. Private specificities of CD8 T cell responses control patterns of heterologous immunity. *J Exp Med* (2005) **201**(4):523–33. doi:10.1084/jem.20041337

76. Nilges K, Höhn H, Pilch H, Neukirch C, Freitag K, Talbot PJ, et al. Human papillomavirus type 16 E7 peptide-directed CD8+ T cells from patients with cervical cancer are cross-reactive with the coronavirus NS2 protein. *J Virol* (2003) **77**(9):5464–74. doi:10.1128/JVI.77.9.5464-5474.2003

77. Reinherz EL, Tan K, Tang L, Kern P, Liu J, Xiong Y, et al. The crystal structure of a T cell receptor in complex with peptide and MHC class II. *Science* (1999) **286**(5446):1913–21. doi:10.1126/science.286.5446.1913

78. Reiser J-B, Darnault C, Grégoire C, Mosser T, Mazza G, Kearney A, et al. CDR3 loop flexibility contributes to the degeneracy of TCR recognition. *Nat Immunol* (2003) **4**(3):241–7. doi:10.1038/ni891

79. Newell EW, Ely LK, Kruse AC, Reay PA, Rodriguez SN, Lin AE, et al. Structural basis of specificity and cross-reactivity in T cell receptors specific for cytochrome c-I-E(k). *J Immunol* (2011) **186**(10):5823–32. doi:10.4049/ jimmunol.1100197

80. Su LF, Kidd BA, Han A, Kotzin JJ, Davis MM. Virus-specific CD4(+) memory-phenotype T cells are abundant in unexposed adults. *Immunity* (2013) **38**(2):373–83. doi:10.1016/j.immuni.2012.10.021

81. Heath WR, Miller JF. Expression of two alpha chains on the surface of T cells in T cell receptor transgenic mice. *J Exp Med* (1993) **178**(5):1807–11. doi:10.1084/jem.178.5.1807

82. Padovan E, Casorati G, Dellabona P, Meyer S, Brockhaus M, Lanzavecchia A. Expression of two T cell receptor alpha chains: dual receptor T cells. *Science* (1993) **262**(5132):422–4. doi:10.1126/science.8211163

83. Petrie HT, Livak F, Schatz DG, Strasser A, Crispe IN, Shortman K. Multiple rearrangements in T cell receptor alpha chain genes maximize the production of useful thymocytes. *J Exp Med* (1993) **178**(2):615–22. doi:10.1084/jem.178. 2.615

84. Selin LK, Varga SM, Wong IC, Welsh RM. Protective heterologous antiviral immunity and enhanced immunopathogenesis mediated by memory T cell populations. *J Exp Med* (1998) **188**(9):1705–15. doi:10.1084/jem.188.9.1705

85. Yang HY, Joris I, Majno G, Welsh RM. Necrosis of adipose tissue induced by sequential infections with unrelated viruses. *Am J Pathol* (1985) **120**(2): 173–7.

86. Chen HD, Fraire AE, Joris I, Brehm MA, Welsh RM, Selin LK. Memory CD8+ T cells in heterologous antiviral immunity and immunopathology in the lung. *Nat Immunol* (2001) **2**(11):1067–76. doi:10.1038/ni727

87. Mason D. A very high level of crossreactivity is an essential feature of the T-cell receptor. *Immunol Today* (1998) **19**(9):395–404. doi:10.1016/S0167-5699(98) 01299-7

88. Sewell AK. Why must T cells be cross-reactive? *Nat Rev Immunol* (2012) **12**(9):669–77. doi:10.1038/nri3279

89. Jenkins MK, Chu HH, McLachlan JB, Moon JJ. On the composition of the preimmune repertoire of T cells specific for peptide-major histocompatibility complex ligands. *Annu Rev Immunol* (2010) **28**:275–94. doi:10.1146/annurev-immunol-030409-101253

90. Murali-Krishna K, Ahmed R. Cutting edge: naive T cells masquerading as memory cells. *J Immunol* (2000) **165**(4):1733–7.

91. Ernst B, Lee DS, Chang JM, Sprent J, Surh CD. The peptide ligands mediating positive selection in the thymus control T cell survival and homeostatic proliferation in the periphery. *Immunity* (1999) **11**(2):173–81. doi:10.1016/S1074-7613(00)80092-8

92. Goldrath AW, Bevan MJ. Low-affinity ligands for the TCR drive proliferation of mature CD8+ T cells in lymphopenic hosts. *Immunity* (1999) **11**(2):183–90. doi:10.1016/S1074-7613(00)80093-X

93. Kieper WC, Jameson SC. Homeostatic expansion and phenotypic conversion of naïve T cells in response to self peptide/MHC ligands. *Proc Natl Acad Sci U S A* (1999) **96**(23):13306–11. doi:10.1073/pnas.96.23.13306

94. Muranski P, Chmielowski B, Ignatowicz L. Mature CD4+ T cells perceive a positively selecting class II MHC/peptide complex in the periphery. *J Immunol* (2000) **164**(6):3087–94.

95. Lee JK, Stewart-Jones G, Dong T, Harlos K, Di Gleria K, Dorrell L, et al. T cell cross-reactivity and conformational changes during TCR engagement. *J Exp Med* (2004) **200**(11):1455–66. doi:10.1084/jem.20041251

96. Kosmrlj A, Read EL, Qi Y, Allen TM, Altfeld M, Deeks SG, et al. Effects of thymic selection of the T-cell repertoire on HLA class I-associated control of HIV infection. *Nature* (2010) **465**(7296):350–4. doi:10.1038/nature08997

97. Cameron BJ, Gerry AB, Dukes J, Harper JV, Kannan V, Bianchi FC, et al. Identification of a titin-derived HLA-A1-presented peptide as a cross-reactive target for engineered MAGE A3-directed T cells. *Sci Transl Med* (2013) **5**(197):197ra103. doi:10.1126/scitranslmed.3006034

98. Linette GP, Stadtmauer EA, Maus MV, Rapoport AP, Levine BL, Emery L, et al. Cardiovascular toxicity and titin cross-reactivity of affinity-enhanced T cells

in myeloma and melanoma. *Blood* (2013) **122**(6):863–71. doi:10.1182/blood-2013-03-490565

99. Fujinami RS, Oldstone MB. Amino acid homology between the encephalitogenic site of myelin basic protein and virus: mechanism for autoimmunity. *Science* (1985) **230**(4729):1043–5. doi:10.1126/science.2414848

100. Oldstone MB. Molecular mimicry and immune-mediated diseases. *FASEB J* (1998) **12**(13):1255–65.

101. Wucherpfennig KW, Strominger JL. Molecular mimicry in T cell-mediated autoimmunity: viral peptides activate human T cell clones specific for myelin basic protein. *Cell* (1995) **80**(5):695–705. doi:10.1016/0092-8674(95)90348-8

102. Burrows SR, Silins SL, Khanna R, Burrows JM, Rischmueller M, McCluskey J, et al. Cross-reactive memory T cells for Epstein-Barr virus augment the alloresponse to common human leukocyte antigens: degenerate recognition of major histocompatibility complex-bound peptide by T cells and its role in alloreactivity. *Eur J Immunol* (1997) **27**(7):1726–36. doi:10.1002/eji.1830270720