Ecological analysis of antigen-specific CTL repertoires defines the relationship between naïve and immune T-cell populations

Paul G. Thomas[a], Andreas Handel[b], Peter C. Doherty[a,c,1], and Nicole L. La Gruta[c,1]

[a]Department of Immunology, St. Jude Children's Research Hospital, Memphis, TN 38105; [b]Department of Epidemiology and Biostatistics, College of Public Health, University of Georgia, Athens, GA 30602; and [c]Department of Microbiology and Immunology, University of Melbourne, Parkville, VIC 3010, Australia

Contributed by Peter C. Doherty, December 20, 2012 (sent for review October 15, 2012)

Ecology is typically thought of as the study of interactions organisms have with each other and their environment and is focused on the distribution and abundance of organisms both within and between environments. On a molecular level, the capacity to probe analogous questions in the field of T-cell immunology is imperative as we acquire substantial datasets both on epitope-specific T-cell populations through high-resolution analyses of T-cell receptor (TCR) use and on global T-cell populations analyzed via high-throughput DNA sequencing. Here, we present the innovative application of existing statistical measures (used typically in the field of ecology), together with unique statistical analyses, to comprehensively assess how the naïve epitope-specific CD8$^+$ cytotoxic T lymphocyte (CTL) repertoire translates to that found following an influenza-virus–specific immune response. Such interrogation of our extensive, cumulated TCR CDR3β sequence datasets, derived from both naïve and immune CD8$^+$ T-cell populations specific for four different influenza-derived epitopes (D^{b}NP$_{366}$, influenza nucleoprotein amino acid residues 366–374; D^{b}PA$_{224}$, influenza acid polymerase amino acid residues 224–233; D^{b}PB1-F2$_{62}$, influenza polymerase B 1 reading frame 2 amino acid residues 62–70; K^{b}NS2$_{114}$, and influenza nonstructural protein 2 amino acid residues 114–121), demonstrates that epitope-specific TCR use in an antiviral immune response is the consequence of a complex interplay between the intrinsic characteristics of the naïve cytotoxic T lymphocyte precursor pool and extrinsic (likely antigen driven) influences, the contribution of which varies in an epitope-specific fashion.

killer T cells | antiviral immunity

For CD8$^+$ cytotoxic T lymphocytes (CTLs), the clonotypic T-cell receptor (TCR) recognizes "nonself" peptides (p) in complex with "self" class I MHC glycoproteins (MHCI). To recognize and respond to the vast array of novel pMHCI determinants encountered by a given individual (or population), the adaptive immune system uses the recombination of variable (V), junctional (J), diversity (D), and constant (C) somatic gene segments to establish an extraordinary spectrum of TCRs. Most of the observed TCR diversity results from imprecise joining of these gene segments and the addition of nontemplate-encoded nucleotides at V(D)J junctions (1). The consequence is that TCR diversity is principally a function of the CDR3 regions, meaning that clonal uniqueness is reliably defined by the TCRαβ CDR3 sequences.

The nature of the highly specific TCR-pMHC recognition event likely determines functional outcomes after T-cell activation (2–6). Thus, the complexity and distribution of TCR use *within* a responding CTL population is likely to have a very real impact on the efficacy of T-cell–mediated pathogen clearance. Certainly, both virus control and the prevention of mutational escape operate more effectively if the host response is characterized by a diverse, rather than a limited, spectrum of pMHCI-specific TCRs (7–12). Further, TCR diversity has also been associated with effective cross-reactive immunity against different virus infections (13).

Single cell RT-PCR analysis of CDR3 nucleotide sequences has provided an unprecedented level of resolution for characterizing TCR use, especially in the case of CTL responses to pathogens. Such detailed dissection of epitope-specific T-cell populations means that shifts in clonal prevalence can be determined over the course of infection, between primary and recall immune responses, or, more recently, between naïve and immune repertoires (14–18). Given the vast amount of data generated by this strategy (if we assume that a rigorously sampled immune repertoire is represented by 50–200 sequences), and the fact that the sequences obtained can only ever provide a snapshot of the total spectrum (necessitating extrapolation to populations), it is imperative to use sophisticated statistical analyses to probe these datasets.

We, and others, have applied established statistical analyses for determining relative levels of diversity and sharing within and between T-cell responses (15, 19). Many of these tests rely on similar measures in the field of ecology, where there is a history of interpreting analogous questions (20). For example, determination of species richness—the number of species in a given area—or species diversity—the number and distribution of species within a given area—has direct correspondence to the interrogation of TCR use within naïve or responding antigen-specific TCR repertoires. The application of such ecology-based analyses to T-cell sets provides unique possibilities for predicting pathogen-specific CTL response efficacy both within and between individuals and populations.

Prior analysis has defined immune TCR repertoires specific for the major pMHCI (p+H-2^b) restricted epitopes in the B6 mouse model of influenza virus infection (14, 15, 17, 21, 22). More recently, using a sensitive enrichment technique (23, 24), we have been able to isolate these influenza epitope-specific CTL precursors directly from naïve mice and characterize their TCR use profiles (18). The present analysis provides a precise and robust analysis of the relationship between endogenous naïve and immune cytotoxic T lymphocyte precursor (CTLp) populations, enhances understanding of the mechanisms behind virus-driven CTL recruitment and expansion, and highlights statistical tools that are useful additions to the analysis of such population data.

Results

Primary Immune Repertoires Show Selective Clonal Expansion/Recruitment. The recently developed capacity to isolate and characterize naïve CTLps from previously uninfected mice (23) has greatly enhanced our capacity to identify key determinants of antiviral CD8$^+$ T-cell responses. Our previous analysis compared TCR repertoire characteristics for four epitope-specific CTL sets (D^{b}NP$_{366}$, D^{b}PA$_{224}$, D^{b}PB1-F2$_{62}$, and K^{b}NS2$_{114}$) in naïve mice

IMMUNOLOGY

Author contributions: P.G.T., A.H., and N.L.L.G. designed research; P.G.T., A.H., and N.L.L.G. performed research; A.H. contributed new reagents/analytic tools; P.G.T., A.H., and N.L.L.G. analyzed data; and P.G.T., P.C.D., and N.L.L.G. wrote the paper.

The authors declare no conflict of interest.

[1]To whom correspondence may be addressed. E-mail: nllg@unimelb.edu.au or pcd@unimelb.edu.au.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1222149110/-/DCSupplemental.

and following primary virus challenge (18). In the present study, we have applied several statistically robust measures to address the extent of selective recruitment/expansion of particular CTL clones during these primary CTL responses.

Firstly, we used Simpson's Diversity Index (SDI), which is a function of both the relative number of clonotypes present (richness) and the relative abundance or distribution of each clonotype (evenness) (20, 25). Analysis of SDI within the naïve and immune D^bNP_{366}-, D^bPA_{224}-, $D^bPB1\text{-}F2_{62}$-, and K^bNS2_{114}-specific populations, established that all epitope-specific immune sets were significantly less diverse than the corresponding naïve CTLp pools (Fig. 1*A*). This is not due to any decrease in clonotype richness in the antigen-expanded repertoires (18) (Table 1), reflecting rather a less even distribution of all clonotypes (and thus an overrepresentation of particular clones) in all of the immune populations analyzed.

The same effect can be observed if we analyze the size distributions of clonotypes within naïve and immune T-cell populations (Fig. 1*B*). Clonotypes were first ranked according to size (largest to smallest), and the cumulative number of clonotypes making up a given proportion of the response was then calculated. This was then normalized to the percentage of total clonotypes, allowing us to determine how clones contribute to the total epitope-specific CTL population. Due to the predominance of singletons within the naïve CTLp repertoires there is, broadly speaking, a one-to-one ratio between the proportion of clonotypes and their overall contribution to the response (e.g., ~60% of clonotypes make up 60% of the total response). This is somewhat skewed for the naïve D^bNP_{366}-specific set (Fig. 1*B*) due to the presence of a greater number of repeated clonotypes (i.e., present at a frequency of >1). Interestingly, looking at immune repertoires, it is apparent that a smaller proportion of the clonotypes contributes far more to the total population, compared with the naïve repertoires. For example, in the D^bPA_{224}- and $D^bPB1\text{-}F2_{62}$-specific immune sets, ~50% of all clonotypes account for 80% of the response (Fig. 1*B*). This unevenness is even more pronounced in the D^bNP_{366}- and K^bNS2_{114}-specific populations, which are characterized by a marked hierarchical structure in which one or a few clonotypes dominate the response. In this case, only around 35% of all clonotypes account for 80% of the total response. Thus, for all four pMHCI specificities, there is clear evidence for the selective expansion of particular clonotypes such that they become significantly overrepresented in the immune, compared with the naïve, repertoire.

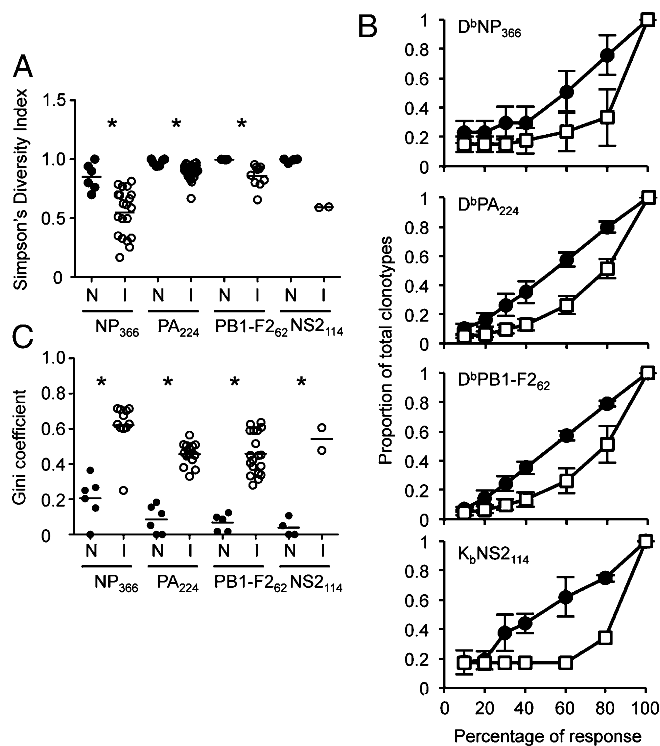

Fig. 1. Selective clonotype expansion in immune epitope-specific responses. (*A*) Clonotypic diversity within naïve (N, closed symbols) and immune (I, open symbols) mice was measured using Simpson's diversity index (*$P < 0.01$ using Student's unpaired *t* test). (*B*) Evenness of clonotype distribution in naïve and immune mice was determined by ranking clonotypes from largest to smallest for each mouse and the cumulative proportion of total clonotypes was plotted against the cumulative percentage of the response (percentage of the total sequences). Shown are the means ± SD of data obtained from 6, 6, 5, and 4 naïve mice (closed symbols) and 11, 13, 19, and 2 immune mice (open symbols), for D^bNP_{366}, D^bPA_{224}, $D^bPB1\text{-}F2_{62}$, and K^bNS2_{114}, respectively. (*C*) Statistical comparison of clonotype distribution between naïve (N, closed symbols) and immune (I, open symbols) populations was performed using the Gini coefficient (26) (*$P < 0.001$, comparing naïve to immune sets).

The Gini coefficient is commonly used in economics and ecology to determine the degree to which some commodity is distributed among individuals (26) and is, as such, a direct measure of equality of distribution. For our purposes, we can use this approach to measure the equality of distribution for individual TCR clones in naïve and immune epitope-specific CTL populations. The Gini coefficient yields a value between 0 and 1: a Gini coefficient of 0 represents a perfectly even distribution, and a higher value represents a less even distribution. As shown in Fig. 1*C*, in all epitope specific responses there is a substantial and statistically significant shift toward unevenness (i.e., where the relative abundance of various clonotypes is substantially different) in the immune, compared with the naïve, compartment.

Collectively, these data demonstrate biased accumulation of individual clones as a consequence of selective expansion from the naïve compartment that cannot be explained by their relative frequencies in the naïve CTLp set. Clearly, extrinsic factors (antigen, inflammatory environment) and/or the unique sensitivities of particular TCRs are key determinants of T-cell clonal composition in any given immune response.

Naïve Repertoire Characteristics Define "Public" vs. "Private" Immune Repertoires. In epitope-specific immune responses in both mice and humans, the same clonotypes (defined by CDR3β sequence) may be observed in multiple MHC-matched individuals (reviewed in refs. 27, 28). The extent of such clonotype "sharing results in epitope-specific responses being classified as predominantly public (heavily shared) or private (largely unique to individuals) (29–31). We determined the extent of clonotype sharing for three epitope-specific naïve and immune repertoires (D^bNP_{366}, D^bPA_{224}, and $D^bPB1\text{-}F2_{62}$) by calculating the proportions of particular CDR3β clonotypes found in at least 33% of mice analyzed (Fig. 2*A*). The K^bNS2_{114} epitope was excluded from this analysis because immune repertoires were available for only two mice. Note that the abundance of a given clonotype, and thus the contribution to the total detected response, is taken into account. Corresponding to their well-characterized immune counterparts, the naïve D^bNP_{366}-specific repertoires showed the greatest degree of sharing between mice, with ~65% of an individual's repertoire identified in other mice. In contrast, the naïve D^bPA_{224}- and $D^bPB1\text{-}F2_{62}$-specific repertoires were substantially private, with less than 20% overlap in clonotype use (Fig. 2*A*). Comparable results were observed when the sharing threshold was set to 66% of mice analyzed (Fig. S1). Broadly speaking, therefore, the public or private nature of any given response is largely determined by the characteristics of the naïve repertoire.

Table 1. Naïve and immune epitope-specific TCR clonotype richness

	Naïve repertoire			Immune repertoire		
Epitope	No. of mice	Seqs/mouse	Clonotypes/mouse*	No. of mice	Seqs/mouse	Clonotypes/mouse*
NP_{366}	6	7.2 (8)	4.8 ± 2.1 (5.7 ± 2.7)	11	65	7.6 ± 3.2
PA_{224}	6	16.7 (17.5)	15.3 ± 3.3 (15.8 ± 2.4)	13	58	21.8 ± 4.3
PB1-$F2_{62}$	5	47.4	43.6 ± 12.3	19	79	25.6 ± 9.8
$NS2_{114}$	4	8.75	8.25 ± 6.7	2	50.5	6

Numbers in parentheses include cells that yielded PCR product but from which no TCRβ sequence was obtained.
*Values represent mean ± SD.

Clonotype Dominance Does Not Necessarily Predict Sharing. The extent to which a particular TCR clonotype is shared between individuals has been closely linked to the dominance of that clonotype within individuals. Both are thought to be a consequence of convergent recombination, which suggests that the CDR3 amino acid (aa) sequences generated more frequently by the recombination machinery (influenced predominantly, but not exclusively, by the number of different nucleotide sequences that can encode the same amino acid sequence) are likely to dominate the naïve repertoire and thus the immune response after infection (31–33). This model thus predicts a three-way correlation between clonotype frequency in naïve populations, within immune populations, and between immune populations (sharing). Using our unique database of TCRβ clonotypes from a number of naïve epitope-specific CTLp populations (18), along with a large database of corresponding immune TCR CDR3β sequences (14, 15, 17), we sought to investigate these relationships. Our previous demonstration of significant clonotype overrepresentation in immune (compared with naïve) repertoires (Fig. 1) indicated that clonal prevalence in the immune response could not be solely explained by the frequency with which particular clonotypes were generated.

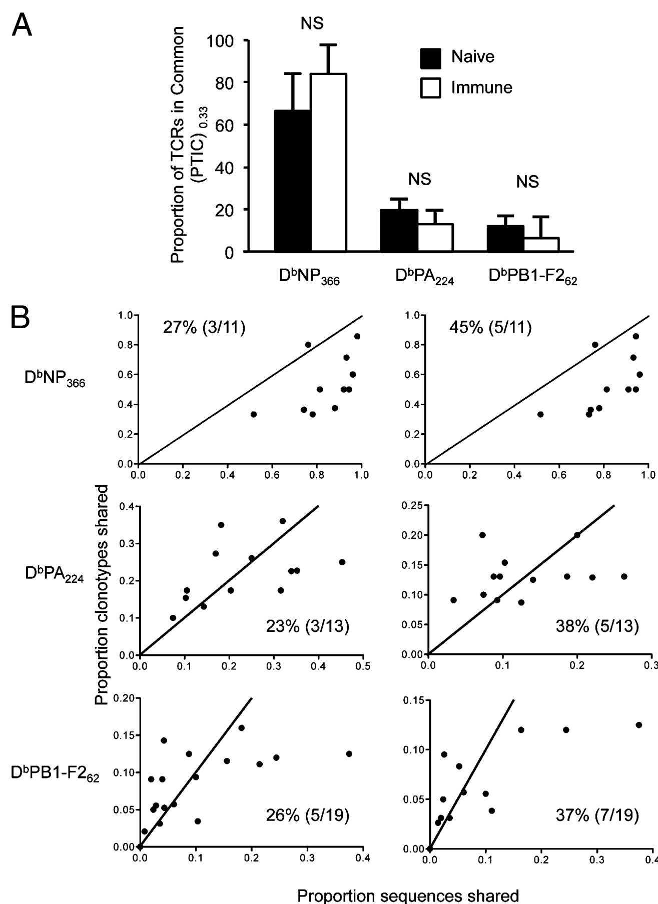

Fig. 2. Clonotype sharing versus abundance in virus-specific CTL responses. (*A*) PTIC is the proportion of the clonotypic response from individual immune mice that is shared by at least 33% of mice sampled. Shown is mean ± SD for sequence data described in Fig. 1. Statistical analyses were performed using the Mann–Whitney test. NS, not significant, $P > 0.05$. (*B*) Relative abundance of shared clonotypes was determined by plotting, for each mouse, the proportion of clonotypes shared in ~25% (first column) or ~40% (second column) of mice (actual percentage of mice is noted within each plot and is determined by the number of mice analyzed) against the proportion of the total response those clonotypes contribute (proportion sequences shared).

So, what determines clonotype prevalence within immune individuals versus the extent of clonotype sharing between these individuals? Looking more closely at clonotype sharing in the D^bNP_{366}-specific sets (Fig. 2*A*), we observed a small increase in the proportion of the response shared between immune (compared with naïve) individuals. This reflected the fact (observed previously) that clonotype sharing in the D^bNP_{366}-specific repertoire largely involves the dominant (or overrepresented) clonotypes (14, 31). However, such increases in response sharing were not observed for the D^bPA_{224}- and $D^bPB1–F2_{62}$-specific populations (Fig. 2*A*), suggesting the large clonotypes are not being shared in these populations. To look at this directly, we plotted (for each mouse) the proportion of clonotypes shared (in a designated percentage of mice) against the proportion of all shared sequences (Fig. 2*B*). An $x = y$ plot has been used to show a theoretical linear relationship between these parameters. If (as an example) we take a mouse for which 100 TCR sequences were obtained, revealing 10 unique clonotypes, the data point for this individual would fall on the $x = y$ line if each shared clonotype had a frequency of 10/100 (i.e., the same frequency at which the clonotype was detected). However, if shared clonotypes in this mouse were large, then, whereas their clonotype frequency would remain at 1/10, their contribution to the total sequences would be increased and thus the data point for this mouse would be to the right of the $x = y$ line. Similarly, data points to the left of the line not only indicate that the shared clonotypes in a given mouse are not, on average, abundant within the individual, but also that there are large clonotypes that are not shared.

As expected, shared D^bNP_{366}-specific clones were generally highly abundant within mice (Fig. 2*B*), verifying the link demonstrated previously between clonotype abundance within an individual and the likelihood of that clonotype being shared across the population (14, 31). However, this correlation was not observed for the D^bPA_{224}- and $D^bPB1–F2_{62}$-specific immune repertoires, because shared clonotypes appeared equally likely to be over- or underrepresented within mice (Fig. 2*B*). Thus, for these two pMHCI specificities, the factors influencing clonotype frequency within immune individuals appear to be distinct from those determining clonotype prevalence between immune individuals (sharing).

Nucleotide Sequence Flexibility Promotes Clonal Prevalence Between, Rather than Within, Naïve Individuals. As mentioned above, a key tenet of convergent recombination is that the prevalence of

a particular TCR amino acid clonotype is defined in large part by the number of different nucleotide sequences that can be used to encode it (30–33). Thus, it is hypothesized that particular TCR clonotypes will be more prevalent if they can be encoded by a larger number of nucleotide sequences. We investigated the number and distribution of nucleotide sequences for all naïve TCRβ clonotypes (defined by amino acid) that were observed at a prevalence of >1 (Table S1). As predicted, and shown previously for several of the public D^bNP_{366}-specific TCRβ clonotypes (31), many of these clonotypes could be encoded by more than one nucleotide sequence. To determine whether the diversity of nucleotide sequences influenced the prevalence of a clonotype within or between individuals, we generated statistics to quantify the extent to which nucleotide sequences recur within or between mice (*Materials and Methods*). This statistic measures the proportion of duplicated nucleotide sequences for a given epitope either within an individual animal or across animals. To determine whether these values were statistically different from a random distribution, we used two methods to randomly distribute the nucleotide sequences that were identified experimentally. Firstly, all of the observed nucleotide sequences were randomly assigned to one of six theoretical mice. The extent of nucleotide sequence recurrence was then calculated within and between mice for this artificially produced dataset. This simulation was repeated 1,000 times and plotted alongside the sharing statistic obtained from the experimentally acquired data (Fig. 3*A*). Secondly, we used an epitope-specific sampling approach, where the nucleotide sequence sampling was done with replacement. Here, the number of sequences within each mouse, for each epitope, was determined by random drawing of a number from a Poisson distribution centered on the mean size of the response in the experimental data. Then each mouse was populated with sequences drawn from the experimental data with equal probability. This simulation was repeated 1,000 times and the results are shown in Fig. 3*B*, alongside the actual sharing data.

This analysis demonstrates that the extent of nucleotide sequence recurrence observed within mice is significantly higher than would be predicted from a random distribution, and conversely, that the extent of nucleotide sequence recurrence between mice is significantly less than would be predicted. Thus, these data indicate that, whereas nucleotide diversity is likely to influence the prevalence of clonotypes between mice (sharing), factors other than nucleotide diversity are responsible for the overrepresentation of clonotypes within mice (clonal dominance). This fits with (and likely explains) our previous observation that clonotype abundance within an individual does not necessarily correlate with abundance between individuals. In fact, the repetition of particular nucleotide sequences within individuals may reflect the duplication of particular clones as a result of CTL division, although the possibility that each is a unique clonotype (as defined by a particular TCRα sequence) cannot be excluded based on this analysis.

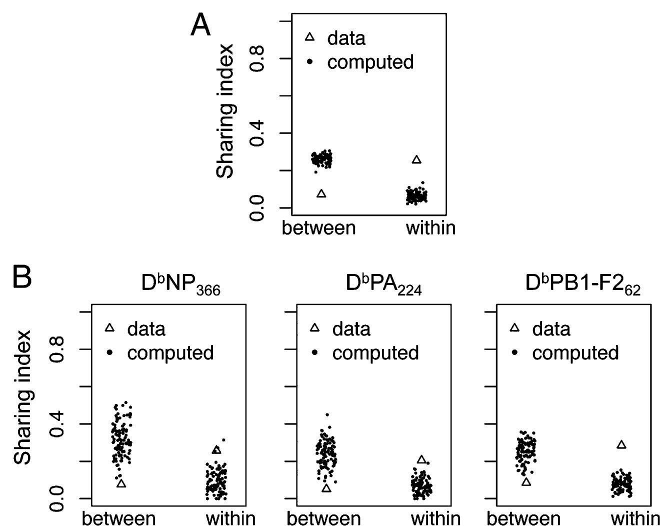

Fig. 3. Nucleotide flexibility within and between naïve individuals. Model sharing indices were generated using two methods of rarefaction to test how the measures of within- and between-mouse sharing compared with a random distribution. (*A*) All experimentally observed nucleotide sequences from naïve animals were randomly assigned to one of six theoretical "mice" (without replacement) and the within-mouse and between-mice sharing measures were calculated for this artificially produced dataset, with 1,000 total simulations. (*B*) For each epitope specificity, sequences were randomly distributed among theoretical mice, by randomly drawing from the pool of sequences, with each sequence having an equal probability of being chosen in each draw (i.e., sampling with replacement). Within-mouse and between-mice sharing is then calculated (as described) for the theoretical set (dots; a representative 100 of a total 1,000 simulations) or actual data (triangles).

Discussion

Studies that have directly enumerated naïve $CD4^+$ and $CD8^+$ T-cell precursors specific for a range of antigens (reviewed in ref. 34) have largely indicated that precursor frequency is a key predictor of response magnitude after antigen challenge (23, 24, 35–38), implying that naïve populations are comprehensively recruited and expanded in the immune response (39). Recently, however, our work and that of others has shown a disconnect between naïve CTLp frequency and immune magnitudes (18, 40) which, in our study, resulted from incomplete recruitment and/or expansion of T-cell populations after antigen challenge. Given that T-cell precursor recruitment and expansion is not necessarily uniform following antigen challenge (even where relative precursor frequencies and immune magnitudes correlate, the relationship is rarely absolute), the current study sought to elucidate how naïve CTLp repertoires influence corresponding immune repertoires after influenza virus infection. The interrogation of extensive TCR CDR3β sequence data (naïve and immune) using statistical analyses that have been adapted from different disciplines, or developed specifically for this study, has allowed us to identify the key drivers that dictate the nature of virus-specific immune populations after antigen challenge.

A key finding in the present study is that the prevalence of particular T-cell clones across multiple individuals (or the extent of clone sharing) is not intrinsically linked to the abundance of those clones within a given immune response. This appears to contradict the convergent recombination hypothesis, which suggests that TCR clonotypes that can be generated from numerous different nucleotide sequences will be present at a high frequency, both within and between individuals (31–33). Similar discordances in clonal prevalence within and between individuals have been previously noted in, for example, human $CD8^+$ T-cell populations specific for the Epstein Barr Virus BMLF $1_{280–288}$ epitope or the Cytomegalovirus pp65 $_{495–503}$ epitope (41). In these cases it was proposed that clonal dominance within immune individuals may be due to the multitude of antigen-driven factors that influence the differential involvement of T-cell clones in an immune response, despite their prevalence within the CTLp pool. In fact, our data show that the influence of convergent recombination is exerted predominantly on clonal prevalence between, rather than within, individuals, even in the starting naïve repertoire. This provides an explanation for the fact that heavily shared T-cell clones are not necessarily abundant within individuals.

Intriguingly, we observed selective expansion of particular clones from naïve to immune repertoires for all of the epitope-specific populations studied. This is unlikely to be explained by convergent recombination for D^bPA_{224} or $D^bPB1\text{-}F2_{62}$ (D^bNP_{366} is discussed below) because of the lack of correlation between immune

clonotype abundance and sharing in the immune response. The fact that (within individuals) TCR clonotypes tend to be encoded by the same nucleotide sequence suggested that dominance within mice may instead be a consequence of low-level cycling of particular T-cell clones before cognate antigen recognition. This is supported by previous studies showing that a small, but significant, proportion of epitope-specific $CD8^+$ T cells from naïve mice express high levels of CD44, a marker of antigen experience (18, 42). Such $CD44^{hi}$ naïve cells are thought to arise as a consequence of homeostatic proliferation and respond more rapidly in vitro after antigen challenge (42). Thus, we are currently investigating the possibility that T-cell clonotypes that are overrepresented within naïve mice derive from single clones and have the capacity to respond rapidly after infection, resulting in their selective prominence in the immune repertoire. Alternatively, the dominance of any clone in the immune response may be a consequence of preferential recruitment or expansion that is conferred by superior TCR/pMHC avidity or "fit" (4, 43–47), or by stochastic effects that occur during early priming and proliferation. Indeed, clonal dominance may arise as a consequence of distinct influences depending on viral dose, with a recent study indicating that the relative size of epitope-specific CTL responses is predominantly influenced by antigen presentation at low viral doses, and by TCR avidity and CTLp frequency when antigen is no longer limiting (48). Importantly, it remains possible that recurrent nucleotide sequences within naïve individuals do reflect unique clonotypes, differentiated based on TCRα chain sequence. In that case, the recurrence of sequences within (and not between) mice becomes difficult to explain.

The strong correlation between clonal prevalence within and between mice in the immune D^bNP_{366}-specific response (with the large clones being statistically the most likely to be shared) suggests that, for this epitope at least, convergent recombination can influence clonal prevalence at both the individual and the population levels. The difference between D^bNP_{366} and other epitope-specific populations analyzed here and elsewhere (41) is likely to be the vast number of different nucleotide sequences that can be used to encode the large public TCR clonotypes (Table S1) (31). In this case, such extreme nucleotide flexibility may indeed influence within-individual prevalence (albeit to a significantly smaller extent than that found between individuals). Thus, it would appear that the preferential expansion of these D^bNP_{366}-specific T-cell clones may have more to do with their exceptionally high frequency in the naïve population (rather than any differences in TCR avidity or fit), conferring an early proliferative advantage.

Indeed, the D^bNP_{366}-specific immune repertoire shows an extremely pronounced hierarchical organization of clonotypes where one clonotype can contribute up to 70–80% of the TRBV13-1^+ immune response (14, 49). Furthermore, we have shown that, in mice lacking the enzyme terminal deoxynucleotidyl transferase (TdT), the almost exclusive use of the public TCRβ SGGANTGQQL TRBV13-1^+ TCR corresponded to a low avidity response (29). Thus, for D^bNP_{366}, and possibly other epitope-specific repertoires where an exceptionally large number of possible nucleotide sequences can encode a given TCR, the elevated frequency of naïve CTLps expressing this TCR may override any other stochastic or antigen-driven influences.

In summary, we have performed a biologically relevant analysis of naïve and immune epitope-specific CTL populations, using appropriate and robust statistical measures typically reserved for the study of ecology or economics. The need for such statistical measures in the analysis of TCR repertoire data is becoming more pressing as vast datasets are being generated through bulk and single cell RT-PCR analyses of TCR use, as well as the rapid emergence of high-throughput DNA sequencing data (50–54). The findings from these analyses, most notably the selective expansion of particular TCRβ clonotypes in all epitope-specific populations, as well as the determinants of clonotype dominance within and between individuals, provide key insights into the mechanisms controlling $CD8^+$ T-cell use in an antiviral immune response that are likely to be broadly applicable to a range of CTL specificities.

Materials and Methods

The CDR3β datasets that are analyzed here are derived from earlier published work (14, 15, 17, 18) and additional data.

Statistical Analyses. SDI was used to measure clonotype diversity within individuals (20, 25). SDI is calculated as $D = \Sigma_i[(n_i(n_i - 1))/(N(N - 1))]$, where n_i is the number of sequences in the *i*th clonotype and *N* is the total number of sequences in the whole population. This is a measure of diversity, which is defined as a combination of *richness* (the number of different clonotypes within a population) and *evenness* (the distribution of the clonotypes) and expressed as 1-D so that a higher number corresponds to greater diversity. To measure clonotype sharing, we determined the proportion of sequences that are found in a specific percentage of the mice analyzed. This is referred to as proportion of TCRs in common (PTICq) (29), with the q referring to the percentage of mice where the sequences must be found. For example, here we have used PTIC0.33, which is calculated as $\Sigma_q(n_q/N)$ for all sequences "n" where "n" is present in at least 33% (q) of the mice. The Mann–Whitney test, or Student's unpaired *t* test, was used to determine significance for all individual comparisons.

Gini Coefficients. Calculated within individuals and then averaged for all mice, the Gini coefficient defines the distribution of clonotype use for each epitope. This provides a measure of inequality, based on a curve plotting the cumulative proportion of total sequences on the *y* axis relative to the individual clonotype use on the *x* axis, with the coefficient being calculated using the ineq package in R (55). The quantity varies between 0 and 1, with 1 representing extreme inequality and 0 representing an equal distribution of sequences across all clonotypes.

Nucleotide Sequence Sharing. We computed measures for CDR3β nucleotide sequence sharing both within and between mice for naïve CTLp populations, as follows (detailed examples of these calculations are provided in *SI Materials and Methods*).

Within-mouse sharing. For each amino acid sequence, we counted the number of duplicate nucleotide sequences within a mouse. All duplicates were summed and divided by the total number of nucleotide sequences for a given amino acid sequence.

Between-mice sharing. For each amino acid sequence, we counted the number of duplicate mice in which a given nucleotide sequence appeared. Duplicates for each nucleotide sequence were summed and divided by the total number of nucleotide sequences for a given amino acid sequence.

Finally, we take the mean of the within-mouse and between-mice sharing values over all amino acid sequences to arrive at overall measures of within-mouse and between-mice sharing for a given dataset.

To test how these measures of within- and between-mouse sharing compare with a random distribution, we used two sampling methods to generate model-sharing indices. In the first, we took an "all data sampling approach," where all experimentally observed nucleotide sequences from naïve animals were completely randomly assigned to one of six "theoretical mice," in line with the six mice analyzed in the experiments (five for $D^bPB1\text{-}F2_{62}$). We then computed the within-mouse and between-mice sharing measures for this artificially produced dataset. The random assignment of nucleotide sequences and sharing computation was repeated 1,000 times.

In the second sampling approach, we performed sampling for each epitope individually. We start by constructing six "empty mice." Each artificial mouse has slots for a given number of nucleotide sequences. The number of slots is determined by drawing an integer number (≥1) from a Poisson distribution, the mean of which is determined by the average of the number of samples observed in the six real mice. Next, for each mouse we fill the slots by sampling, with replacement nucleotide sequences from the pool of experimentally observed nucleotide sequences, with equal probability for each sequence, until the mouse is "full."

Once a set of artificial mice is constructed, within-mouse and between-mice sharing is computed as described above.

ACKNOWLEDGMENTS. We thank Prof. Stephen Turner for helpful discussions. This work was supported by National Health and Medical Research Council (NHMRC) Project Grant AI628316 (to N.L.L.G.), NHMRC Program Grant AI567122 (to P.C.D.), an NHMRC R. D. Wright Career Development Award (to N.L.G.), a Sylvia and Charles Viertel Senior Medical Research Fellowship (to N.L.L.G.), and National Institutes of Health Grants AI70251 (to P.C.D.), AI065097 and AI077714 (to P.G.T.), and AI072193 (to A.H.).

1. Cabaniols JP, Fazilleau N, Casrouge A, Kourilsky P, Kanellopoulos JM (2001) Most alpha/beta T cell receptor diversity is due to terminal deoxynucleotidyl transferase. *J Exp Med* 194(9):1385–1390.
2. Faroudi M, et al. (2003) Lytic versus stimulatory synapse in cytotoxic T lymphocyte/target cell interaction: Manifestation of a dual activation threshold. *Proc Natl Acad Sci USA* 100(24):14145–14150.
3. Malherbe L, Hausl C, Teyton L, McHeyzer-Williams MG (2004) Clonal selection of helper T cells is determined by an affinity threshold with no further skewing of TCR binding properties. *Immunity* 21(5):669–679.
4. Zehn D, Lee SY, Bevan MJ (2009) Complete but curtailed T-cell response to very low-affinity antigen. *Nature* 458(7235):211–214.
5. Schmid DA, et al. (2010) Evidence for a TCR affinity threshold delimiting maximal CD8 T cell function. *J Immunol* 184(9):4936–4946.
6. Gett AV, Sallusto F, Lanzavecchia A, Geginat J (2003) T cell fitness determined by signal strength. *Nat Immunol* 4(4):355–360.
7. Messaoudi I, Guevara Patiño JA, Dyall R, LeMaoult J, Nikolich-Zugich J (2002) Direct link between mhc polymorphism, T cell avidity, and diversity in immune defense. *Science* 298(5599):1797–1800.
8. Cornberg M, et al. (2006) Narrowed TCR repertoire and viral escape as a consequence of heterologous immunity. *J Clin Invest* 116(5):1443–1456.
9. Price DA, et al. (2004) T cell receptor recognition motifs govern immune escape patterns in acute SIV infection. *Immunity* 21(6):793–803.
10. Charini WA, et al. (2001) Clonally diverse CTL response to a dominant viral epitope recognizes potential epitope variants. *J Immunol* 167(9):4996–5003.
11. Chen H, et al. (2012) TCR clonotypes modulate the protective effect of HLA class I molecules in HIV-1 infection. *Nat Immunol* 13(7):691–700.
12. Wang GC, Dash P, McCullers JA, Doherty PC, Thomas PG (2012) T cell receptor alphabeta diversity inversely correlates with pathogen-specific antibody levels in human cytomegalovirus infection. *Sci Transl Med* 4(128):128ra142.
13. Welsh RM, Che JW, Brehm MA, Selin LK (2010) Heterologous immunity between viruses. *Immunol Rev* 235(1):244–266.
14. Kedzierska K, Turner SJ, Doherty PC (2004) Conserved T cell receptor usage in primary and recall responses to an immunodominant influenza virus nucleoprotein epitope. *Proc Natl Acad Sci USA* 101(14):4942–4947.
15. La Gruta NL, et al. (2008) Epitope-specific TCRbeta repertoire diversity imparts no functional advantage on the CD8+ T cell response to cognate viral peptides. *Proc Natl Acad Sci USA* 105(6):2034–2039.
16. Maryanski JL, Jongeneel CV, Bucher P, Casanova JL, Walker PR (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(1):47–55.
17. Turner SJ, Diaz G, Cross R, Doherty PC (2003) Analysis of clonotype distribution and persistence for an influenza virus-specific CD8+ T cell response. *Immunity* 18(4):549–559.
18. La Gruta NL, et al. (2010) Primary CTL response magnitude in mice is determined by the extent of naive T cell recruitment and subsequent clonal expansion. *J Clin Invest* 120(6):1885–1894.
19. Venturi V, et al. (2008) Method for assessing the similarity between subsets of the T cell receptor repertoire. *J Immunol Methods* 329(1-2):67–80.
20. Magurran AE (2004) *Measuring Biological Diversity* (Blackwell Publishing, Oxford, UK).
21. Dash P, et al. (2011) Paired analysis of TCRα and TCRβ chains at the single-cell level in mice. *J Clin Invest* 121(1):288–295.
22. Zhong W, et al. (2007) CTL recognition of a protective immunodominant influenza A virus nucleoprotein epitope utilizes a highly restricted Vbeta but diverse Valpha repertoire: functional and structural implications. *J Mol Biol* 372(2):535–548.
23. Moon JJ, et al. (2007) Naive CD4(+) T cell frequency varies for different epitopes and predicts repertoire diversity and response magnitude. *Immunity* 27(2):203–213.
24. Obar JJ, Khanna KM, Lefrançois L (2008) Endogenous naive CD8+ T cell precursor frequency regulates primary and memory responses to infection. *Immunity* 28(6):859–869.
25. Venturi V, Kedzierska K, Turner SJ, Doherty PC, Davenport MP (2007) Methods for comparing the diversity of samples of the T cell receptor repertoire. *J Immunol Methods* 321(1-2):182–195.
26. Sadras V, Bongiovanni R (2004) Use of Lorenz curves and Gini coefficients to assess yield inequality within paddocks. *Field Crops Res* 90(2–3):303–310.
27. Li H, Ye C, Ji G, Han J (2012) Determinants of public T cell responses. *Cell Res* 22(1):33–42.
28. Miles JJ, Douek DC, Price DA (2011) Bias in the αβ T-cell repertoire: Implications for disease pathogenesis and vaccination. *Immunol Cell Biol* 89(3):375–387.
29. Kedzierska K, et al. (2008) Terminal deoxynucleotidyltransferase is required for the establishment of private virus-specific CD8+ TCR repertoires and facilitates optimal CTL responses. *J Immunol* 181(4):2556–2562.
30. Venturi V, Chin HY, Price DA, Douek DC, Davenport MP (2008) The role of production frequency in the sharing of simian immunodeficiency virus-specific CD8+ TCRs between macaques. *J Immunol* 181(4):2597–2609.
31. Venturi V, et al. (2006) Sharing of T cell receptors in antigen-specific responses is driven by convergent recombination. *Proc Natl Acad Sci USA* 103(49):18691–18696.
32. Quigley MF, et al. (2010) Convergent recombination shapes the clonotypic landscape of the naive T-cell repertoire. *Proc Natl Acad Sci USA* 107(45):19414–19419.
33. Venturi V, Price DA, Douek DC, Davenport MP (2008) The molecular basis for public T-cell responses? *Nat Rev Immunol* 8(3):231–238.
34. Jenkins MK, Moon JJ (2012) The role of naive T cell precursor frequency and recruitment in dictating immune response magnitude. *J Immunol* 188(9):4135–4140.
35. Schmidt J, et al. (2011) Immunodominance of HLA-A2-restricted hepatitis C virus-specific CD8+ T cell responses is linked to naive-precursor frequency. *J Virol* 85(10):5232–5236.
36. Kotturi MF, et al. (2008) Naive precursor frequencies and MHC binding rather than the degree of epitope diversity shape CD8+ T cell immunodominance. *J Immunol* 181(3):2124–2133.
37. Tan AC, La Gruta NL, Zeng W, Jackson DC (2011) Precursor frequency and competition dictate the HLA-A2-restricted CD8+ T cell responses to influenza A infection and vaccination in HLA-A2.1 transgenic mice. *J Immunol* 187(4):1895–1902.
38. Flesch IE, et al. (2010) Altered CD8(+) T cell immunodominance after vaccinia virus infection and the naive repertoire in inbred and F(1) mice. *J Immunol* 184(1):45–55.
39. van Heijst JW, et al. (2009) Recruitment of antigen-specific CD8+ T cells in response to infection is markedly efficient. *Science* 325(5945):1265–1269.
40. Ruckwardt TJ, et al. (2011) Neonatal CD8 T-cell hierarchy is distinct from adults and is influenced by intrinsic T cell properties in respiratory syncytial virus infected mice. *PLoS Pathog* 7(12):e1002377.
41. Venturi V, et al. (2008) TCR beta-chain sharing in human CD8+ T cell responses to cytomegalovirus and EBV. *J Immunol* 181(11):7853–7862.
42. Haluszczak C, et al. (2009) The antigen-specific CD8+ T cell repertoire in unimmunized mice includes memory phenotype cells bearing markers of homeostatic expansion. *J Exp Med* 206(2):435–448.
43. Callan MF, et al. (1996) Large clonal expansions of CD8+ T cells in acute infectious mononucleosis. *Nat Med* 2(8):906–911.
44. Price DA, et al. (2005) Avidity for antigen shapes clonal dominance in CD8+ T cell populations specific for persistent DNA viruses. *J Exp Med* 202(10):1349–1361.
45. Busch DH, Pamer EG (1999) T cell affinity maturation by selective expansion during infection. *J Exp Med* 189(4):701–710.
46. Kjer-Nielsen L, et al. (2003) A structural basis for the selection of dominant alphabeta T cell receptors in antiviral immunity. *Immunity* 18(1):53–64.
47. Stewart-Jones GB, McMichael AJ, Bell JI, Stuart DI, Jones EY (2003) A structural basis for immunodominant human T cell receptor recognition. *Nat Immunol* 4(7):657–663.
48. Luciani F, Sanders MT, Oveissi S, Pang KC, Chen W (2013) Increasing viral dose causes a reversal in CD8+ T Cell immunodominance during primary influenza infection due to differences in antigen presentation, T cell avidity, and precursor numbers. *J Immunol* 190(1):36–47.
49. Zhong W, Reinherz EL (2004) In vivo selection of a TCR Vbeta repertoire directed against an immunodominant influenza virus CTL epitope. *Int Immunol* 16(11):1549–1559.
50. Freeman JD, Warren RL, Webb JR, Nelson BH, Holt RA (2009) Profiling the T-cell receptor beta-chain repertoire by massively parallel sequencing. *Genome Res* 19(10):1817–1824.
51. Robins HS, et al. (2009) Comprehensive assessment of T-cell receptor beta-chain diversity in alphabeta T cells. *Blood* 114(19):4099–4107.
52. Boyd SD, et al. (2009) Measurement and clinical monitoring of human lymphocyte clonality by massively parallel VDJ pyrosequencing. *Sci Transl Med* 1(12):12ra23.
53. Wang C, et al. (2010) High throughput sequencing reveals a complex pattern of dynamic interrelationships among human T cell subsets. *Proc Natl Acad Sci USA* 107(4):1518–1523.
54. Venturi V, et al. (2011) A mechanism for TCR sharing between T cell subsets and individuals revealed by pyrosequencing. *J Immunol* 186(7):4285–4294.
55. Zeileis A (2009) ineq: Measuring Inequality, Concentration, and Poverty. R package version 0.2-9. Available at http://CRAN.R-project.org/package=ineq). Accessed January 9, 2013.

Supporting Information

Thomas et al. 10.1073/pnas.1222149110

SI Materials and Methods

Example of Calculation for Within-Mouse Nucleotide Sequence Sharing. As an example, for the D^bNP_{366}-specific amino acid sequence SGGANTGQL, the nucleotide sequence *agtgggggggccaacaccgggcagctc* is repeated twice in mouse M1, which is counted as 1 duplicate. The nucleotide sequence *agtgggggggcaaacaccgggcagctc* is repeated twice in mouse M5 and three times in mouse M6, which is counted as 1 + 2 duplicates. No other duplicates within a mouse occur for any of the other nucleotide sequences for this amino acid sequence. This leads to a total of 1 + 1 + 2 = 4 duplicates, divided by a total of 14 nucleotide sequences for this amino acid sequence, leading to a within-host sharing index for this amino acid sequence of 4/14.

Example of Calculation for Between-Mouse Nucleotide Sequence Sharing. Again, consider as example, the D^bNP_{366}-specific amino acid sequence, SGGANTGQL. For this aa sequence, the nucleotide sequence *agtgggggggcaaacaccgggcagctc* appears in four mice, which is counted as 3 duplicates. No other nucleotide sequence shows up in more than one mouse. This leads to a total of 3 between-mice duplicates, divided by a total of 14 nucleotide sequences for this amino acid sequence, leading to a between-host sharing index for this amino acid sequence of 3/14.

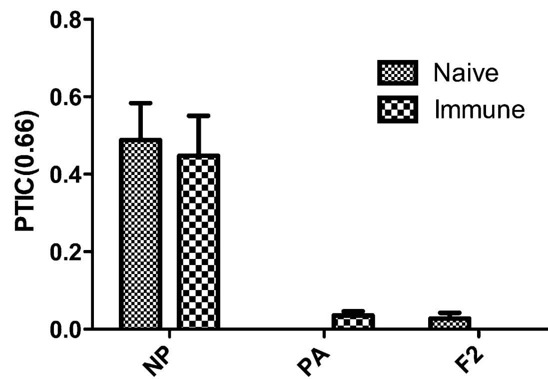

Fig. S1. The proportion of TCRs in common (PTIC) shows the proportion of the clonotypic response from individual mice that is shared by at least 66% of mice sampled. Shown is mean ± SD for sequence data described in Fig. 1.

Table S1. Nucleotide sequences for high frequency amino acid clonotypes in naïve D^bNP_{366}-, D^bPA_{224}-, and $D^bPB1–F2_{62}$-specific TCRβ repertoires

CDR3β	M1	M2	M3	M4	M5	M6
D^bNP_{366}						
SGGANTGQL	3	3		3	2	3
Agtgggggggccaacaccgggcagctc	*2*					
agtgggggggcaaacaccgggcagctc	*1*			*1*	*2*	*3*
agtgggggggcgaacaccgggcagctc				*1*		
agtggcggagcaaacaccgggcagctc				*1*		
Agtggaggggcgaacaccgggcagctc		*1*				
gtggaggggcaaacaccgggcagctc		*1*				
agtgggggcgcaaacaccgggcagctc		*1*				
SGGSNTGQL	1		1		1	
tctgggggtcaaacaccgggcagctc	*1*					
agtgggggtcaaacaccgggcagctc			*1*			
agtgggggttcaaacaccgggcagctc					*1*	
RGGANTGQL		1	1	1	1	1
cgggggggagcaaacaccgggcagctc		*1*				
aggggggcgcaaacaccgggcagctc			*1*			
agggggggggcaaacaccgggcagctc				*1*		
agggggggtgcaaacaccgggcagctc					*1*	
agaggggggcaaacaccgggcagctc						*1*
SGGGNTGQL	1					5
agtggggggggaacaccgggcagctc	*1*					*4*
agtgggggggcaacaccgggcagctc						*1*
KGGSNTGQL	2					
agcaaggggggtcaaacaccgggcagctc	*2*					
D^bPA_{224}						
SWGGEQ	2		1		1	
agttgggggggggaacag	*2*					
agctgggggggtgaacag			*1*			
agttggggtggggaacag					*1*	
SWGERL	2				1	
agttggggcgaaagatta	*2*				*1*	
SLGDEQ	1				1	1
agtttaggggatgaacag	*1*					*1*
agtttgggggatgaacag					*1*	
SLGAEQ		1		1		
Agtttgggggctgagcag		*1*				
agtttaggtgctgagcag				*1*		
SLGTEV		1		1		
agccttgggacagaagtc		*1*				
agtctcgggacagaagtc				*1*		
SSGEAP		1			1	
agttccggggaggctccg		*1*				
agttcgggggaggctccg					*1*	
SGGDEQ			1		1	
Tcgggggcgatgaacag			*1*			
tcaggggcgatgaacag					*1*	
SPDRGAL	2					
agtcccgacagggggggccttg	*2*					
TPGAEQ				3		
acccccggtgctgaacaa				*1*		
acccccggtgctgagcag				*2*		
SLGGYEQ				2		
agtttaggggggtatgaacag				*2*		
TGGERL				1		1
acgggggcgaaagatta				*1*		
acaggggcgaaagatta						*1*
SWGDEQ					2	
agttgggggatgaacag					*1*	
agttggggagatgaacag					*1*	
SDQ						4
agcgaccag						*4*
$D^bPB1\text{-}F2_{62}$						

Table S1. Cont.

CDR3β	M1	M2	M3	M4	M5	M6
SPGTANTEV	3					
agtcccgggacagcaaacacagaagtc	*1*					
agccccgggactgcaaacacagaagtc	*1*					
agccccgggacagcaaacacagaagtc	*1*					
SMGNTEV	2	2				
agtatggggaacacagaagtc	*1*	*1*				
tccatggggaacacagaagtc	*1*					
agtatgggaaacacagaagtc		*1*				
SPGQNTEV	2		1	2		
agcccgggacagaacacagaagtc	*1*					
agtccgggacaaaacacagaagtc	*1*					
agcccgggacaaaacacagaagtc			*1*	*2*		
SPGTTNTEV	2			2		
agtcccgggacaacaaacacagaagtc	*2*			*1*		
agccccgggacaacaaacacagaagtc				*1*		
SMGANTEV	2					
agtatgggggcaaacacagaagtc	*2*					
SPGTNTEV	2					
agccctgggacaaacacagaagtc	*2*					
SNWGTNTGQL	1	1				
agtaactgggggacaaacaccgggcagctc	*1*	*1*				
SPGTDTEV	1		1	2		
agtcccgggacagacacagaagtc	*1*					
agccccgggacagacacagaagtc			*1*			
agcccagggacagacacagaagtc				*2*		
SMGAQDTQ	1		1			
agtatgggggcccaagacacccag	*1*		*1*			
SMGTEV	1					1
agtatggggacagaagtc	*1*					
agtatgggcacagaagtc						*1*
SPGINTEV		1	1			
agtccagggatcaacacagaagtc		*1*				
agtccagggataaacacagaagtc			*1*			
SQGFTEV			2			
agtcagggtttcacagaagtc			*2*			
SRDISYNSPL			1	1		
agtcgggacatttcctataattcgcccctctac			*1*			
agtagggacatttcctataattcgcccctctac				*1*		
SAGTNSDY				2		
agcgccggaacaaactccgactac				*2*		
SPGLGGAETL				2		
agtccgggactgggggtgcagaaacgctg				*2*		
STGLGDTL				2		
agtacgggactgggggacaccttg				*2*		
SPGQSNTGQL					2	
agcccgggacagtcaaacaccgggcagctc					*2*	
SIGDWGGQNTL					2	
agtataggggactgggggggccaaaacaccttg					*2*	
SMGNANTEV					2	
agtatgggcaatgcaaacacagaagtc					*2*	
TSGTGDYAEQ					2	
acctccgggactggggactatgctgagcag					*2*	

CDR3β aa sequences are shown in capitals; nucleotide sequences encoding aa sequences are shown underneath in italics.