ELSEVIER

Journal of Immunological Methods 329 (2008) 67–80

JIM Journal of Immunological Methods

www.elsevier.com/locate/jim

Research paper

Method for assessing the similarity between subsets of the T cell receptor repertoire

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

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

Received 18 June 2007; received in revised form 27 September 2007; accepted 28 September 2007
Available online 29 October 2007

Abstract

The $CD8^+$ T cell response is important in the control of many viral and other infections. There have been many studies aimed at better understanding the influence of T cell receptor diversity on immune responses and the evolution of the T cell receptor repertoire over time and through the various stages of immune responses to infection. In recent years, there has been an increase in both the number of studies using T cell receptor data and the volume of T cell receptor data generated per study. Appropriate analytical tools are required to analyse this data. We present a robust approach to assessing the similarity between samples of the T cell receptor repertoire, which we demonstrate on published data of subsets of the influenza A virus D^bNP_{366}- and D^bPA_{224}-specific $CD8^+$ T cell responses in mice sorted on the expression of CD62L, which is a marker distinguishing central and effector memory cells.

Keywords: T cell receptor; Repertoire; Immune memory

1. Introduction

The $CD8^+$ T cell response to infection is often focused on a few immunodominant epitopes of the virus. The response to an individual viral epitope is often composed of a large number of different T cell clonotypes, identified on the basis of their different T cell receptor (TCR) sequences. This diversity of TCR sequences is produced during the maturation of the T cell in the thymus by the random recombination of the variable (V), joining (J), and diverse (D) (in the case of β-chains) TCR germline gene segments and the random deletion and addition of nucleotides in the V(D)J junction. It has been estimated that the large diversity of TCR-αβ receptors ($>10^{15}$ (Davis and Bjorkman, 1988) and $>10^{18}$ (Janeway et al., 2005) for mice and humans, respectively) produced in the thymus, gives rise to $\sim 2\times 10^6$ (Casrouge et al., 2000) and $>2\times 10^7$ (Arstila et al.,

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

* Corresponding author. Centre for Vascular Research, UNSW, Kensington 2052, Australia. 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.2007.09.016

1999; Naylor et al., 2005) different CDR3-αβs of naïve T cells in a mouse and a human, respectively, at any given time.

There have been many studies (reviewed in (Nikolich-Zugich et al., 2004) and (Davenport et al., 2007)) aimed at better understanding the role of TCR diversity in immune responses and the factors that influence the evolution of the TCR repertoire after infection or vaccination. For example, studies have compared TCR samples obtained at different time points in acute infection and memory (McHeyzer-Williams and Davis, 1995; McHeyzer-Williams et al., 1999; Turner et al., 2003; Kedzierska et al., 2004), TCR repertoires responding to different viral epitopes (Price et al., 2004), or TCR repertoires involved in primary and secondary responses (Maryanski et al., 1996; Lin and Welsh, 1998; Sourdive et al., 1998; Blattman et al., 2000; Turner et al., 2003; Kedzierska et al., 2004). Improved technology and decreasing sequencing costs have led to many more groups investigating TCR repertoire using direct sequencing of TCRs and an increase in the amount of TCR data being produced in studies. However, there is a lack of appropriate analytical methods to both quantify the characteristics of the TCR repertoire and determine whether there are any statistically significant differences or similarities between samples of the TCR repertoire.

The most common current methods for analyzing TCR repertoire involve the counting of clonotypes, and this can be extremely sensitive to sampling constraints. In most studies, the TCR repertoire is far from exhaustively sampled. It has been estimated that between 15 and 900 different TCRs respond to a given epitope (Maryanski et al., 1996; Pewe et al., 1999; Kedzierska et al., 2006a) and that the T cell population can expand >1000 fold (Butz and Bevan, 1998) in a viral response. Thus, randomly sampling 50–100 TCRs does not guarantee that all epitope-specific TCR clonotypes (i.e.: distinct nucleotide or amino acid sequences of the CDR3 of TCR α- or β-chains) will be included in a sample and there is the possibility that a larger sample would have included more distinct clonotypes. In such cases, it is important that the methods of analysis not be highly sensitive to differences in size between samples. Another important requisite of methods for comparing samples of the TCR repertoire is that they account for the clone size distribution (i.e.: the number of copies of each clonotype) or clonal dominance within an epitope-specific response. The clonal dominance hierarchy can not only bias the observations when non-quantitative approaches are used but may also play an important role in controlling viral growth or promoting the escape of the virus from T cell recognition (Price et al., 2004) (For example, a highly focused T cell response makes it easier for a virus mutant to escape recognition).

We have previously reported analytical methods for comparing the diversity of samples of the TCR repertoire (Venturi et al., 2007). These methods were demonstrated on published data from a study of primary $CD8^+$T cell responses to two viral epitopes, nucleoprotein 366–374 peptide (D^bNP_{366}) and acid polymerase 224–233 peptide (D^bPA_{224}), of the influenza A virus in C57BL/6 mice (Kedzierska et al., 2006b). The aim of this study was to better understand the relationship between subsets of the T cell population characterized by their expression of CD62L, which has been used as a marker to distinguish central memory from effector memory T cells. In this paper we present an analytical approach to assessing the similarity between samples of the TCR repertoire and we demonstrate the proposed method on this same set of data (Kedzierska et al., 2006b).

The T cell memory response to infection broadly comprises the central memory and effector memory subsets (Sallusto et al., 1999; Masopust et al., 2001; Reinhardt et al., 2001; Wherry et al., 2003), which reside in different anatomical locations and have different functions in the response to secondary infection. Central memory T cells predominantly reside in secondary lymphoid organs and, upon restimulation, have poor immediate effector function but high proliferative potential. Effector memory T cells are mostly found in non-lymphoid tissues and have rapid effector function in response to secondary infection but have low proliferative capacity. The central memory and effector memory subsets can be distinguished by their level of expression of several surface molecules (Sallusto et al., 2004). One of these surface molecules is CD62L, which is required for T cells to traverse the high endothelial venules (Tripp et al., 1995; Stein et al., 1999) during migration to the lymph nodes. Thus, central memory T cells are often classified as having high expression of CD62L ($CD62L^{hi}$), while effector memory T cells have low expression of CD62L ($CD62L^{lo}$). It has also been shown that other indicators can be used to partition memory T cells into more complex subsets (Unsoeld and Pircher, 2005).

There have been several studies (recently reviewed in (Lefrancois and Marzo, 2006) and (Lefrancois, 2006)) of the lineage relationship between the central ($CD62L^{hi}$) and effector ($CD62L^{lo}$) memory T cell subsets but it has not yet been resolved whether central and effector memory T cells arose from either distinct (Baron et al., 2003; Marzo et al., 2005) or identical (Wherry et al., 2003; Bouneaud et al., 2005) lineages. One approach to

better understanding the development of the central and effector memory T cell subsets is to study the similarities and differences between the TCR repertoires of the central and effector memory T cell subsets in immune responses.

In a recent paper, we studied the differences in diversity and the similarities between the $CD62L^{hi}$ (central memory) and $CD62L^{lo}$ (effector memory) $CD8^{+}$ T cell subsets in the primary immune responses to the D^bNP_{366} and D^bPA_{224} epitopes of the influenza A virus in mice (Kedzierska et al., 2006b). The data collected for this study (Kedzierska et al., 2006b) is summarized in Table 1. We assessed the similarity between the $CD62L^{hi}$ and $CD62L^{lo}$ subsets of the TCR repertoire by determining the percentage of clonotypes, defined by the CDR3 β-chain nucleotide sequence, in each subset that was common to both subsets. At least one common clonotype was found in all paired $CD62L^{hi}/CD62L^{lo}$ samples (Table 1) and, as a result of the $CD62L^{hi}$ subset being more diverse than the $CD62L^{lo}$ subset, the percentage of common clonotypes was found to be consistently higher in the $CD62L^{lo}$ subset than in the $CD62L^{hi}$ subset in the responses to both the D^bNP_{366} and D^bPA_{224} epitopes. This analysis had the limitations that it accounted for neither the differences in sample size between the two subsets nor the clonal dominance (i.e.: a large number of copies were sampled for some clonotypes) within each of the subsets. We overcome these limitations of the previous study in our reanalysis of the data using a novel approach to determining the similarity between the two subsets. Furthermore, this analytical approach allows the application of tests for determining the statistical significance of the similarity across samples of the TCR repertoire. We apply this methodology to address some questions of interest about the relationship between the $CD62L^{hi}$ and $CD62L^{lo}$ TCR subsets, including (i) Do the $CD62L^{hi}$ and $CD62L^{lo}$ subsets become more or less similar over time? and (ii) Are the $CD62L^{hi}$ and $CD62L^{lo}$ subsets more or less similar in the D^bNP_{366}-specific response than in the D^bPA_{224}-specific response? The latter question in particular is difficult to address because the D^bNP_{366}-specific response is typically dominated by one or a few TCR amino acid sequences (i.e.: there is a strong clonal dominance hierarchy) whereas the D^bPA_{224}-specific response is typically more diverse (Turner et al., 2003; Kedzierska et al., 2004). Thus, differences in the overlap between the $CD62L^{hi}$ and $CD62L^{lo}$ subsets may arise as a result of differences in the clonal dominance hierarchy. We also propose an

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 influenza virus infection

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

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

analytical approach to address the issue of how likely it is that the $CD62L^{hi}$ and $CD62L^{lo}$ subsets randomly differentiated from a common precursor.

In this paper we present some novel methods for assessing the similarity between samples of the TCR repertoire that overcome some of the problems associated with current methods of analysis, and we demonstrate the application of these methods on recently published data.

2. Materials and methods

2.1. Experimental data

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

2.2. Computational analysis

The standardization of sample size, the randomization test, and the calculation of the similarity 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 generation of the reference similarity distribution for two TCR sets randomly sampled from the pooled TCR populations was performed using the permutation method. For the standardization of the sample size, the permutation method involved first linearly arranging the data. The random selection of a reduced subset was achieved by randomly permuting the arrangement of the data (i.e.: shuffling the order of the data) and then selecting the required sample size from one end of the rearranged data. For the generation of the reference similarity distributions, the two data sets being compared were first combined and then the data was linearly arranged. The random distribution of the data into the two subsets was achieved by randomly permuting the arrangement of the two combined data sets and then selecting, from one end of the rearranged data, two sets of the required sample size.

The random permutation of the data was performed using the Matlab randperm function to randomly generate a new order for the data. The data was then rearranged according to this randomly generated order. Similar calculations can be performed using the random number generators of other software.

2.3. 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.

2.4. Measures of similarity: the Morisita–Horn similarity index

The similarity, or overlap, between samples of the TCR repertoire is often assessed by calculating either the proportion of the number of clonotypes in each sample that are common to both samples or the proportion of the total number of clonotypes in the two samples that are common to both samples. Such measures of similarity are sensitive to the sample sizes and disregard the distribution of clone sizes. Similarity measures that do not account for clonal dominance within an epitope-specific response can be misleading because they equally weight the presence of dominant and subdominant clonotypes. Thus, for example, if only a single copy of a TCR is seen in one sample, it is possible that it may not be seen in the other sample. However, if 20 copies of a sequence were present in one sample, the absence of the sequence in the other sample is likely to be highly significant. We illustrate the importance of accounting for clonal dominance using the theoretical example in Table 2A. Furthermore, the clonal dominance hierarchy may play an important role in an immune response and it is therefore advantageous to use methods of analysis that use information about the clonal dominance hierarchy within TCR samples.

In order to assess the overlap in species composition between populations, the field of ecology has developed various measures of similarity (reviewed in (Magurran, 2004)) that account for both the number of different species and the number of individuals of each species. We have chosen to use the Morisita–Horn similarity index, which has been used previously to compare TCR repertoire data (Hsieh et al., 2004; Hsieh et al., 2006). The Morisita–Horn similarity index accounts for both the number of common clonotypes and the distribution of clone sizes, and it is most sensitive to the clone sizes

of the dominant clonotypes. This is illustrated in Example A in Table 2A, where a higher similarity measure is achieved for dominant clonotypes common between two samples than for common subdominant clonotypes. This is appropriate for TCR repertoire studies because it is less likely that subdominant clonotypes would be found in multiple samples when only a small portion of the total repertoire may be represented in any one sample. The Morisita–Horn similarity index is given by:

$$C_{MH} = \frac{2\sum_{i=1}^{c} f_i g_i}{\sum_{i=1}^{c}(f_i^2 + g_i^2)},$$

where $f_i = n_{1i} / N_1$ and $g_i = n_{2i} / N_2$, n_{1i} and n_{2i} are the clone sizes of the ith clonotype (i.e.: number of copies of each distinct CDR3 sequence for the TCR α- or β-chain) in samples 1 and 2, and N_1 and N_2 are the total number of TCRs in samples 1 and 2, respectively. The summations in the numerator and the denominator are over all c clonotypes in both samples. This index ranges between 0 and 1, with 0 and 1 representing minimal and maximal similarity, respectively.

Table 2
Theoretical examples that illustrate the limitations of some methods of analysis for comparing samples of the TCR repertoire

A
Example A

	Number of TCR sequences		
Clonotype	Sample P	Sample Q	Sample R
Clonotype 1	90	90	0
Clonotype 2	5	5	90
Clonotype 3	2	2	5
Clonotype 4	2	0	2
Clonotype 5	1	0	0
Clonotype 6	0	2	0
Clonotype 7	0	1	2
Clonotype 8	0	0	1
Total	100	100	100

B
Example B

	Number of TCR sequences		
Clonotype	Sample X	Sample Y	Sample Z
Clonotype 1	30	10	60
Clonotype 2	15	10	25
Clonotype 3	5	10	10
Clonotype 4	0	10	1
Clonotype 5	0	10	1
Clonotype 6	0	0	1
Clonotype 7	0	0	1
Clonotype 8	0	0	1
Total	50	50	100

C
Analysis for example A

Comparison of samples	% Common clonotypes	Morisita–Horn similarity index
P vs. Q	42.9%	1.00
Q vs. R	42.9%	0.06
P vs. R	42.9%	0.06

D
Analysis for example B

		With sample size correction	
Comparison of samples	% Common clonotypes	% Common clonotypes	Morisita–Horn similarity index
X vs. Y	60%	60%	0.61
Y vs. Z	62.5%	62.5%	0.61
X vs. Z	37.5%	60%	0.99

Notes to Table 2
Example A in panels A and C illustrates the importance of accounting for clonal dominance within an epitope-specific response in determining the similarity between TCR samples. Comparison of each pair of samples (panel C) shows that 3 of the 7 clonotypes present in each sample pair are common to both samples. Thus, measures of similarity that are based on percentages of clonotypes common between two samples would conclude that samples P and Q, Q and R, and P and R are equally similar, even though the dominant clonotype (shaded) in samples P and Q is not dominant in sample R. It is clear that if clonal dominance was taken into account then samples P and Q are more similar than samples P and R or samples Q and R, as demonstrated by the Morisita–Horn similarity index. Example B in panels B and D shows three samples where not all samples are of the same size (i.e.: they contain different numbers of TCR sequences). It appears that, based on the percentage of the number of clonotypes in the two samples that are common to both samples, sample X is more similar to sample Y than to sample Z. However, if we correct for the differences in sample size using the method described in Fig. 1, we find that the percentage of the number of clonotypes in the two samples that are common to both samples is comparable between all three samples. Furthermore, if we calculate the Morisita–Horn similarity index and account for both the differences in sample size and the clone size distributions, we find that samples X and Z are the most similar of all pairs of samples. The high similarity between samples X and Z is due to the similar clonal dominance hierarchy of the 3 common clonotypes and the subdominance of the clonotypes present in sample Z but not in sample X. If 50 TCRs had been sequenced for sample Z, instead of 100 TCRs, it is likely that not all of these subdominant TCRs would have been found. This theoretical example illustrates that differences in sample size make it difficult to assess the similarity between samples.

2.5. Comparing many samples of the T cell receptor repertoire

2.5.1. Correcting for unequal sample sizes

While most studies aim to obtain an equal number of TCR sequences in all samples, this is not always possible due to experimental limitations. This is particularly

the case with single cell PCR. However, differences in sample size can bias observed differences or similarities between samples. When using methods of analysis that are highly sensitive to the sample size, it can be particularly difficult to determine whether the observed differences or similarities could be due to differences in the size of the samples. We provide a theoretical example in Table 2B, Example B, to illustrate some of the problems associated with different sample sizes.

We recently applied a method for standardizing the sample sizes prior to comparing the diversity of TCR samples (Venturi et al., 2007) and here we apply this method to samples of different size to estimate the similarity measures had the samples been of the same size. This method, illustrated in Fig. 1A, involves reducing larger samples to a standard sample size by randomly drawing, without replacement, a subset of the smaller standard sample size. With each random draw the similarity measure is calculated between two samples of the same size. A distribution of the estimated similarity is obtained by performing this random draw multiple times and estimating the similarity with each random draw. The median of the distribution of similarity resulting from multiple random draws provides an estimate of the similarity between the two samples, had the two samples been of the same size.

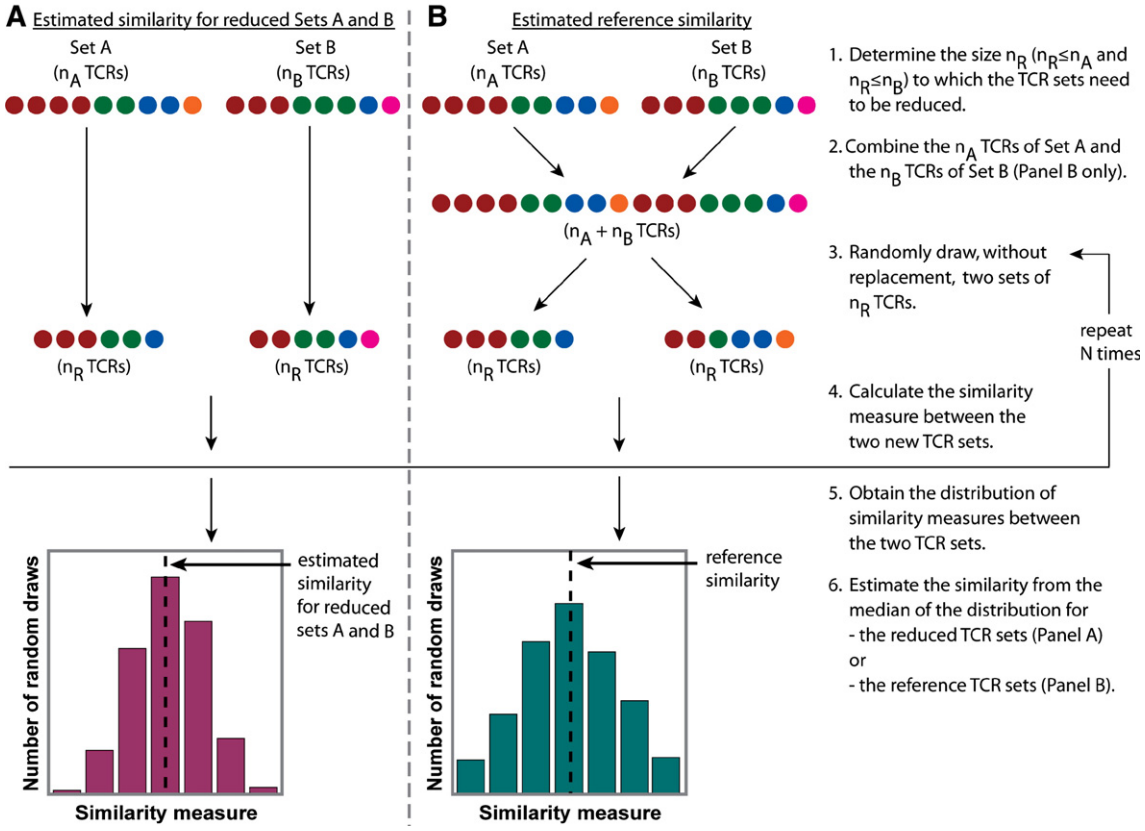

Fig. 1. Schematic of the methods for estimating similarity measures. In this case, n_A TCRs and n_B TCRs were observed experimentally in sets A and B, respectively. Panel A shows the method for estimating what the similarity between the two sets would have been if we had obtained a smaller number of TCRs (n_R), so that we can compare the similarities between multiple paired TCR sets. This involves randomly drawing a sample of size n_R from the n_A TCRs of set A and a sample of size n_R from the n_B TCRs of set B, and calculating the similarity measure for these two smaller sample sets. Drawing the reduced number of TCRs from both sets many times, and estimating the similarity each time (i.e.: repeat steps 3 and 4 multiple times), we obtain a distribution of the similarity measure for the reduced TCR samples. The median of this distribution provides an estimate of the similarity had the two observed TCR samples been of size n_R. Panel B shows the steps involved in determining the reference similarity from the distribution generated under the null hypothesis that sets A and B were randomly assigned from the same TCR population. To determine the reference similarity between two sets of size n_R, we first pool the n_A TCRs of set A and the n_B TCRs of set B. Two TCR sets of size n_R are then randomly drawn from the pooled populations of sets A and B, and the similarity measure calculated for these two randomly sampled sets. Repeating the procedure many times (i.e.: repeat steps 3 and 4 multiple times), we obtain a reference distribution of the similarity measure. The median of this distribution provides a reference similarity for two TCR sets of size n_R randomly sampled under the null hypothesis that the two observed TCR sets were randomly assigned from the same TCR population.

For example, for the three theoretical samples in Example B in Table 2B, we used this method to estimate the similarity between any two of the three samples as if only 50 TCRs (i.e.: the size of samples X and Y) had also been sequenced for sample Z. This involved randomly drawing 50 TCRs from the 100 TCRs of sample Z. The similarity was calculated between the 50 TCRs of sample X and the 50 TCRs randomly drawn from sample Z and between the 50 TCRs of sample Y and the 50 TCRs randomly drawn from sample Z. The medians of the distributions generated from many random draws provide estimates of the similarity between sample X and Z and sample Y and Z. The similarity between samples X and Y was calculated directly using the 50 TCRs in each sample, without any need for corrections for differences in sample size. The results of the analysis both with and without the corrections for differences in sample sizes are shown on Table 2D, and demonstrate that the proposed method eliminates a number of biases that can arise as a result of unequal sample sizes. The application of this method is also demonstrated in Section 3.1. for the comparison of the $CD62L^{hi}$ and $CD62L^{lo}$ subsets of the D^bNP_{366}- and D^bPA_{224}-specific $CD8^+$ T cell responses.

Using the above method to estimate similarity measures between samples, where differences in sample size have been accounted for, we can apply various statistical tests to address a range of questions about the overlap between samples of the TCR repertoire.

2.5.2. Are paired samples of the T cell receptor repertoire in one group more similar than in another group?

Assessments of similarity, by their very nature, involve comparisons between paired samples. However, it is often of interest in immunological studies to compare the similarity between sample pairs in one group with the similarity between sample pairs in another group and such comparisons can be useful in addressing important immunological questions. For example: are virus-specific repertoires at two successive post-infection time points more or less similar in vaccinated animals than in unvaccinated animals? This example would involve assessing the similarity between paired samples at the two post-infection time points for each animal and then comparing these similarities between the vaccinated and unvaccinated groups of animals. Another example, presented in Section 3.2., asks the question: Is the $CD62L^{hi}$ subset more similar to the $CD62L^{lo}$ subset in the influenza D^bNP_{366}-specific responses than in the D^bPA_{224}-specific responses in mice? For this example, the similarity is assessed between paired samples of the $CD62L^{hi}$ and $CD62L^{lo}$ subsets for each epitope-specific response within each animal and then the similarities compared between the D^bNP_{366}- and D^bPA_{224}-specific responses.

In order to compare the similarities of paired samples of the TCR repertoire between groups, all samples must either be of equal size or the similarities between paired samples must be estimated as if all samples had had the same number of TCRs as obtained for the smallest sample. This latter option is performed using the method described in Section 2.5.1. The measures of similarity between paired samples can then be compared between the two groups using either a Mann–Whitney test (if the sample pairs are not paired and thus the similarity measures are not paired between groups) or a Wilcoxon rank sum test (if the sample pairs are paired and thus the similarity measures are paired between groups) to determine whether the similarity values in one group are significantly higher than in the other group. A minimum of four similarity measures in each group is required to obtain a two-tailed p-value less than 0.05 with a Mann–Whitney test and a minimum of six similarity measures paired between the two groups is required to determine a two-tailed p-value less than 0.05 with a Wilcoxon test.

2.5.3. Does the similarity between paired samples of the T cell receptor repertoire increase or decrease with the change in some experimental variable?

The effect of some other experimental variable on the similarity between paired samples of the TCR repertoire may also be of interest in some studies. For example, in Section 3.3. we apply the proposed methods to look at whether the $CD62L^{hi}$ and $CD62L^{lo}$ subsets in each of the influenza D^bNP_{366}- and D^bPA_{224}-specific responses in mice become more or less similar with time since infection. If sample sizes are all equal, we can simply correlate the similarity of the paired samples with the experimental variable (for example, time) using a standard test such as a Spearman rank correlation. However, if sample sizes are unequal we first need to estimate the similarities between paired samples as if all samples had had the same number of TCRs as obtained for the smallest sample (described in the Section 2.5.1.). Once the differences in sample sizes have been accounted for, the correlation between the estimates of similarity and the experimental variable can be tested using a Spearman rank sum test.

2.5.4. Are paired samples of the T cell receptor repertoire significantly less similar than expected?

A common question that may arise when comparing paired samples of the TCR repertoire is whether the observed level of similarity between different TCR sets

is consistent with these sets having been randomly assigned from the same TCR population, and whether any differences between the two TCR sets can be explained by random sampling. This is a relevant question in understanding the lineage relationship of two TCR samples. Possible relationships might be (i) the two populations are derived from a common precursor, and cells derived from the different clonotypes randomly differentiated into the two compartments, independent of their TCR, (ii) the two populations were derived from a common precursor but the differentiation process was TCR-specific, depending on factors such as TCR affinity or level of stimulation (so the distribution of clonotypes into the two compartments is non-random with respect to TCR), or (iii) the two populations arose from different precursors. It is important to note that even if the TCRs were randomly sorted into two populations (relationship (i) above), we do not expect the TCR samples to be identical because of the effects of random sampling. Thus, if the two samples are identical our task is easy, but if the samples are different, we need to know whether these differences could be explained by random sampling. The corresponding statistical question is 'Under the null hypothesis that the two observed TCR sets were randomly assigned from the same TCR population, is the similarity between the two samples lower than expected by chance?' As an example of this type of question, in Section 3.4., we determine the likelihood that the $CD62L^{hi}$ and $CD62L^{lo}$ TCR subsets, in the influenza D^bNP_{366}- and D^bPA_{224}-specific primary responses in mice, were randomly assigned from the same $CD8^+$ T cell population.

One approach to assessing the statistical significance of the similarity between two TCR sets is to use a randomisation test (described in (Edgington, 1995)) on each pair of TCR samples. The randomisation test involves first generating a reference distribution of similarity measures under the null hypothesis that the two TCR sets were randomly assigned from the same TCR population. This involves pooling the two TCR samples, randomly distributing them into two sets of the same size as the original samples, and then calculating the similarity between the two sets. Repeating this procedure many times produces the reference distribution of the similarity between two TCR sets, assuming the null hypothesis that the two TCR sets were randomly assigned from the same TCR population. The proportion of the distribution that has similarity measures less than or equal to the similarity between the two observed TCR sets is an estimate of the *p*-value and represents the probability that a similarity less than or equal to the observed similarity could have been achieved by random sampling from the same TCR population. This approach results in a *p*-value for each pair of TCR samples. However, when there are several pairs of samples, how do we interpret the results if the similarities between the observed TCR paired samples are significantly less than predicted by the reference distribution for only some of the samples? What if the similarities for all the observed TCR paired samples are consistently low by comparison with the reference distribution, but this is not significant for the individual pairs? If there are many sample pairs, the interpretation of the results may not be straight forward. Thus, we have developed another approach for assessing the significance of the similarity between paired TCR sets across many sample pairs.

For studies involving many sample pairs, the observed similarity for each TCR paired sample can be compared with the corresponding reference similarity distribution expected if the two samples were pooled and then randomly redistributed into two sets. Under the null hypothesis that the observed paired TCR samples were randomly assigned from the same TCR population, there is $\sim$50% chance that the similarity between the two observed TCR sets will be larger (and $\sim$50% chance that it will be smaller) than the median of the reference similarity distribution (henceforth 'reference similarity'). Thus, we can apply a standard statistical test for paired values to compare the similarities between all the observed paired TCR samples with the corresponding reference similarities. If the similarities between the observed TCR sample pairs are consistently lower than the reference similarities across all TCR sample pairs, then we may be able to reject the null hypothesis (even if this difference is not significant for any single pair).

Based on the reasoning presented above, we propose that the median of the reference similarity distribution (or 'reference similarity') generated under the null hypothesis that the two TCR sets were randomly assigned from the same TCR population, be used as a 'standard' against which to compare observed similarity measures. That is, if the two observed TCR samples were randomly assigned from the same population, then pooling them and randomly redistributing them back into two sets should give us a similarity measure comparable to that observed. This approach also requires that, if there are differences in sample size between the paired TCR samples, we first account for these differences by estimating the similarity between samples of a standard sample size. The similarities between the observed TCR sample pairs can be estimated using the approach described in Section 2.5.1. The reference similarity can be estimated for samples of a standard sample size by drawing two samples of a standard sample size from the

pooled TCR populations of the two observed TCR sets (Fig. 1B), instead of drawing two samples of the same sizes as the observed TCR samples, as done in the randomisation test. A Wilcoxon rank sum test can then be applied across all pairs of similarity measurements to compare the similarities for the observed paired TCR samples with the corresponding reference similarities. Thus, it can be determined whether the similarities for the observed TCR sample pairs are lower than expected by chance, if the sample pairs had been randomly assigned from the same TCR population. The Wilcoxon test requires a minimum of six pairs of samples to determine a two-tailed p-value less than 0.05.

3. Results and discussion

We have applied the analytical methods for comparing samples of the TCR repertoire, presented in Sections

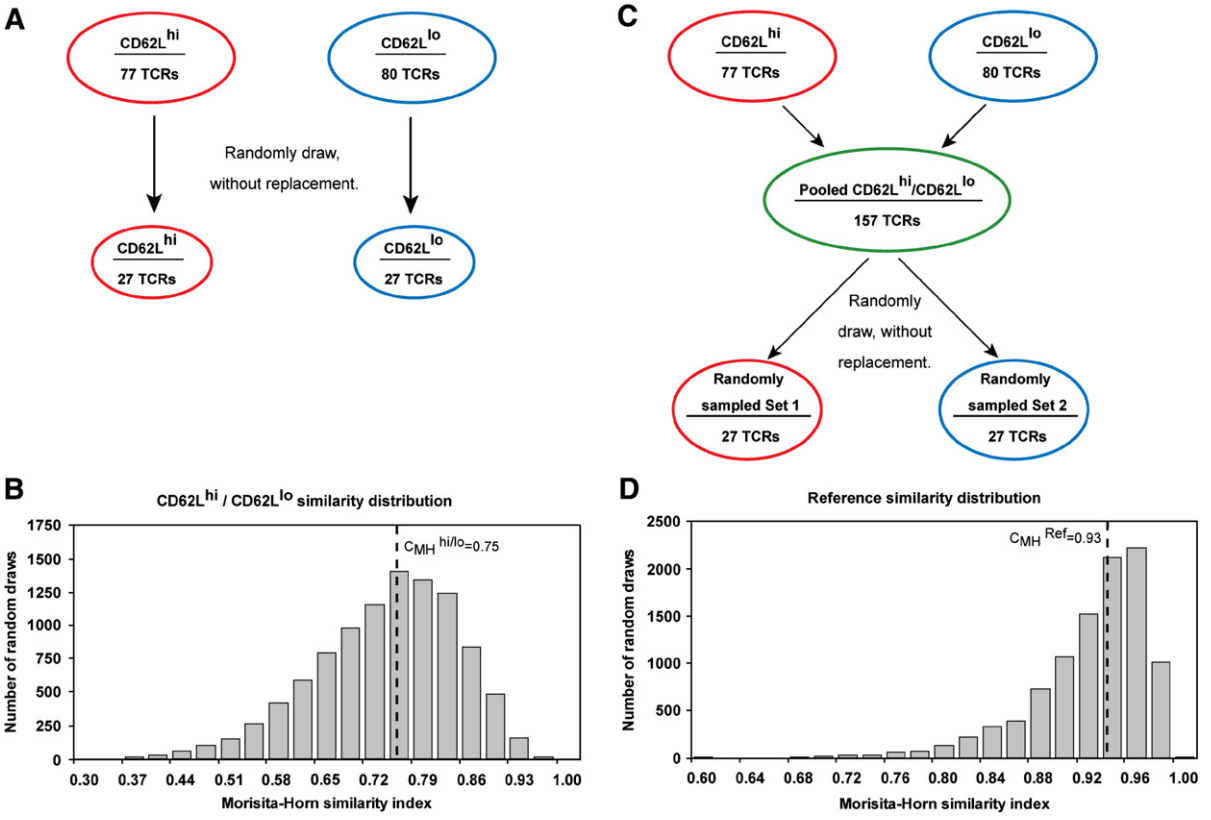

Fig. 2. Estimating the similarity between the CD62L^{hi} and CD62L^{lo} TCR subsets and the corresponding reference similarity determined under the null hypothesis that the CD62L^{hi} and CD62L^{lo} subsets were randomly assigned from the same TCR population. Comparisons involving the similarity of paired CD62L^{hi}/CD62L^{lo} TCR samples require that the similarity of CD62L^{hi}/CD62L^{lo} sample pairs be estimated as if a standard 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^{hi} and CD62L^{lo} subsets for the D^{b}NP$_{366}$-specific 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 (Panel A). The median of the distribution of the Morisita–Horn similarity index resulting from 10,000 random draws of two reduced samples was used to represent the Morisita–Horn similarity index between the CD62L^{hi} and CD62L^{lo} subsets, $C_{MH}^{hi/lo}$ (Panel B). Under the assumption that the CD62L^{hi} and CD62L^{lo} subsets were randomly assigned from the same TCR population, a reference similarity was determined by randomly drawing two sets of the standard sample size of 27 TCR sequences from the pooled 77 TCRs from the CD62L^{hi} subset and 80 TCRs from the CD62L^{lo} subset (Panel C). The median of the distribution of the Morisita–Horn similarity index resulting from 10,000 random draws of two TCR sets from the pooled CD62L^{hi} and CD62L^{lo} subsets provides a reference Morisita–Horn similarity index, C_{MH}^{Ref}, representing the similarity for two TCR sets randomly sampled under the assumption that the CD62L^{hi} and CD62L^{lo} subsets were randomly assigned from the same TCR population (Panel D). The Morisita–Horn similarity 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 horizontal axis, and determined the number of random draws that resulted in an index with a value falling within each discrete interval. The similarity derived from sampling of the experimentally observed CD62L^{hi} and CD62L^{lo} subsets ($C_{MH}^{hi/lo}$=0.75, panels A and B) is lower than the reference similarity obtained from sampling the pooled CD62L^{hi} and CD62L^{lo} subsets (C_{MH}^{Ref}=0.93, panels C and D), suggesting that the experimentally observed CD62L^{hi} and CD62L^{lo} subsets were less similar than could be explained by random sampling from the same population.

2.4. and 2.5., to assessing the similarity for samples of the CD62L^{hi} and CD62L^{lo} subsets of the TCR in immune responses to influenza in mice. The data for this study is summarized in Table 1.

3.1. Determining similarity measures for samples of standard size

The summary of data in Table 1 shows that there were large variations in the sample sizes of the CD62L^{hi} and CD62L^{lo} subsets both within and between sample pairs. Thus, to facilitate the similarity analysis across all sample pairs it was necessary to first estimate the similarities between the CD62L^{hi} and CD62L^{lo} subsets for each sample pair for a standardized number of TCRs. We excluded paired samples 12 and 13 for D^{b}NP$_{366}$-specific responses from the analysis because of the low number (<20) of TCR sequences in the CD62L^{lo} subsets of these paired samples. The standard sample size, for which all CD62L^{hi} / CD62L^{lo} similarities were estimated, was the minimum sample size of 27 (i.e.: the sample size of the CD62L^{lo} subset of D^{b}NP$_{366}$ paired sample 11) from the remaining 11 and 9 paired samples for the D^{b}NP$_{366}$- and D^{b}PA$_{224}$-specific responses, respectively.

As an example of the standardization of the number of TCRs and calculation of the similarity between reduced CD62L^{lo} and CD62L^{hi} subsets we focus on the first D^{b}NP$_{366}$ 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^{hi} and CD62L^{lo} subsets to 27 by randomly drawing, without replacement, 27 from the 77 CD62L^{hi} and 27 from the 80 CD62L^{lo} TCR sequences (Fig. 2A). With each pair of random draws, the Morisita–Horn similarity index for the two reduced TCR subsets was calculated. The distribution, from 10,000 random draws, of the Morisita–Horn similarity index between the reduced CD62L^{hi} and CD62L^{lo} subsets is shown in Fig. 2B. The representative Morisita–Horn similarity index, $C_{MH}^{hi/lo}=0.75$, was estimated from the median of the distribution.

The Morisita–Horn similarity indices between the paired CD62L^{hi} and CD62L^{lo} subsets of the D^{b}NP$_{366}$- and D^{b}PA$_{224}$-specific responses, where all samples have been reduced to a standard sample size of 27 TCRs, are displayed in Fig. 3. In the following sections we apply various statistical tests to these estimated similarity measures to address several questions about the relationship between the CD62L^{hi} and CD62L^{lo} TCR subsets. These tests demonstrate the approach required to address several typical types of questions that may arise in TCR repertoire studies.

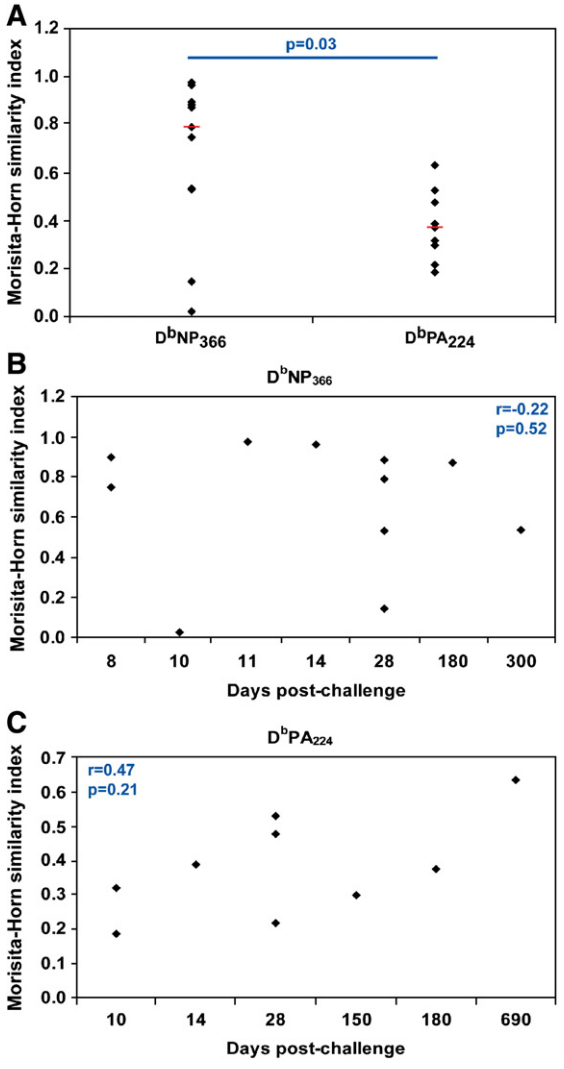

Fig. 3. Analysis of the similarity between the CD62L^{hi} and CD62L^{lo} subsets of the D^{b}NP$_{366}$-and D^{b}PA$_{224}$-specific TCR repertoires. The similarity between TCR subsets is quantified using the Morisita–Horn similarity index, which varies between 0 (minimal similarity) and 1 (maximal similarity). The Morisita–Horn similarity indices were estimated for a standard sample size of 27 TCR sequences (as described in Figs. 1 and 2). Panel A shows the Morisita–Horn similarity indices for the paired samples of the CD62L^{hi} and CD62L^{lo} subsets for the D^{b}NP$_{366}$-and D^{b}PA$_{224}$-specific responses. The CD62L^{hi} / CD62L^{lo} similarities of the samples from the D^{b}NP$_{366}$-specific TCR repertoires were significantly higher than for the D^{b}PA$_{224}$-specific responses. In Panels B and C, the Morisita–Horn similarity indices for the CD62L^{hi} and CD62L^{lo} subsets are displayed for the times post-challenge that the paired samples were obtained. The correlations were performed using a Spearman rank correlation and demonstrate no significant trend in similarity over time.

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

The immune responses to the D^{b}NP$_{366}$ and D^{b}PA$_{224}$ epitopes have different clonal dominance hierarchies, with the D^{b}PA$_{224}$-specific responses being clonotypically more diverse than the D^{b}NP$_{366}$-specific responses, which are mostly dominated by a few TCR clonotypes. It is therefore of interest to know whether the overlap between the CD62L^{hi} and CD62L^{lo} CD8$^+$ T cell subsets differs between the responses to the D^{b}NP$_{366}$ and D^{b}PA$_{224}$ epitopes.

In this comparison we used the Morisita–Horn similarity indices for the paired CD62L^{hi} and CD62L^{lo} subsets, shown in Fig. 3A, that were estimated for a standard sample size of 27 TCRs across all samples. We found the similarity between the CD62L^{hi} and CD62L^{lo} subsets, as quantified by the Morisita–Horn similarity index, was significantly higher for the D^{b}NP$_{366}$-than for the D^{b}PA$_{224}$-specific responses (medians: 0.79 vs. 0.37, $p=0.03$, Mann–Whitney). However, this result raises the question of whether the strong clonal dominance hierarchy within the responses to the D^{b}NP$_{366}$ epitope plays a large role in the CD62L^{hi} and CD62L^{lo} subsets being more similar in the D^{b}NP$_{366}$-specific than in the D^{b}PA$_{224}$-specific responses. This question will be addressed in Section 3.5.

3.3. Is the similarity between the samples of the CD62L^{hi} and CD62L^{lo} subsets of the T cell receptor repertoire correlated with time?

The summary of data in Table 1 shows that paired samples of the CD62L^{hi} and CD62L^{lo} subsets of the D^{b}NP$_{366}$- and D^{b}PA$_{224}$-specific responses were obtained at various time points ranging from day 8 to day 690 post-infection. The similarity estimates in Fig. 3A show that there is a large variation in the similarity measures for paired samples of the CD62L^{hi} and CD62L^{lo} subsets within each of the D^{b}NP$_{366}$- and D^{b}PA$_{224}$-specific responses. This raises the interesting question of whether or not the similarity between the CD62L^{hi} and CD62L^{lo} subsets alters over time, such that there is a trend for the similarity between the two subsets to increase or decrease with time. Fig. 3 B-C show the Morisita–Horn similarity indices, estimated for paired samples of the CD62L^{hi} and CD62L^{lo} subsets with a standard sample size across all samples, obtained at various times following primary challenge. There was no significant correlation between time since infection and the similarity between the CD62L^{hi} and CD62L^{lo} subsets for either the D^{b}NP$_{366}$-specific ($r=-0.22$, $p=0.52$, Spearman) or the D^{b}PA$_{224}$-specific ($r=0.47$, $p=0.21$) repertoires.

3.4. Are the samples of the CD62L^{hi} and CD62L^{lo} subsets of the T cell receptor repertoire significantly less similar than expected?

To better understand the lineage relationship between the CD62L^{hi} and CD62L^{lo} compartments of CD8$^+$ T cell memory, we assessed the likelihood that the CD62L^{hi} and CD62L^{lo} subsets resulted from a random assignment of cells from the CD8$^+$ T cell population into the two subsets. The corresponding statistical question is: Assuming that the CD62L^{hi} and CD62L^{lo} subsets were randomly assigned from the same TCR population, is the similarity between the CD62L^{hi} and CD62L^{lo} subsets lower than expected by chance? To answer this question, we compared the similarity for each CD62L^{hi}/CD62L^{lo} paired sample with the reference similarity expected of two subsets randomly sampled from the pooled population of CD62L^{hi} and CD62L^{lo} TCR sequences. Under the null hypothesis that the CD62L^{hi} and CD62L^{lo} subsets were produced by the random assignment of T cells from the same population, then two TCR sets randomly drawn from the pooled CD62L^{hi} and CD62L^{lo} TCR subsets should have a comparable similarity measure to that of the experimentally observed CD62L^{hi} and CD62L^{lo} subsets.

As an example of the calculation of the reference similarity for the standard sample size (i.e.: 27 TCRs) we again focus on the first D^{b}NP$_{366}$ paired sample, for which 77 and 80 TCRs were sequenced for the CD62L^{hi} and CD62L^{lo} subsets, respectively (Table 1). Two sets of 27 TCRs were randomly drawn, without replacement, from the pooled 77 CD62L^{hi} and 80 CD62L^{lo} TCR sequences (i.e.: a total pool of 157 TCRs) (Fig. 2C) and the Morisita–Horn similarity index calculated. The reference distribution of the Morisita–Horn similarity index for two TCR sets randomly sampled from the pooled population, determined from 10,000 random draws, is shown in Fig. 2D. The reference Morisita–Horn similarity index for two TCR sets randomly sampled from the pooled CD62L^{hi} and CD62L^{lo} TCRs was estimated from the median of the reference distribution to be $C_{MH}^{Ref}=0.93$.

This analysis was repeated for each pair of CD62L^{hi} and CD62L^{lo} TCR subsets to determine a reference similarity for two TCR sets randomly sampled from the pooled CD62L^{hi} and CD62L^{lo} subsets (Fig. 4). The significance of the differences between the similarity estimates for the CD62L^{hi} and CD62L^{lo} subsets (determined in Section 3.1.) and the reference similarities was assessed using a Wilcoxon rank sum test on the Morisita–Horn similarity indices shown in Fig. 4. The

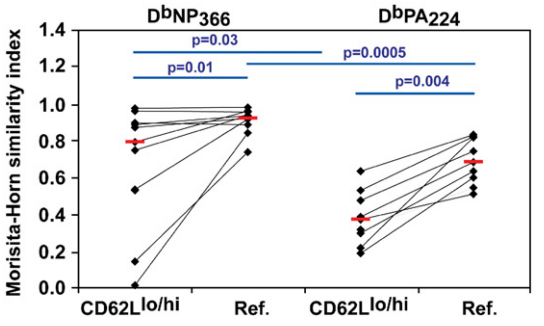

Fig. 4. Comparison of the similarity between the CD62L^{hi} and CD62L^{lo} subsets of the D^{b}NP$_{366}$-and D^{b}PA$_{224}$-specific TCR repertoires with the reference similarity for two TCR sets randomly sampled from the pooled CD62L^{hi} and CD62L^{lo} TCRs, under the assumption that the CD62L^{hi} and CD62L^{lo} subsets were randomly assigned from the same TCR population. The similarity between TCR subsets is quantified using the Morisita–Horn similarity index, which varies between 0 (minimal similarity) and 1 (maximal similarity). All similarities were estimated for a standard sample size of 27 TCR sequences (as described in Figs. 1 and 2). The Morisita–Horn similarity indices for the CD62L^{hi} and CD62L^{lo} subsets (labelled as CD62L$^{lo/hi}$) were compared with the reference Morisita–Horn similarity indices (labelled as Ref.) using a Wilcoxon rank sum test, demonstrating that the experimentally observed CD62L^{hi} and CD62L^{lo} subsets were significantly less similar than could be explained by random sampling from the same population. As was shown in Fig. 3A, CD62L^{hi}/CD62L^{lo} similarities of the samples from the D^{b}NP$_{366}$-specific TCR repertoires were significantly higher than for the D^{b}PA$_{224}$-specific responses. However, the reference similarity measures were also significantly higher for the D^{b}NP$_{366}$-specific responses than for the D^{b}PA$_{224}$-specific responses, suggesting that the lower similarity between the CD62L^{hi} and CD62L^{lo} subsets for the D^{b}PA$_{224}$-specific responses arose, at least in part, due to differences in the clonal dominance hierarchy.

similarity between the CD62L^{hi} and CD62L^{lo} subsets was significantly lower than the references similarities for both the D^{b}NP$_{366}$-specific responses (median: 0.79 vs. 0.93, $p=0.01$) and the D^{b}PA$_{224}$-specific responses (median: 0.37 vs. 0.68, $p=0.004$). Thus, the results of our repertoire analysis show that the CD62L^{hi} and CD62L^{lo} subsets in the primary responses to both the D^{b}NP$_{366}$ and D^{b}PA$_{224}$ epitopes are significantly less similar than expected if the CD62L^{hi} and CD62L^{lo} subsets had been randomly assigned from the same TCR population. This suggests that it is unlikely that the CD8$^+$ T cell population was randomly assigned into the CD62L^{hi} and CD62L^{lo} subsets.

3.5. Does clonal dominance hierarchy play a role in samples of the CD62L^{hi} and CD62L^{lo} subsets of the T cell receptor repertoire being more similar for the D^{b}NP$_{366}$-than for the D^{b}PA$_{224}$-specific responses?

The analysis in Section 3.2. found that the similarity between the CD62L^{hi} and CD62L^{lo} subsets was significantly higher for the D^{b}NP$_{366}$- than for the D^{b}PA$_{224}$-specific responses, raising the question of how much of a role the difference in the clonal dominance hierarchy between the two responses plays in the similarity between the CD62L^{hi} and CD62L^{lo} subsets. A comparison between the reference similarities (i.e.: the similarities for two TCR sets randomly sampled from the pooled CD62L^{hi} and CD62L^{lo} subsets) for the D^{b}NP$_{366}$- and D^{b}PA$_{224}$-specific responses (described in Section 3.4. and displayed in Fig. 4) provides us with some insight into this question. That is, under the null hypothesis that the CD62L^{hi} and CD62L^{lo} subsets were randomly assigned from the same TCR population, would there still be a significant difference in the similarity of the TCR sets between the D^{b}NP$_{366}$- and D^{b}PA$_{224}$-specific responses? It was found that the reference similarities were also significantly higher for the D^{b}NP$_{366}$- than for the D^{b}PA$_{224}$-specific responses (medians: 0.92 vs. 0.68, $p=0.0005$, Mann–Whitney). This suggests that the clonal dominance hierarchy within the responses to the D^{b}NP$_{366}$ epitope plays a large role in the CD62L^{hi} and CD62L^{lo} subsets being more similar in the D^{b}NP$_{366}$- than in the D^{b}PA$_{224}$-specific responses.

One way to better understand how the clonal structure of the response influences the reference similarities is to consider the ‘singletons’ (i.e.: TCR sequences found only once in one but not both observed TCR samples and thus present in only one copy in the pooled population of both observed TCR samples). In the random distribution of TCRs from the pooled TCR population into the two TCR sets, a singleton by definition must be allocated to one set or the other, and thus will never be shared between TCR sets. If there are a high number of singletons in the pooled TCR population, then the two randomly drawn TCR sets will have many ‘unshared’ TCR clonotypes and a low similarity. In contrast, if the observed TCR sample pair, and hence the pooled population of TCRs from both samples, consist mainly of larger clonotypes (i.e.: TCR sequences observed many times in a sample), then both TCR sets randomly drawn from the pooled population of the TCR sample pair will tend to have multiple copies of these larger clonotypes, leading to more shared clonotypes and a higher similarity.

In the context of the occurrence of singleton clonotypes, it is easy to understand why there is a significant difference in the reference similarities for the CD62L^{hi} and CD62L^{lo} subsets between the D^{b}NP$_{366}$- and D^{b}PA$_{224}$-specific responses. The CD8$^+$ T cell response to the D^{b}PA$_{224}$ epitope tends to be clonotypically diverse, consisting of many different TCR clonotypes with a relatively even distribution of clone sizes. Thus,

samples of the D^bPA_{224}-specific TCR repertoire often consist of many singleton clonotypes and many smaller clonotypes, of which only a few copies were sampled. In contrast, the immune response to the D^bNP_{366} epitope is mostly dominated by a few larger TCR clonotypes, and hence, smaller clonotypes are less likely to be sampled and are therefore observed much less frequently.

4. Summary

We have presented a robust approach for assessing the similarity between samples of the T cell receptor repertoire that overcomes some of the common problems encountered in the analysis of TCR data. This approach avoids biases that may arise from inconsistent sampling efforts, includes valuable information about the clonal dominance hierarchy within an epitope-specific response, and enables standard statistical tests to be performed to determine the statistical significance of the observations. Furthermore, this approach is non-parametric and thus does not require any assumptions about the distribution of the data. We propose the Morisita–Horn similarity index for quantifying the overlap between TCR sets and, for cases when consistent sample sizes cannot be obtained by experimental means, provide a method for estimating similarity measures for reduced sample sizes.

We have applied this approach to assessing the similarity between the $CD62L^{hi}$ (central memory) and $CD62L^{lo}$ (effector memory) subsets of D^bNP_{366}- and D^bPA_{224}-specific $CD8^+$ T cell responses in primary influenza infection in mice. Our results suggest that the $CD62L^{hi}$ and $CD62L^{lo}$ subsets are significantly less similar than would be expected if these two subsets had been randomly assigned from the same $CD8^+$ T cell population. We also found that the similarity between the $CD62L^{hi}$ and $CD62L^{lo}$ subsets was not significantly correlated with the time since infection. However, there were only a few samples, with sufficient sample sizes, in long-term memory and only one paired sample (for the D^bPA_{224}-specific $CD8^+$ T cell response at day 690) from a mouse of the age where age-related impairment of the immune responses is most readily observed (Po et al., 2002). Furthermore, we identified a significant difference between the D^bNP_{366}- and D^bPA_{224}-specific T cell responses in the similarity of the $CD62L^{hi}$ and $CD62L^{lo}$ subsets. Our analysis suggests that this difference is, in part, due to the strong clonal dominance hierarchy within the D^bNP_{366}-specific response.

Finally, this study demonstrates that quantitative methods for analysing TCR data can not only provide assurance of the statistical significance of observations but can also enrich the analysis of TCR data by making it possible to address questions that cannot be answered by either simply looking at the data or using methods that rely on ‘counting clonotypes’.

Acknowledgements

We thank Hui Yee Chin for review of this manuscript. This work was supported by the James S. McDonnell Foundation 21st Century Research Award/ Studying Complex Systems, the Australian Research Council (ARC), 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.

Baron, V., Bouneaud, C., Cumano, A., Lim, A., Arstila, T.P., Kourilsky, P., Fesrradini, 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.

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.

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

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.

Davenport, M.P., Price, D.A., McMichael, A.J., 2007. The T cell repertoire in infection and vaccination: implications for control of persistent viruses. Curr. Opin. Immunol. 19, 294.

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.

Hsieh, C.S., Liang, Y., Tyznik, A.J., Self, S.G., Liggitt, D., Rudensky, A.Y., 2004. Recognition of the peripheral self by naturally arising $CD25^+$ $CD4^+$ T cell receptors. Immunity 21, 267.

Hsieh, C.S., Zheng, Y., Liang, Y., Fontenot, J.D., Rudensky, A.Y., 2006. An intersection between the self-reactive regulatory and nonregulatory T cell receptor repertoires. Nat. Immunol. 7, 401.

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., Day, E.B., Pi, J., Heard, S.B., Doherty, P.C., Turner, S.J., Perlman, S., 2006a. Quantification of repertoire diversity of influenza-specific epitopes with predominant public or private TCR usage. J. Immunol. 177, 6705.

Kedzierska, K., Venturi, V., Field, K., Davenport, M.P., Turner, S.J., Doherty, P.C., 2006b. 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.

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.

Lin, M.Y., Welsh, R.M., 1998. Stability and diversity of T cell receptor repertoire usage during lymphocytic choriomeningitis virus infection of mice. J. Exp. Med. 188, 1993.

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

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.

Marzo, A.L., Klonowski, K.D., Le Bon, A., Borrow, P., Tough, D.F., Lefrancois, L., 2005. Initial T cell frequency dictates memory $CD8^+$ T cell lineage commitment. Nat. Immunol. 6, 793.

Masopust, D., Vezys, V., Marzo, A.L., Lefrancois, L., 2001. Preferential localization of effector memory cells in nonlymphoid tissue. Science 291, 2413.

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

McHeyzer-Williams, L.J., Panus, J.F., Mikszta, J.A., McHeyzer-Williams, M.G., 1999. Evolution of antigen-specific T cell receptors in vivo: preimmune and antigen-driven selection of preferred complementarity-determining region 3 (CDR3) motifs. J. Exp. Med. 189, 1823.

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.

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

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.

Po, J.L., Gardner, E.M., Anaraki, F., Katsikis, P.D., Murasko, D.M., 2002. Age-associated decrease in virus-specific $CD8^+$ T lymphocytes during primary influenza infection. Mech. Ageing Dev. 123, 1167.

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.

Reinhardt, R.L., Khoruts, A., Merica, R., Zell, T., Jenkins, M.K., 2001. Visualizing the generation of memory CD4 T cells in the whole body. Nature 410, 101.

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.

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.

Stein, J.V., Cheng, G., Stockton, B.M., Fors, B.P., Butcher, E.C., von Andrian, U.H., 1999. L-selectin-mediated leukocyte adhesion in vivo: microvillous distribution determines tethering efficiency, but not rolling velocity. J. Exp. Med. 189, 37.

Tripp, R.A., Hou, S., Doherty, P.C., 1995. Temporal loss of the activated L-selectin-low phenotype for virus-specific $CD8^+$ memory T cells. J. Immunol. 154, 5870.

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.

Unsoeld, H., Pircher, H., 2005. Complex memory T-cell phenotypes revealed by coexpression of CD62L and CCR7. J. Virol. 79, 4510.

Venturi, V., Kedzierska, K., Turner, S.J., Doherty, P.C., Davenport, M.P., 2007. Methods for comparing the diversity of samples of the T cell receptor repertoire. J. Immunol. Methods 321, 182.

Wherry, E.J., Teichgraber, V., Becker, T.C., Masopust, D., Kaech, S.M., Antia, R., von Andrian, U.H., Ahmed, R., 2003. Lineage relationship and protective immunity of memory CD8 T cell subsets. Nat. Immunol. 4, 225.