ELSEVIER

Journal of Immunological Methods 321 (2007) 182–195

JIM
Journal of Immunological Methods

www.elsevier.com/locate/jim

Research paper

Methods for comparing the diversity of samples of the T cell receptor repertoire

Vanessa Venturi [a,*], Katherine Kedzierska [b], Stephen J. Turner [b], Peter C. Doherty [b,c], Miles P. Davenport [a]

[a] *Department of Haematology, Prince of Wales Hospital and, Centre for Vascular Research, University of New South Wales, Kensington NSW 2052, Australia*
[b] *Department of Microbiology and Immunology, University of Melbourne, Parkville 3010, Melbourne, Australia*
[c] *Department of Immunology, St. Jude Children's Research Hospital, Memphis, TN, USA*

Received 21 November 2006; received in revised form 19 January 2007; accepted 21 January 2007
Available online 21 February 2007

Abstract

Analysis of T cell receptor (TCR) data has become a crucial element in many studies aimed at better understanding the evolution of the T cell repertoire and the role of TCR diversity in immune responses. In this paper we focus on comparing the diversity between samples of the TCR repertoire. We discuss some of the limitations and potential problems inherent in some of the more popular approaches to comparing samples of the TCR repertoire and we suggest alternate methods that both avoid these problems and enrich the analysis of TCR data. Examples from published studies of the $CD8^+$ T cell responses to the influenza A virus D^bNP_{366} and D^bPA_{224} epitopes in mice are used to demonstrate the implementation of these methods. One example involves a comparison between the central and effector memory T cell subsets, defined on the basis of CD62L expression, and the other examines changes in the TCR repertoire over time.

Keywords: T cell receptor; Repertoire; Diversity; Immune memory

1. Introduction

The immune response to infection involves the recognition of bacterial and viral proteins by both antibodies secreted from B cells and by T cell receptors (TCRs) on the surface of T cells, which act to neutralize or kill the infectious agent. Antibodies and TCRs comprise a unique family of recognition molecules that are formed by random recombination of genetic elements during the maturation of the B cell or T cell. For example, it has been estimated for mice that the number of different TCRs that can be formed is $>10^{15}$ (Davis and Bjorkman, 1988) and that there are $\sim 2\times10^6$ different TCRs (Casrouge et al., 2000) in an uninfected mouse. For humans, an estimated $>10^{18}$ different TCRs can be produced (Janeway et al., 2005) and the diversity of TCRs in a human at any given time has been estimated to be $>2\times10^7$ (Arstila et al., 1999; Naylor et al., 2005). This enormous diversity allows the recognition of many different infectious agents.

Abbreviations: TCR, T cell receptor; SIV, Simian Immunodeficiency Virus; NP, viral nucleoprotein; PA, viral acid polymerase; CDR, complementarity determining region.

* Corresponding author. Tel.: +61 2 9385 8234; fax: +61 2 9385 1389.
E-mail address: v.venturi@unsw.edu.au (V. Venturi).

0022-1759/$ - see front matter © 2007 Elsevier B.V. All rights reserved.
doi:10.1016/j.jim.2007.01.019

Most TCR repertoire studies focus on either the nucleotide or amino acid sequence of the CDR3 of the TCR α- and/or β-chain. A distinct TCR sequence is referred to as a clonotype and the number of copies of that particular sequence is referred to as its clone size. A subset of around 15 to 900 different TCRs responds to a given epitope (Maryanski et al., 1996; Pewe et al., 1999). The responding T cells may divide many times and increase in number >1000 fold (Butz and Bevan, 1998). The immune response to a particular viral epitope is sometimes highly focused and dominated by cells of a particular TCR clonotype, presumably due to some growth advantage for these cells. This phenomenon of the preponderance of T cells of a single clonotype in a particular response is known as 'clonal dominance'. In other cases, no single clonotype dominates and many clonotypes make a comparable contribution to the response, resulting in a relatively even distribution of clone sizes (Turner et al., 2003; Kedzierska et al., 2004).

Recent developments in sorting and purifying T cells combined with developments in high throughput TCR sequencing have led to a large number of studies of TCR repertoires in immune responses. Some studies have endeavoured to estimate the total diversity of the repertoire, either prior to infection (Wagner et al., 1998; Arstila et al., 1999; Casrouge et al., 2000; Bouneaud et al., 2005; Naylor et al., 2005) or responding to a particular epitope (Pewe et al., 1999; Blattman et al., 2002; Pewe et al., 2004). However, most studies investigate changes in the TCR repertoire over time or with infection by comparing samples of the repertoire.

There are many factors (for example: ageing and persistent infection) that can alter the 'shape' of the TCR repertoire (Nikolich-Zugich et al., 2004) and thus influence the immune response to a particular epitope. The diversity or clonal dominance of an immune response is important for limiting the growth of some viruses. It has been suggested that a highly focused immune response facilitates the 'escape' from T cell recognition of viral mutants (Price et al., 2004). In this study of Simian Immunodeficiency Virus (SIV) in macaques, investigators compared the diversity of T cell responses in the same animals to two viral epitopes, one that did and one that did not escape immune recognition. This comparison determined that the T cell response to the viral epitope that did escape was significantly less diverse (i.e.: more focused), in terms of the number of different clonotypes, than the response to the viral epitope that did not escape.

Another fundamental question in immunology relates to the development and maintenance of T cell memory. A large number of studies have investigated the relationship between the T cell clonotypes in the acute response to infection and those that survive into memory. Some studies have reported that entry into the pool of long-lived memory cells is simply stochastic, implying that the memory pool is a representative snapshot of the acute response (Sourdive et al., 1998; Blattman et al., 2000). It has been suggested by other studies that there is immune focusing in the memory pool, with mostly dominant clonotypes surviving into the memory pool (McHeyzer-Williams and Davis, 1995; Bachmann et al., 1997; Busch et al., 1998; Busch and Pamer, 1999; Savage et al., 1999), while other studies have suggested that there is a turnover of different clonotypes from the acute response to the memory response (Callan et al., 2000; Davenport et al., 2002). Many studies have also shown that the memory T cell response is not homogenous, and instead comprises two subsets of 'central memory' and 'effector memory' cells ((Sallusto et al., 1999) and subsequent studies reviewed in (Sallusto et al., 2004) and (Lefrancois, 2006). These cells can be identified based on the expression of different surface molecules and also on their function in subsequent infection. Central memory cells are $CD62L^{hi}$ (i.e.: express high levels of CD62L), predominantly localized to secondary lymphoid organs, and are thought to have poor immediate function and high proliferative potential. By contrast, effector memory cells are $CD62L^{lo}$ (i.e.: express low levels of CD62L), more widely distributed in tissues, and have rapid effector function but low proliferative capacity. There is also significant debate about the origins of these two types of memory T cells, with investigators arguing that they are either from different or the same lineages (reviewed in Lefrancois and Marzo, 2006).

Analysis of TCR data during physiological responses to infection is fundamental in addressing many basic questions on immune recognition and response. However, appropriate analytical tools are required to facilitate this analysis. Many of the methods currently used can be strongly affected by sampling constraints, since (i) usually only a small proportion of the total repertoire can be sampled, and (ii) due to methodological limitations, the sample size is often different between the samples requiring comparison. Another concern is that many studies focus on the number of T cell clonotypes contributing to an immune response without accounting for the clonal dominance of the response. The extent of clonal dominance may play an important role in the control of infection or immune escape. An additional concern in some studies is that differences or similarities between TCR samples are reported without any statistical assurance that the observed differences or similarities are

significant (i.e.: How likely was it that the observed differences or similarities could have been achieved by chance, given that only a small and random sample of the total TCR repertoire was analysed?). In this paper we present some simple and robust methods for comparing the diversity of TCR repertoires involved in immune responses, and apply these to some topical questions about T cell memory.

2. Measures of diversity and dominance: Simpson's diversity index

TCR repertoire studies often rely upon counting the number of different clonotypes (either nucleotide or amino acid sequences) as the measure of diversity to compare between TCR samples. While the number of different clonotypes is a useful measure of diversity, it is highly sensitive to sample size and it neglects the relative abundance of the clonotypes, or the clone size distribution. For example, a response in which 91% of responding cells are of one clonotype and 9 other clonotypes contribute 1% has a total of ten clonotypes. By clonotype counting alone, this would be considered as equally diverse as a response in which 10 clonotypes each contributed 10% of the total repertoire. The clone size distribution contains important information about the TCR repertoire, and in particular is important for understanding clonal dominance. We suggest the use of diversity measures that better utilize information on the clone size distribution in conjunction with the number of different clonotypes.

The field of ecology has developed many different measures of diversity (reviewed in Magurran, 2004) that incorporate information on both the number of different species and the number of individuals of each species. These diversity measures often have a different sensitivity to the variety of species and the dominance or the evenness of the abundances of the species. Each diversity measure also has a different sensitivity to sample size and how well the sample represents the population.

We have chosen to use the Simpson's diversity index, which has widely been used in ecology (Magurran, 2004), population genetics (Nei, 1973), and molecular epidemiology (Hunter and Gaston, 1988) studies, and has previously been used in immunological studies of TCR repertoire (Mahajan et al., 2005) and immune responses to viral strain diversity (Nowak and May, 1992). This diversity index has the advantages that it has a direct interpretation, is easy to calculate, is not highly sensitive to sample size, and can be used to compare diversities with smaller sample sizes, proportional to the total population size, than required by other methods. The latter point is an important one for TCR repertoire studies, because the size of repertoire means that only a very small portion of the repertoire is usually sampled. Simpson's diversity index also tends to be more sensitive to the dominant clonotypes and less sensitive to the number of clonotypes (Magurran, 2004), making it an appropriate measure of diversity if we wish to assess clonal dominance.

In the context of TCR repertoire studies, Simpson's diversity index is the probability that any two TCRs chosen at random from the sample will have different clonotypes (i.e.: distinct CDR3 sequences for the TCR α- or β-chain):

$$D_S = 1 - \sum_{i=1}^{c} \frac{n_i(n_i-1)}{n(n-1)},$$

where n_i is the clone size of the ith clonotype (i.e.: number of copies of each clonotype), c is the number of different clonotypes in the TCR sample, and n is the total number of TCR sequences sampled. This index ranges between 0 and 1, with 0 and 1 representing the minimal and maximal diversity, respectively. A sample of TCRs containing more than one clonotype will have maximum diversity when all clonotypes have equal clone sizes.

3. Comparing many samples of the T cell receptor repertoire

3.1. Overcoming the problem of unequal sample sizes

The experimental methods involved in obtaining TCR repertoire data make it difficult to obtain consistent numbers of TCR sequences across all individuals or time points sampled, and for many studies the differences in sample size are large. Differences in sample size are a particular problem for repertoire studies because the sample sizes are relatively small compared with the size of the total repertoire and many of the popular approaches to both qualitatively or quantitatively studying diversity can be biased by differences in sample size. This will be demonstrated in one of the examples to which the proposed methods are applied later in the paper.

In designing an ideal study we therefore recommend that the sample size be standardized at the time of collection. When this is not possible, then the approach of estimating the diversity for a reduced, standard sample size is useful. That is, when two samples of different size are compared the sample size of the larger

TCR set can be reduced by randomly drawing, without replacement, a subset of the same size as the smaller sample. We can then apply diversity measures to this sampled TCR subset. By performing this random draw multiple times and estimating the diversity for each subset we obtain a distribution of the reduced TCR set diversity (Fig. 1). The median of the distribution provides an estimate of the diversity of the larger TCR set if the same number of TCR sequences had been sampled as in the smaller TCR set.

For studies comparing multiple samples, the recommended approach is to standardize the number of TCRs in every sample down to the smallest number of TCRs in any sample. However, this requires that all samples have a reasonable sample size or that there are a sufficient number of samples to discard those with very small sample sizes. In this approach, we use the observed diversity for the sample that has the smallest sample size, and for all other samples, we estimate the diversity had the samples been of the same size as the smallest sample (i.e.: the median of the diversity distribution resulting from the random sampling of a subset without replacement).

With diversity estimates for all samples, had they been of equal sample size, we can apply various statistical tests to compare the diversity measures between different groups of samples or different subsets of the TCR repertoire.

3.2. Are the diversities of samples of the TCR repertoire in one group significantly higher than in another group?

There are many immunological questions which require comparing the diversity of TCR samples taken from two sources. One such example is: Are virus-specific repertoires post-infection more or less diverse in vaccinated animals than in unvaccinated animals? When diversity measures have been estimated for a standard sample size across all samples (described in Section 3.1), we can compare the diversity measures for the samples in the two groups using a Mann–Whitney test. A minimum of four samples in each group is required to obtain a two-tailed p-value less than 0.05 with a Mann–Whitney test.

3.3. Are the diversities of paired samples of the TCR repertoire significantly different?

In many situations, the samples of the TCR repertoire will be paired. These are usually two samples of different

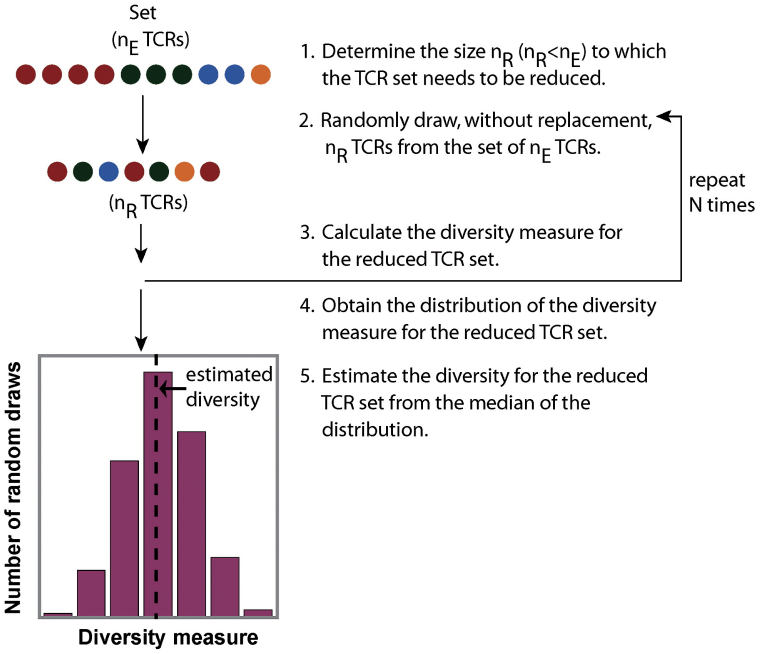

Fig. 1. Schematic of the method for estimating diversity measures for samples of reduced sample size. In this case a certain number of TCRs were observed experimentally (n_E), however we wish to know what the diversity would have been if we had obtained a smaller number of TCRs (n_R). This involves randomly drawing a sample of size n_R from the n_E TCRs, and calculating the diversity measure for this smaller sample set. Drawing the reduced number of TCRs many times, and estimating the diversity each time (i.e.: repeat steps 2 and 3 multiple times), we obtain a distribution of the diversity measures for the reduced TCR set.

TCR subsets within one animal (for example: $CD62L^{hi}$ and $CD62L^{lo}$ T cell subsets for central and effector memory), or samples from the same animal at two time points (for example: pre-infection and post-infection repertoires).

As discussed in Section 3.1, the recommended approach is to standardize the number of TCRs in every sample down to the smallest number of TCRs in any sample. When the sample size has been standardized across all sample pairs it is appropriate to apply a Wilcoxon signed rank sum test to determine the significance of the differences in diversity between the two subsets of the TCR repertoire. The Wilcoxon test requires a minimum of six pairs of samples to determine a two-tailed p-value less than 0.05.

There is another option available for pair-wise comparisons of TCR samples for cases where there are several samples with small sample sizes and discarding these would result in too few paired samples to perform a Wilcoxon test. The sign test is less powerful than a Wilcoxon test, so it is preferable to use the Wilcoxon test where possible. The sign test option requires that the diversity is estimated for standard sample sizes between the two samples in each pair. This involves reducing the sample size for the larger TCR set to the same size as the smaller TCR set in each pair and estimating the diversity (as described in Section 3.1) (Note that the major methodological difference here is that rather than reduce the size of all samples down to that of the smallest sample as described earlier, here we work only within a pair of samples and reduce the size of the larger sample down to that of the smaller sample within the pair). For each pair the observed diversity of the smaller sample is compared with the estimated diversity of a reduced set of the larger sample and it can be concluded which sample had the higher or lower diversity. A sign test can be performed to see whether the diversity for sample A is higher than for sample B in more of the paired samples than expected. That is, if there is truly no difference in diversity, then we expect that sample A will have higher diversity than sample B in half of the paired samples. The null hypothesis of the sign test assumes that both samples have an equal probability of having the higher diversity and uses a binomial distribution, with $p=q=0.5$, to determine the probability that the diversity for sample A would be higher than for sample B in as many paired samples as observed. A minimum of six pairs of samples is required to determine a two-tailed p-value less than 0.05.

3.4. Does the difference in diversity between paired TCR samples increase or decrease with the change in some experimental variable?

An additional advantage of having sufficiently large sample sizes to standardize the sample size across all samples is that when the samples correspond with some additional experimental variable, such as time, a Spearman rank correlation can be used to evaluate correlations between this variable and the difference in diversity. For example, we may be interested in determining whether the difference in diversity between the $CD62L^{hi}$ and $CD62L^{lo}$ subsets for central and effector memory TCR repertoire increases or decreases with the time since infection.

4. Example: Comparing the diversity between many samples of central and effector T cell memory in immune responses to influenza in mice

The early immune response to influenza virus in C57BL/6J mice is dominated by the $CD8^+$ T cell responses to two viral epitopes, nucleoprotein 366–374 peptide (D^bNP_{366}) and acid polymerase 224–233 peptide (D^bPA_{224}). These responses appear co-dominant in the acute response to infection, and also in the memory compartment, but during a secondary infection the D^bNP_{366} response dominates over the D^bPA_{224} response. The mechanisms for this are unclear, although a variety of suggestions have been put forward (Crowe et al., 2003; La Gruta et al., 2006). Another important difference between the two responses is that the D^bNP_{366} response tends to be highly focused (dominated by one or a few clonotypes) whereas the response to D^bPA_{224} tends to be more diverse (Turner et al., 2003; Kedzierska et al., 2004).

We have recently published a study (Kedzierska et al., 2006) comparing the central and effector memory T cell subsets in acute infection and in the memory phase of the D^bNP_{366} and D^bPA_{224} $CD8^+$ T cell responses in mice. It has been suggested by previous studies that the central and effector subsets ($CD62L^{hi}$ and $CD62L^{lo}$, respectively) of T cell memory are either quite independent (suggesting that cells arose from separate lineages (Baron et al., 2003) or identical (suggesting that a single progenitor randomly differentiated into the two different pools (Bouneaud et al., 2005)). Our study involved the sorting of single cells according to their phenotype and specificity, and the sequencing of the CDR3 β-chains of over 2000 T cell receptors. This was not an exhaustive sampling of the total TCR repertoire and there were also different numbers of

cells sampled for each subset in each animal, because of the limitations of single cell sampling (i.e.: although the same number of cells were sent for sequencing, due to some failure rate of single cell sorting and/or synthesizing cDNA from a single cell, different numbers of successful sequences were obtained). We investigated the diversity of the $CD62L^{hi}$ and $CD62L^{lo}$ subsets through counting the number of TCR clonotypes, defined by the nucleotide sequence, and determining the average number of clonotypes per mouse. The conclusion of our study was that there was a higher diversity in the central memory ($CD62L^{hi}$) subset than the effector memory ($CD62L^{lo}$) subset. This approach has limitations because it (i) does not account for the different sample sizes taken per mouse, (ii) it pools the data from all mice which results in a loss of potentially valuable information, and (iii) it does not account for clonal dominance, which may be the more relevant biological variable. In this paper, we use a variety of statistical tests to analyze these data to provide a rigorous methodology for application to other datasets.

We use the data from our previous study to demonstrate the application of the diversity measures and statistical tests for a case involving many samples of the TCR repertoire. The study involved TCR samples from the spleens of 13 and 9 mice for the D^bNP_{366} and D^bPA_{224} responses, respectively. The number of TCR sequences obtained for each of the $CD62L^{lo}$/$CD62L^{hi}$ paired samples are shown in Table 1, together with the number of different clonotypes among the TCR sequences for each sample. To demonstrate the application of the proposed approaches to addressing each of the three questions raised in Sections 3.2–3.4, we compare the diversities of samples of the $CD62L^{lo}$ and $CD62L^{hi}$ subsets between the D^bNP_{366}- and D^bPA_{224}-specific responses, compare the diversities between samples of the $CD62L^{lo}$ and $CD62L^{hi}$ TCR subsets, and investigate whether the differences in diversity between the samples of the $CD62L^{lo}$ and $CD62L^{hi}$ TCR subsets increase or decrease over time.

Table 1
Summary of the $CD62L^{hi}$ and $CD62L^{lo}$ TCR subsets of the D^bNP_{366}- and D^bPA_{224}-specific responses following primary infection

Epitope	Sample pair	Day	Number of TCR sequences		Number of different clonotypes	
			$CD62L^{hi}$	$CD62L^{lo}$	$CD62L^{hi}$	$CD62L^{lo}$
D^bNP_{366}	1	8	77	80	16	7
	2	8	57	40	10	8
	3	10	32	29	10	7
	4	11	46	58	11	5
	5	14	71	70	10	5
	6	28	56	61	14	6
	7	28	60	74	9	4
	8	28	37	48	35	6
	9	28	42	31	8	5
	10	180	43	62	11	7
	11	300	57	27	10	5
	12	450	28	12	14	7
	13	690	28	17	6	5
D^bPA_{224}	1	10	59	56	29	22
	2	10	50	40	30	16
	3	14	70	67	36	14
	4	28	38	36	19	9
	5	28	40	49	21	11
	6	28	31	36	16	12
	7	150	51	54	36	15
	8	180	78	73	33	25
	9	690	54	52	18	14

Shown are the day following primary infection that the samples were obtained, the number of TCR CDR3 β-chain sequences sampled and the number of different clonotypes (i.e.: the number of unique nucleotide sequences) obtained for each of the $CD62L^{hi}$ and $CD62L^{lo}$ subsets.

4.1. Standardizing the sample sizes

The data provided in Table 1 shows that there are substantial differences in the number of TCR sequences sampled both between the two subsets in most of the paired samples and between paired samples, which may bias the observations. For example, it appears that the $CD62L^{hi}$ subset has consistently more clonotypes than the $CD62L^{lo}$ subset for all paired samples studied in the responses to both epitopes. However, for some of the paired samples, the difference in the number of TCRs between the $CD62L^{lo}$ and $CD62L^{hi}$ subsets could easily be responsible for the observed difference in the number of clonotypes. An example of this is the second paired sample for the D^bNP_{366} epitope for which the sample size for the $CD62L^{hi}$ subset was larger than the $CD62L^{lo}$ subset by 17 TCR sequences while the difference in the number of clonotypes was only two. Hence, it is necessary to standardize the sample sizes before comparison of the subsets. Furthermore, it is difficult to assess by eye whether the observed differences are significantly different from what would be expected if the two subsets of TCR sequences had been randomly sampled from the same TCR pool.

These data include a sufficient number of sample pairs (>5) to examine the observed differences between the $CD62L^{lo}$ and $CD62L^{hi}$ subsets, and between the D^bNP_{366}- and D^bPA_{224}-specific responses, across all samples. However, the sample sizes for the $CD62L^{lo}$ subsets of D^bNP_{366} paired samples 12 and 13 are very low (<20 TCR sequences) so we choose to exclude

these two sample pairs from the analysis. This still leaves us with a sufficient number of sample pairs to perform statistical tests across all sample pairs. To enable our comparisons we first estimate the diversity for each sample by standardizing the number of TCR sequences to the minimum sample size of 27 across all $CD62L^{lo}$/$CD62L^{hi}$ sample pairs (i.e.: the sample size of the $CD62L^{lo}$ subset of D^bNP_{366} paired sample 11).

As an example of the standardization of the number of TCRs and calculation of diversity for the reduced $CD62L^{lo}$ and $CD62L^{hi}$ subsets (illustrated in Fig. 1 and discussed in Section 3.1) we focus on the first D^bNP_{366} paired sample, for which 77 and 80 TCRs were sequenced for the $CD62L^{hi}$ and $CD62L^{lo}$ subsets, respectively (Table 1). The number of TCR sequences was standardized for both samples to 27. We reduced the number of TCRs in the $CD62L^{lo}$ subset to 27 by randomly drawing, without replacement, 27 from the 80 $CD62L^{lo}$ TCR sequences. The distributions from 10,000 random draws of the diversity measures for the reduced $CD62L^{lo}$ subset are shown in Fig. 2A, C. The number of clonotypes, $N_c^{lo}=5$, and the Simpson's diversity index, $D_S^{lo}=0.50$, were estimated from the medians of the distributions. Similarly, the $CD62L^{hi}$ subset was reduced to 27 by randomly drawing, without replacement, 27 from the 77 $CD62L^{hi}$ TCR sequences. The number of clonotypes, $N_c^{hi}=9$, and the Simpson's diversity index, $D_S^{hi}=0.82$, were estimated from the medians of the distributions from 10,000 random draws (Fig. 2B, D).

The diversity measures for the samples of the $CD62L^{lo}$ and $CD62L^{hi}$ subsets of the D^bNP_{366}- and D^bPA_{224}-specific responses, where all samples have been reduced to a standard sample size of 27 TCRs, are displayed in Fig. 3A, B. With the problem of unequal sample sizes now resolved, we apply various statistical tests to these diversity measures to demonstrate the approach to several typical types of questions (discussed in Sections 3.2–3.4) that may arise in studies of TCR repertoire.

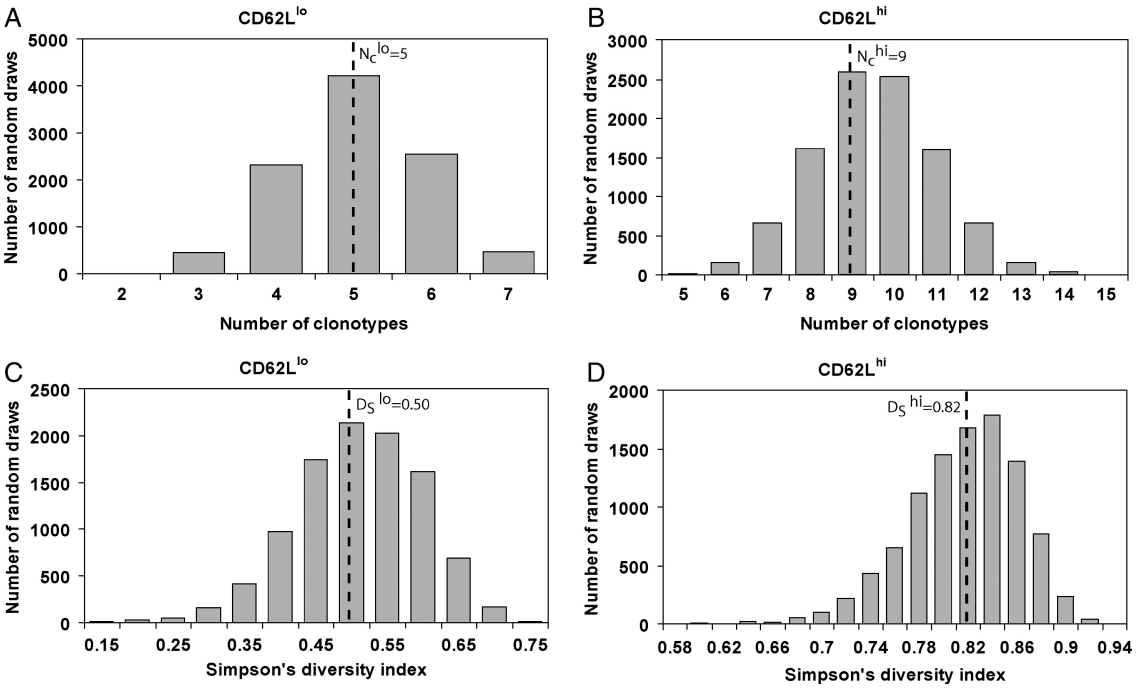

Fig. 2. Estimating the diversity of $CD62L^{lo}$ and $CD62L^{hi}$ subsets resulting from reductions in sample size. To compare the diversity of TCRs in different samples or different individuals, it is first necessary to estimate the diversity of each sample if the same number of TCRs had been obtained in all samples. Hence, the size of all samples must be reduced to the size of the smallest sample (in this case, 27 TCR sequences). The sample sizes of the $CD62L^{lo}$ and $CD62L^{hi}$ subsets for the D^bNP_{366} response in the first mouse shown in Table 1 were standardized by reducing the $CD62L^{hi}$ subset from 77 to 27 TCR sequences and reducing the $CD62L^{lo}$ subset from 80 to 27 TCR sequences. The medians of the distributions of the number of clonotypes (Panels A and B) and Simpson's diversity index (Panel C and D) resulting from 10,000 random draws of a reduced sample were used to represent the diversity measures for the $CD62L^{lo}$ subset (Panels A and C) and the $CD62L^{hi}$ subset (Panels B and D). The Simpson's diversity index has continuous values ranging between 0 and 1. In order to plot these measures we chose discrete intervals, centered on the values shown on the *x*-axis, and determined the number of random draws that resulted in an index with a value falling within each discrete interval.

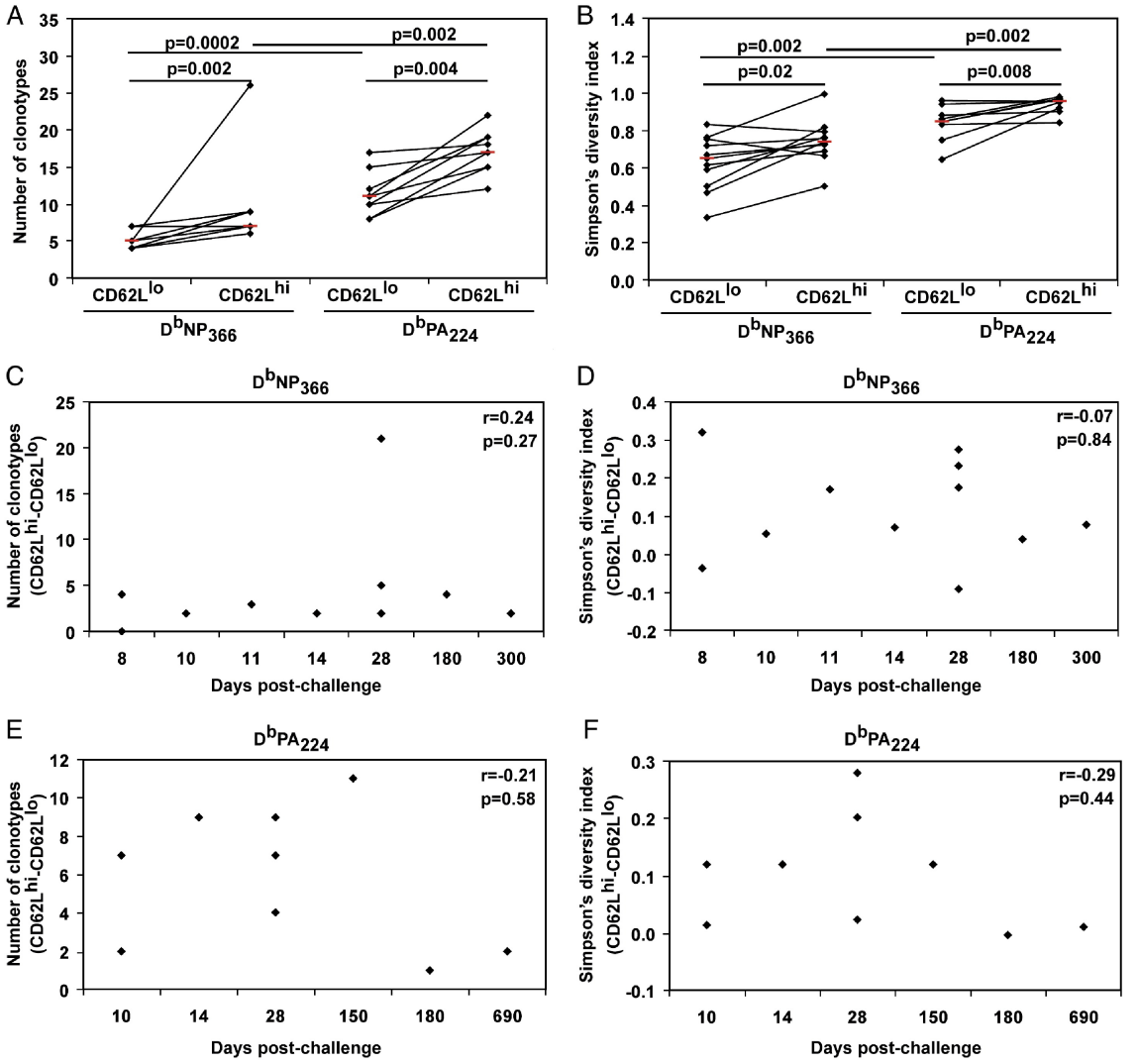

Fig. 3. Analysis of differences in diversity between the CD62L^{lo} and CD62L^{hi} subsets of the D^{b}NP$_{366}$- and D^{b}PA$_{224}$-specific TCR repertoires. The diversity measures were estimated by reducing all samples to a standard sample size of 27 TCR sequences (as described in Figs. 1 and 2). The diversity measures include the number of clonotypes (Panels A, C, E) and Simpson's diversity index (Panels B, D, F). In Panels A and B, the median diversity measures are shown as horizontal bars and the diversity measures of samples of the CD62L^{lo} and CD62L^{hi} subsets were compared using a Wilcoxon rank sum test. The diversities of the samples of the D^{b}NP$_{366}$- and D^{b}PA$_{224}$-specific responses were compared using a Mann–Whitney test. In Panels C–F, the difference in diversity measures between the CD62L^{lo} and CD62L^{hi} subsets are displayed for the times post-challenge that the paired samples were obtained. The correlations were performed using a Spearman rank correlation.

4.2. Are the samples of the CD62L^{hi} and CD62L^{lo} subsets of the T cell receptor repertoire significantly less diverse for the D^{b}NP$_{366}$- than for the D^{b}PA$_{224}$-specific responses?

Previous studies have found that the D^{b}NP$_{366}$-specific response is less diverse than the response to the D^{b}PA$_{224}$ epitope (Turner et al., 2003; Kedzierska et al., 2004). This assessment relied on the differences in diversity of Jβ usage and number of CDR3 amino acid sequences, and did not consider the CD62L expression on the cells. Thus, it is possible that differences in observed diversity between the D^{b}NP$_{366}$- and D^{b}PA$_{224}$-specific responses could arise because of differences in the proportion of cells that were CD62L^{hi} or CD62L^{lo} in the two responses. However, we can exclude this by applying our method for comparing TCR samples between two groups to comparing the diversity of samples of the CD62L^{hi} and CD62L^{lo} subsets between the D^{b}NP$_{366}$- and D^{b}PA$_{224}$-specific responses. Again, the

diversity measures used for this comparison are those determined in Section 4.1 by standardizing the sample size across all samples and shown in Fig. 3A, B. We applied a Mann–Whitney test to samples of the $CD62L^{hi}$ subset and found that those responding to the D^bNP_{366} epitope were significantly less diverse than for the D^bPA_{224}-specific response. The number of D^bNP_{366}-specific clonotypes was significantly less than the number of D^bPA_{224}-specific clonotypes (medians: 7 vs. 17, $p=0.002$) and the Simpson's diversity index was significantly less for the D^bNP_{366}- than for the D^bPA_{224}-specific repertoires (medians: 0.74 vs. 0.95, $p=0.002$). Similarly, the samples of the $CD62L^{lo}$ subset of the D^bNP_{366}-specific repertoire were significantly less diverse than for the D^bPA_{224}-specific repertoire, with significant differences in both the number of clonotypes (medians: 5 vs. 11, $p=0.0002$) and the Simpson's diversity index (medians: 0.65 vs. 0.85, $p=0.002$).

4.3. Are the samples of the $CD62L^{hi}$ subset of the T cell receptor repertoire significantly more diverse than for the $CD62L^{lo}$ subset?

To assess the significance of the differences in diversity between the $CD62L^{lo}$ and $CD62L^{hi}$ subsets, we used a Wilcoxon rank sum test on the diversity measures shown in Fig. 3A, B. The number of clonotypes was significantly higher in the $CD62L^{hi}$ subset than in the $CD62L^{lo}$ subset for both the D^bNP_{366}-specific response (median: 7 vs. 5, $p=0.002$) and the D^bPA_{224}-specific response (medians: 17 vs. 11, $p=0.004$). Using the Simpson's diversity index to account for the clonal dominance hierarchy within each response, the $CD62L^{hi}$ subset was significantly more diverse than the $CD62L^{lo}$ subset for the D^bNP_{366}-specific response (medians: 0.74 vs. 0.65, $p=0.02$) and for the D^bPA_{224} -specific response (medians: 0.95 vs. 0.85, $p=0.008$). Thus, the results of our repertoire analysis show that, regardless of the diversity measure used, the $CD62L^{hi}$ subset was significantly more diverse than the $CD62L^{lo}$ subset of $CD8^+$ T cells in the primary response to both the D^bNP_{366} and D^bPA_{224} epitopes.

4.4. Is the difference in diversity between samples of the $CD62L^{hi}$ and $CD62L^{lo}$ subsets of the T cell receptor repertoire correlated with time?

In the previous section, we demonstrated that the diversity (either number of clonotypes or Simpson's diversity index) of samples of the $CD62L^{hi}$ subset was significantly higher than the $CD62L^{lo}$ subset of both the D^bNP_{366}- and the D^bPA_{224}-specific responses. It is also of interest to know whether there is a trend for the difference in diversity between the $CD62L^{hi}$ and $CD62L^{lo}$ subsets to either increase or decrease over time. Fig. 3C–F shows the differences in the number of clonotypes and Simpson's diversity index between standardized samples of the $CD62L^{hi}$ and $CD62L^{lo}$ subsets obtained at varying times since primary challenge. There was no significant correlation between time and the difference in the number of clonotypes for either the D^bNP_{366}-specific ($r=0.24$, $p=0.47$) or the D^bPA_{224}-specific ($r=-0.21$, $p=0.58$) repertoires. Similarly, differences in Simpson's diversity between the $CD62L^{hi}$ and $CD62L^{lo}$ subsets were not correlated with time (D^bNP_{366}: $r=-0.07$, $p=0.84$; D^bPA_{224}: $r=-0.29$, $p=0.44$).

Table 2
Analysis of the difference in diversity between samples of the D^bPA_{224}-specific TCR repertoire at day 8 and days 35, 100, and 200 after secondary challenge

Time point		Mouse	Number of TCR sequences		Number of clonotypes				Simpson's diversity index			
T1	T2		T1	T2	T1	T2	T2–T1	*p*-value	T1	T2	T2–T1	*p*-value
Day 8	Day 35	10	43	34	8	10	2	0.5862	0.77	0.78	0.02	0.8423
		11	44	41	12	11	−1	0.8393	0.90	0.83	−0.06	0.1812
	Day 100	10	43	28	8	10	2	0.6511	0.77	0.86	0.09	0.0687
		11	44	46	12	16	4	0.1585	0.90	0.91	0.01	0.6011
	Day 200	10	43	35	8	9	1	0.8320	0.77	0.86	−0.09	0.0848
		11	44	36	12	9	−3	0.3467	0.90	0.83	0.07	0.1340

The number of TCR β-chain sequences obtained at each time point, for each mouse, is shown. The diversity measures included the number of clonotypes, defined by the distinct CDR3 β-chain amino acid sequences, and Simpson's diversity index. The *p*-values were calculated using randomization tests. The two-tailed *p*-value for the difference in diversity represent the probability that randomly distributing the pooled TCRs from both time points between two subsets would have resulted in differences in the number of clonotypes/Simpson's diversity index being at least as high as the differences observed between the TCR samples taken at the different time points.

5. Comparing a small number of T cell receptor repertoire samples: the randomization test

The methods described earlier in this manuscript rely on having a minimum number of sample pairs or a minimum number of samples across two groups. However, in some studies a smaller number of samples may be obtained and an appropriate test is required to enable the comparison of samples of the TCR repertoire.

The randomization test (described in (Edgington, 1995)) is a method for determining the significance of some observed test statistic that is aimed at assessing a particular hypothesis. This is achieved by generating a distribution of the test statistic assuming the null hypothesis and determining the proportion of the distribution that is at least as extreme in value as the observed test statistic. The null hypothesis for most studies will be that there is no difference in the test statistic between the two samples, or that both samples are drawn from the same distribution. Assuming our null hypothesis, we can pool the two samples, randomly distribute them into two sets of the same size as the original samples, and then calculate the test statistic between the two sets. Repeating this random distribution of TCRs into two groups of the same size as the original many times produces a distribution of the test statistic (Fig. 3). The proportion of the distribution that is at least as extreme in value as the observed test statistic is an estimate of the *p*-value and represents the probability that the observed test statistic could have been achieved by the random distribution of TCRs between two sets. The more random distributions performed, the more accurate the estimate of the *p*-value. We use 10,000 random distributions to estimate a *p*-value to 4 decimal places. If the observed test statistic is not calculated by the random distribution of TCRs between two sets then the estimated *p*-value is <0.0001.

We should note that there is no need to worry about differences in sample size with the randomization tests.

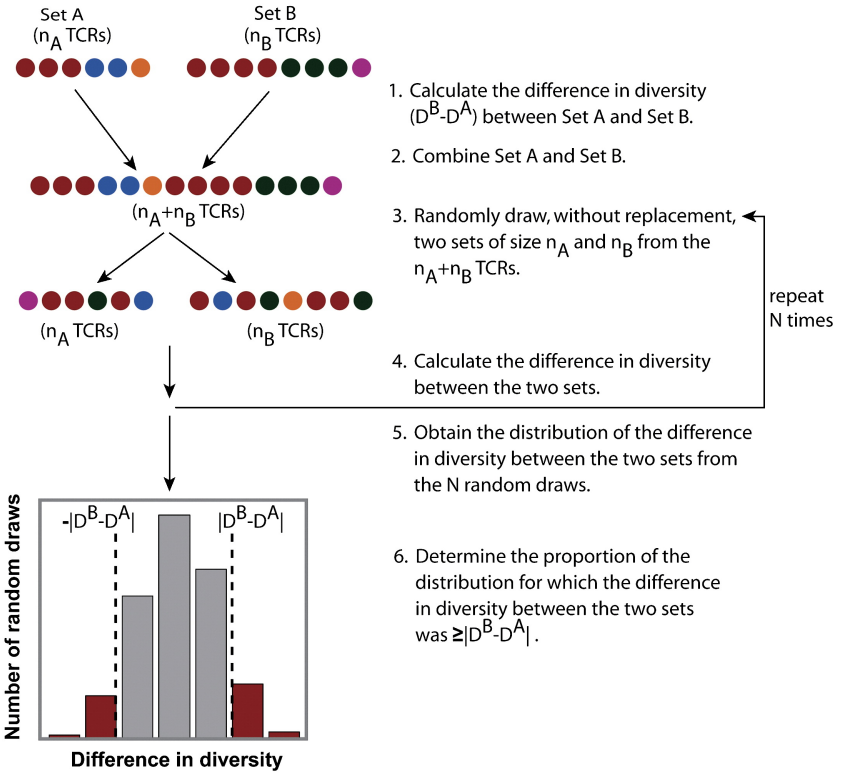

Fig. 4. The randomization test for comparing the diversity of TCR samples. A schematic of the randomization test method for determining the statistical significance of the difference in a diversity measure ($D^B - D^A$) between two TCR sets, A and B. The method involves first pooling all sequences from Set A and Set B and then randomly drawing two new sets (of the same sizes as the original Set A and Set B), and calculating the difference in diversity that arose from this random sampling. This procedure is repeated multiple times (i.e.: repeat steps 3 and 4 multiple times) to obtain a distribution of the difference in diversity measures assuming the null hypothesis that both samples are drawn from the same distribution. The *p*-value for the difference in diversity ($D^B - D^A$) is the proportion of the distribution (highlighted in red) from the random draws for which the difference in diversity was greater than that observed experimentally (i.e.: $\geq (D^B - D^A)$).

The distributions of the test statistic generated assuming the null hypothesis, with which we compare our observed test statistic, are calculated using the same sample sizes as the observed TCR sets. Hence, any biases in the test statistic due to differences in sample size are also present in these distributions.

The randomization test can be used to compare the diversity of two TCR samples and determine whether the observed difference in diversity is statistically significant, or whether it could have arisen by chance due to differences in sample size or random effects. Our test statistic for comparing the diversities of two samples of the TCR repertoire is the difference in diversity (i.e.: difference in the number of clonotypes or difference in Simpson's diversity index) between the two samples. The corresponding null hypothesis is that there is no difference in diversity between the two TCR samples. Using the randomization method described above we can determine the distribution of the difference in diversity that would arise due to random sampling effects (i.e.: the difference we would expect to observe even if there was no true difference in the diversity of the two TCR samples). The proportion of the random trials for which the difference in diversity was at least as large as the observed difference in diversity provides an estimate of the two-tailed *p*-value for the significance of the observed difference in diversity.

6. Example: Comparing the diversity in T cell memory responses to influenza in mice

As an example of a study involving only a few sample pairs, we have used data from a published study of the T cell repertoire responding to the D^bPA_{224} epitope in mice (Turner et al., 2003). This study included a longitudinal analysis of CDR3 β-chain sequences of memory $CD8^+$ T cells, involving samples of the TCR repertoire obtained from the blood of two mice at days 8, 35, 100, and 200 after secondary challenge. One aspect of the stability of the memory TCR repertoire involves determining whether the TCR repertoire becomes more or less diverse over time. We have applied the diversity measures to compare the sample of the TCR repertoire at day 8 with samples obtained at the three later time points. The number of TCR sequences obtained at each time point is shown in Table 2. Substantial differences in sample size between the time points make it difficult to directly compare samples. Furthermore, the significance of any differences across the sample pairs cannot be determined with only two mice, but for each mouse we can use randomization tests to determine whether there is a significant difference in diversity between the repertoires at different time points.

To explain the methodology of the randomization tests (illustrated in Fig. 4) we focus on the comparison of TCR samples between days 8 and 100 for mouse 11. The number of TCR sequences obtained at days 8 and 100 was 44 and 46, respectively. There were differences of four clonotypes, defined by the amino acid sequence, and 0.01 in Simpson's diversity index between the two time points (Table 2). The randomization test involved pooling the TCR sequences from both time points and then randomly drawing from the 90 pooled TCR sequences a set of 44 and a set of 46 TCRs. We performed 10,000 random draws and for each we calculated the

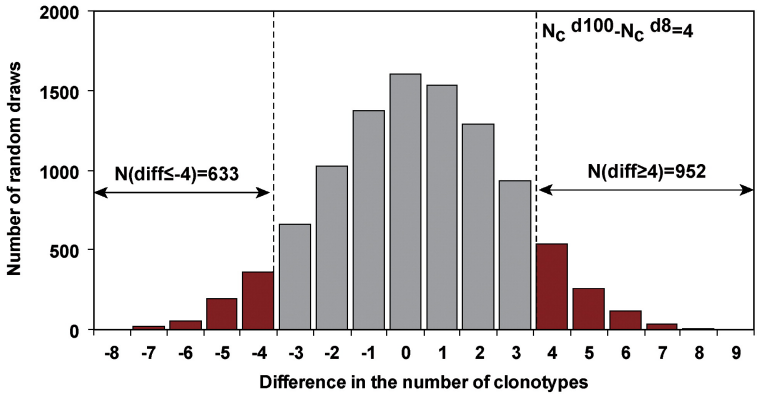

Fig. 5. Distribution of the difference in diversity between TCR sets resulting from a randomization test. The data from the secondary response to the D^bPA_{224} epitope in mouse 11 (Table 2) is used to illustrate the application of the randomization test. In this case, 44 TCRs were sequenced at day 8, including 12 different clonotypes, and 46 TCRs were sequenced at day 100, including 16 different clonotypes. Randomization tests were used to determine whether the difference of 4 in the number of clonotypes between days 8 and 100 was statistically significant. The proportion of random draws of the pooled TCR data from days 8 and 100 that resulted in a difference in the number of clonotypes of at least four (highlighted in red) provides the two-tailed *p*-value (i.e.: $p=0.1585$) for whether the difference in diversity was larger than expected by random distribution.

difference in diversity measures between the two randomly sampled TCR sets. This produced distributions of the differences in the number of clonotypes (Fig. 5) and the differences in Simpson's diversity index.

The distributions of the differences in diversity were used to determine the probability of obtaining the observed differences in diversity (4 clonotypes and 0.01 Simpson's diversity index) by random sampling of the two TCR sets from the same pool of TCR sequences. Of the 10,000 random draws, the number of draws for which the difference in the number of clonotypes for the two randomly sampled sets was at least four was 952+633=1585 (Fig. 5). Hence, the probability of a difference in the number of clonotypes of at least four was 0.1585 (Note: this is a two-tailed *p*-value in which we consider that the diversity could be higher in either set). Similarly, the probability that the difference of 0.01 in the Simpson's diversity index between the two TCR sets could have resulted from the random distribution of the TCR sequences between the two sets was 0.6011.

Our analysis in Table 2 shows that there were no significant differences in the diversity of the TCR responses to the D^bPA_{224} epitope between day 8 and days 35, 100 and 200 in either mouse.

7. Summary

We have discussed some common problems encountered in the analysis of T cell repertoire data, including biases due to inconsistent sampling efforts, the need for statistical tests of the significance of the observations, and the limitations of some commonly used measures of diversity. We have shown via example how differences in sample size can affect the estimates of diversity and thus demonstrated the need to standardize sample sizes prior to the comparison of TCR samples. When consistent sample sizes cannot be obtained by experimental means, a solution to this problem is to estimate diversity measures for TCR sets of reduced sample size. We have proposed measures for quantifying the diversity of samples of T cell repertoires that avoid some of the limitations of the popular 'counting clonotypes' measures of diversity. We have also described several options for determining the statistical significance of the observed differences in diversity measures. Together these methods should be applicable to a broad range of data, in terms of the number of samples and sample sizes, which may be encountered in studies of T cell repertoire data. Furthermore, these tests are non-parametric and thus do not require any assumptions about the distribution of the data. The implementation of these methods has been demonstrated on published data from two studies.

The methods for standardizing the sample size and for establishing the significance of the observations are not only applicable to our preferred measures of diversity and may be used to assess the significance of differences in other measures of diversity. The randomization test may also be used to assess the significance of other test statistics where a distribution of test statistics, under the null hypothesis, can be computationally generated.

The methods presented are by no means the only methods available nor are they necessarily the best methods available for every study. There is a wealth of research into ecological measures of diversity that could be utilized for specific studies of TCR repertoires (Legendre, 2003; Magurran, 2004). We have attempted to present a combination of methods that address some of the concerns we have about popular methods currently used and that are generally applicable to studies of TCR repertoires for which the sample sizes are relatively small compared with the size of the TCR repertoire. We have also chosen methods that require a relatively prescript implementation, with the hope that they can be routinely applied by researchers in their analysis of T cell repertoire data.

8. Materials and methods

8.1. Experimental data

T cell receptor sequences from C57BL/6 mice infected with influenza A were obtained from previously published studies (Turner et al., 2003; Kedzierska et al., 2006). Briefly, TCR sequences were obtained by single cell sorting of $CD62L^{hi}$ or $CD62L^{lo}$ $CD8^+$ $V\beta8.3^+$ D^bNP_{366}-$tetramer^+$ and $CD8^+$ $V\beta7.1^+$ D^bPA_{224}-$tetramer^+$ cells (Kedzierska et al., 2006), or $CD8^+$ $V\beta7.1^+$ D^bPA_{224}-$tetramer^+$ cells (Turner et al., 2003), and subsequent amplification using Vβ-specific primers. The experimental procedures are described in detail in the original publications (Turner et al., 2003; Kedzierska et al., 2006).

8.2. Computational analysis

The standardization of sample size, the randomization test, and the calculation of the diversity measures were performed using Matlab 7.0.1 (The Mathworks, Natick, MA).

The random distribution without replacement required for the standardization of the sample size and the randomization test was performed using the permutation method. (Hence, the randomization test is often referred to as a permutation test instead of randomization test.)

The permutation method involves first linearly arranging the data for the single data set for the standardization of the sample size or the two combined data sets for the randomization test. The random distribution of the data into one set (for the standardization of the sample size) or two sets (for the randomization test) is then achieved by randomly permuting the arrangement of the data (i.e.: shuffling the order of the data) and then, for the standardization of the sample size, selecting the required sample size from one end of the rearranged data or, for the randomization test, dividing the rearranged data according to the sizes of the original two samples.

The random permutation of the data was performed using the Matlab randperm function to randomly generate a new order for the data and then the data was rearranged according to this randomly generated order. To perform similar calculations using other software, random number generators can be used to randomly generate a new order for the data or some software provides a random permutation function which can directly permute the data.

8.3. Statistical analysis

All statistical analyses were performed using GraphPad Prism software (GraphPad Software Inc, San Diego, CA). The Wilcoxon rank sum test was used for pair-wise comparisons of TCR samples, the Mann–Whitney test was used to compare TCR samples within two groups, and correlations were performed using the Spearman rank correlation.

Acknowledgements

We thank Mark Tanaka and Paul Thomas for review of this manuscript. This work was supported by the James S. McDonnell Foundation 21st Century Research Award/Studying Complex Systems, the Australian National Health and Medical Research Council (NHMRC), and a Burnet Award of the NHMRC and Science, Technology, and Innovation funds from the Government of Victoria, Australia (to PCD). MPD is a Sylvia and Charles Viertel Senior Medical Research Fellow. KK is an NHMRC RD Wright Fellow and SJT is a Pfizer Senior Research Fellow.

References

Arstila, T.P., Casrouge, A., Baron, V., Even, J., Kanellopoulos, J., Kourilsky, P., 1999. A direct estimate of the human alphabeta T cell receptor diversity. Science 286, 958.

Bachmann, M.F., Speiser, D.E., Ohashi, P.S., 1997. Functional management of an antiviral cytotoxic T-cell response. J. Virol. 71, 5764.

Baron, V., Bouneaud, C., Cumano, A., Lim, A., Arstila, T.P., Kourilsky, P., Ferradini, L., Pannetier, C., 2003. The repertoires of circulating human CD8(+) central and effector memory T cell subsets are largely distinct. Immunity 18, 193.

Blattman, J.N., Sourdive, D.J., Murali-Krishna, K., Ahmed, R., Altman, J.D., 2000. Evolution of the T cell repertoire during primary, memory, and recall responses to viral infection. J. Immunol. 165, 6081.

Blattman, J.N., Antia, R., Sourdive, D.J., Wang, X., Kaech, S.M., Murali-Krishna, K., Altman, J.D., Ahmed, R., 2002. Estimating the precursor frequency of naive antigen-specific CD8 T cells. J. Exp. Med. 195, 657.

Bouneaud, C., Garcia, Z., Kourilsky, P., Pannetier, C., 2005. Lineage relationships, homeostasis, and recall capacities of central- and effector-memory CD8 T cells in vivo. J. Exp. Med. 201, 579.

Busch, D.H., Pamer, E.G., 1999. T cell affinity maturation by selective expansion during infection. J. Exp. Med. 189, 701.

Busch, D.H., Pilip, I., Pamer, E.G., 1998. Evolution of a complex T cell receptor repertoire during primary and recall bacterial infection. J. Exp. Med. 188, 61.

Butz, E.A., Bevan, M.J., 1998. Massive expansion of antigen-specific CD8+ T cells during an acute virus infection. Immunity 8, 167.

Callan, M.F., Fazou, C., Yang, H., Rostron, T., Poon, K., Hatton, C., McMichael, A.J., 2000. CD8(+) T-cell selection, function, and death in the primary immune response in vivo. J. Clin. Invest. 106, 1251.

Casrouge, A., Beaudoing, E., Dalle, S., Pannetier, C., Kanellopoulos, J., Kourilsky, P., 2000. Size estimate of the alpha beta TCR repertoire of naive mouse splenocytes. J. Immunol. 164, 5782.

Crowe, S.R., Turner, S.J., Miller, S.C., Roberts, A.D., Rappolo, R.A., Doherty, P.C., Ely, K.H., Woodland, D.L., 2003. Differential antigen presentation regulates the changing patterns of CD8+ T cell immunodominance in primary and secondary influenza virus infections. J. Exp. Med. 198, 399.

Davenport, M.P., Fazou, C., McMichael, A.J., Callan, M.F., 2002. Clonal selection, clonal senescence, and clonal succession: the evolution of the T cell response to infection with a persistent virus. J. Immunol. 168, 3309.

Davis, M.M., Bjorkman, P.J., 1988. T-cell antigen receptor genes and T-cell recognition. Nature 334, 395.

Edgington, E.S., 1995. Randomization Tests. Marcel Dekker, New York.

Hunter, P.R., Gaston, M.A., 1988. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. J. Clin. Microbiol. 26, 2465.

Janeway, C.A., Travers, P., Walport, M., Shlomchik, M.J., 2005. Immunobiology. Garland, New York.

Kedzierska, K., Turner, S.J., Doherty, P.C., 2004. Conserved T cell receptor usage in primary and recall responses to an immunodominant influenza virus nucleoprotein epitope. Proc. Natl. Acad. Sci. U. S. A. 101, 4942.

Kedzierska, K., Venturi, V., Field, K., Davenport, M.P., Turner, S.J., Doherty, P.C., 2006. Early establishment of diverse T cell receptor profiles for influenza-specific CD8+ CD62Lhi memory T cells. Proc. Natl. Acad. Sci. U. S. A. 103, 9184.

La Gruta, N.L., Kedzierska, K., Pang, K., Webby, R., Davenport, M., Chen, W., Turner, S.J., Doherty, P.C., 2006. A virus-specific CD8+ T cell immunodominance hierarchy determined by antigen dose and precursor frequencies. Proc. Natl. Acad. Sci. U. S. A. 103, 994.

Lefrancois, L., 2006. Development, trafficking, and function of memory T-cell subsets. Immunol. Rev. 211, 93.

Lefrancois, L., Marzo, A.L., 2006. The descent of memory T-cell subsets. Nat. Rev., Immunol. 6, 618.

Legendre, P.L.A.L., 2003. Numerical Ecology. Elsevier, New York.

Magurran, A.E., 2004. Measuring Biological Diversity. Blackwell.

Mahajan, V.S., Leskov, I.B., Chen, J.Z., 2005. Homeostasis of T cell diversity. Cell. Mol. Immunol. 2, 1.

Maryanski, J.L., Jongeneel, C.V., Bucher, P., Casanova, J.L., Walker, P.R., 1996. Single-cell PCR analysis of TCR repertoires selected by antigen in vivo: a high magnitude CD8 response is comprised of very few clones. Immunity 4, 47.

McHeyzer-Williams, M.G., Davis, M.M., 1995. Antigen-specific development of primary and memory T cells in vivo. Science 268, 106.

Naylor, K., Li, G., Vallejo, A.N., Lee, W.W., Koetz, K., Bryl, E., Witkowski, J., Fulbright, J., Weyand, C.M., Goronzy, J.J., 2005. The influence of age on T cell generation and TCR diversity. J. Immunol. 174, 7446.

Nei, M., 1973. Analysis of gene diversity in subdivided populations. Proc. Natl. Acad. Sci. U. S. A. 70, 3321.

Nikolich-Zugich, J., Slifka, M.K., Messaoudi, I., 2004. The many important facets of T-cell repertoire diversity. Nat. Rev. Immunol. 4, 123.

Nowak, M.A., May, R.M., 1992. Coexistence and competition in HIV infections. J. Theor. Biol. 159, 329.

Pewe, L., Heard, S.B., Bergmann, C., Dailey, M.O., Perlman, S., 1999. Selection of CTL escape mutants in mice infected with a neurotropic coronavirus: quantitative estimate of TCR diversity in the infected central nervous system. J. Immunol. 163, 6106.

Pewe, L.L., Netland, J.M., Heard, S.B., Perlman, S., 2004. Very diverse CD8 T cell clonotypic responses after virus infections. J. Immunol. 172, 3151.

Price, D.A., West, S.M., Betts, M.R., Ruff, L.E., Brenchley, J.M., Ambrozak, D.R., Edghill-Smith, Y., Kuroda, M.J., Bogdan, D., Kunstman, K., Letvin, N.L., Franchini, G., Wolinsky, S.M., Koup, R.A., Douek, D.C., 2004. T cell receptor recognition motifs govern immune escape patterns in acute SIV infection. Immunity 21, 793.

Sallusto, F., Lenig, D., Forster, R., Lipp, M., Lanzavecchia, A., 1999. Two subsets of memory T lymphocytes with distinct homing potentials and effector functions. Nature 401, 708.

Sallusto, F., Geginat, J., Lanzavecchia, A., 2004. Central memory and effector memory T cell subsets: function, generation, and maintenance. Annu. Rev. Immunol. 22, 745.

Savage, P.A., Boniface, J.J., Davis, M.M., 1999. A kinetic basis for T cell receptor repertoire selection during an immune response. Immunity 10, 485.

Sourdive, D.J., Murali-Krishna, K., Altman, J.D., Zajac, A.J., Whitmire, J.K., Pannetier, C., Kourilsky, P., Evavold, B., Sette, A., Ahmed, R., 1998. Conserved T cell receptor repertoire in primary and memory CD8 T cell responses to an acute viral infection. J. Exp. Med. 188, 71.

Turner, S.J., Diaz, G., Cross, R., Doherty, P.C., 2003. Analysis of clonotype distribution and persistence for an influenza virus-specific CD8+ T cell response. Immunity 18, 549.

Wagner, U.G., Koetz, K., Weyand, C.M., Goronzy, J.J., 1998. Perturbation of the T cell repertoire in rheumatoid arthritis. Proc. Natl. Acad. Sci. U. S. A. 95, 14447.