Article:
Goncalves, P, Ferrarini, M, Molina-Paris, C orcid.org/0000-0001-9828-6737 et al. (5 more authors) (2017) A new mechanism shapes the naïve CD8+ T cell repertoire: the selection for full diversity. Molecular Immunology, 85. pp. 66-80. ISSN 0161-5890

https://doi.org/10.1016/j.molimm.2017.01.026

A new mechanism shapes the naïve CD8$^+$ T cell repertoire: the selection for full diversity

Pedro Gonçalves [*,†,§], Marco Ferrarini [‡], Carmen Molina-Paris [‡], Grant Lythe [‡], Florence Vasseur [*,†], Annik Lim[*], Benedita Rocha [*,†] & Orly Azogui[†]

* Lymphocyte Population Biology Unit, CNRS URA 196, Institut Pasteur, Paris, France

† INSERM, U1151, CNRS, UMR8253, Faculté de Médecine Paris Descartes, Paris, France

‡ Department of Applied Mathematics, University of Leeds, Leeds LS29JT, UK

§ Present address: Innate Immunity Unit, INSERM, U668, Institut Pasteur, Paris, France

***Corresponding authors**: P. Gonçalves, Institut Pasteur, 25 Rue du Dr. Roux, 75015 Paris, France, tel: +33601080704, e-mail: pedro.goncalves@pasteur.fr; and B. Rocha, Institut Pasteur, 25 Rue du Dr. Roux, 75015 Paris, France; tel: +33145688582; fax: +33145688921; e-mail: benedita.rocha@inserm.fr

Abstract

During thymic T cell differentiation, TCR repertoires are shaped by negative, positive and agonist selection. In the thymus and in the periphery, repertoires are also shaped by strong inter-clonal and intra-clonal competition to survive death by neglect. Understanding the impact of these events on the T cell repertoire requires direct evaluation of TCR expression in peripheral naïve T cells. Several studies have evaluated TCR diversity, with contradictory results. Some of these studies had intrinsic technical limitations since they used material obtained from T cell pools, preventing the direct evaluation of clone sizes. Indeed with these approaches, identical TCRs may correspond to different cells expressing the same receptor, or to several amplicons from the same T cell. We here overcame this limitation by evaluating TCRB expression in individual naïve $CD8^+$ T cells. Of the 2,269 *Tcrb* sequences we obtained from 13 mice, 99% were unique. Mathematical analysis of this data showed that the average number of naïve peripheral $CD8^+$ T cells expressing the same TCRB is 1.1 cell. Since TCRA co-expression studies could only increase repertoire diversity, these results reveal that the number of naïve T cells with unique TCRs approaches the number of naïve cells. Since thymocytes undergo multiple rounds of divisions after TCRB rearrangement; and 3-5% of thymocytes survive thymic selection events; the number of cells expressing the same TCRB was expected to be much higher. Thus, these results suggest a new repertoire selection mechanism, which strongly selects for full TCRB diversity.

Keywords

$CD8^+$ T cells; mice TCR repertoires; CDR3 sequences; TCR diversity; TCR cross-reactivity.

Abbreviations

BM- bone marrow; CDR3- complementarity determining region; “clonotypes”- cells expressing identical *Tcrb* chains; HP- homeostatic proliferation; LCMV- Lymphocytic Choriomeningitis Virus; LN- lymph nodes; MHC- Major histocompatibility complex; MoAbs- monoclonal antibodies; Ms.- manuscript; SP- spleen; SPF- specific-pathogen-free mice; TCR- T cell receptor;

1. Introduction

The immune system is known to have Promethean properties, i.e., to be able to recognize all types of natural and artificial antigens introduced in the organism. It is yet subject of debate how much this remarkable capability depends on the diversity or on the cross-reactivity of peripheral T cell repertoires is still a subject of debate.

The antigen specificity of $CD8^+$ T cells is determined by a dimer of TCRB and TCRA chains, which binds peptides presented by the major histocompatibility class I complex (pMHC). The TCRB and TCRA chains have three regions of hypervariability, the complementarity determining regions (CDR). The CDR1 and CDR2 loops are encoded by the germline V gene segment, while the CDR3 loop is created by V(D)J recombination (Von Boehmer, 2004). The *Tcrb* rearrangements begin at the $CD44^{low}CD25^+CD4^-CD8^-TCR^-$ (triple negative 3-TN3) thymocyte differentiation stage by recombining one of each of 35 TRBV, 2 TRBD and 12 TRBJ genes in mice (Lefranc, 2001). A semi-random cleavage of the recombination hairpins intermediates results in nibbling at the V-D-J junctions. These events, and the further addition of N and P nucleotides, result in a major increase in CDR3 diversity. TCR diversity studies are often focused on *Tcrb* CDR3 region because this region is the most diverse portion of the TCR and functional/crystallographic analysis shows that the interaction between the pMHC complex and the TCR is predominantly mediated via this region (Das et al., 2015). The expression of an in-frame TCRB induces allelic exclusion, a burst of 6-8 cell

divisions (Kreslavsky et al., 2012; Penit et al., 1995; Penit and Vasseur, 1997; Von Boehmer, 2004), and the transition to the ISP (immature single positive) and the $CD4^+CD8^+$ (DP) thymocyte differentiation stages. Thymocytes divide at the ISP and at least 3 times at the DP stage (Kreslavsky et al., 2012).

A diversity of the peripheral T cell pools also depends on *Tcra* rearrangements and TCRB/TCRA pairing. Indeed, in DP cells *Tcra* chains are rearranged by recombining one of each of 132 TRAV and 60 TRAJ fragments genes in mice (Genolet et al., 2012; Lefranc, 2001). The same events occurring during TCRB V-D-J recombination also occur during *Tcra* V-J rearrangements, inducing a major variability of the *Tcra* CDR3 region. Because the first in-frame TCRA chain may not pair efficiently to the expressed TCRB chain, DP cells have the ability to rearrange multiple *Tcra* chains until a compatible TCRAB dimer is formed. In theory, these events could generate a potential repertoire of more than 10^{15} different TCRs (Von Boehmer, 2004).

It is estimated that in the mouse thymus around $5x10^7$ $TCR\alpha\beta^+$ thymocytes are generated each day (Shortman and Jackson, 1974; Von Boehmer, 2004). These cells undergo negative, positive and agonist selection, or may die by neglect. Lastly, thymic egress is also restricted (Von Boehmer, 2004). Each of these selection events, as well as peripheral T cell survival, is not solely dependent on the TCR-ligand interactions of each individual cell, but is strongly influenced by competition between different T cells (Freitas and Rocha, 2000; Hao et al., 2006). Because it cannot be deduced, in a non-manipulated mouse, from current understanding of these multiple selecting events, the diversity of peripheral T cell repertoires, this diversity must be measured directly.

Several studies addressed this issue, with contradictory results. In the mouse, it was claimed that 10% (Carey et al., 2016), 28% (Casrouge et al., 2000), 55% (Peaudecerf et al., 2012) or 68% (Quigley et al., 2010) of *Tcrb* chains were unique. In humans, the reported

number of unique TCRs ranges from 10^6 to $2x10^7$ (Arstila et al., 1999, 2000; Qi et al., 2014; Robins et al., 2009). These differences could be partially due to differences in the T cell populations studied, or/and by bias introduced by the methodology used to evaluate diversity (PCR amplification, followed by spectrotyping and cloning *versus* PCR amplification followed by next-generation sequence analysis). However, these approaches had common as well as specific technical limitations. None allows the evaluation of sequencing efficiency, i.e., to identify the number of T cells that had their TCR actually amplified. None can exclude bias introduced by primer competition or/and, in next-generation sequencing, by filters used to eliminate potential PCR errors. In most studies these selection filters are not defined. In one study where selected filters were fully described reported that up to 50% of the TCR sequences were eliminated (Nguyen et al., 2011). Apart from these biases, the successive preparation steps may reduce putative diversity by preferentially selecting more abundant TCRs. Importantly, bulk studies are unable to identify the number and the size of different clones. After PCR amplification, it is uncertain if identical TCR sequences correspond to multiple cells sharing the same TCR or to multiple amplicons from the TCR of a single cell. Lastly, in several cases it is not clear how representative the sample was, with respect to the total number of T cells belonging to the same population. To overcome these limitations, TCR expression must be determined in single cells.

Several recent studies used single-cell approaches to determine TCR expression of total naïve or naïve antigen-specific cells from non-immunized mice (Cukalac et al., 2015; Eltahla et al., 2016; Quinn et al., 2016; Stubbington et al., 2016). All these studies report a higher diversity of *Tcrb* expression than that determined by bulk studies. In particular, Quinn *et al* studied over 300 T cells (15-72 cells mouse) specific for a peptide of the influenza virus using tetramers (Quinn et al., 2016). They observed that this naïve repertoire is almost completely diverse but did not determine overall TCRB repertoires. Rigorous analysis of reper-

toire diversity will require studies in which multiple mice are studied and the number of cells sequenced in each mouse is representative of the total population from that mouse. Over more than a decade, we have developed and validated the parameters required for quantifying the expression of multiple mRNAs in single cells (Peixoto et al., 2004), including primer design and concentrations required to prevent primer competition and the conditions of amplification allowing the detection of as little as 2mRNA/cell, while preventing saturation. We now used this experience to develop a single-cell approach allowing evaluating the *Tcrb* expression in single-cells. Here, we describe the evaluation of the repertoire diversity in single $CD8^+$ T cells from specific-pathogen-free (SPF) adult mice. In contrast to bulk cell approaches, this single-cell analysis allows evaluation of the sequencing efficiency since we directly determined the number of cells where an in-frame *Tcrb* chain was sequenced. Our approach prevents primer competition, since a single primer pair is used for the PCR amplification of the *Tcrb* in each individual cell. It allows direct evaluation of PCR errors, by sequencing simultaneously the *Tcrb* of monoclonal TCR-Tg single-cells expressing known *Tcrb* chains. Of the 2,269 *Tcrb* chains we sequenced, 99% were unique. Mathematical analysis of representative samples indicate that, solely based on *Tcrb* expression, the average number of naïve $CD8^+$ T cells expressing the same *Tcrb* is 1.1 cells. This average "clonotype" size is unexpected, taking into account the number of divisions of TCRB expressing immature thymocytes. At the DN3 thymocyte differentiation stage, the expression of an in-frame TCRB induces allelic exclusion, a burst of 6-8 cell divisions (Kreslavsky et al., 2012; Penit et al., 1995; Penit and Vasseur, 1997; Von Boehmer, 2004), and the transition to the ISP (immature single positive) and the $CD4^+CD8^+$ (DP) thymocyte differentiation stages. Thymocytes divide at the ISP and at least 3 times at the DP stage (Kreslavsky et al., 2012). Since immature thymocytes undergo at least 11 divisions after *Tcrb* rearrangement (Kreslavsky et al., 2012; Penit et al., 1995; Penit and Vasseur, 1997; Von Boehmer, 2004), if all these divisions were productive up to 2^{11} cells ex-

pressing the same *Tcrb* could be generated. Even if only 3-5% survive thymus selection events (Huesmann et al., 1991), an average 62 to 102 cells should express the same *Tcrb*. Therefore, our results suggest that, superimposed on the known mechanisms of repertoire selection (negative, positive, agonist and death by neglect), a remarkable selection for full TCRB diversity also occurs.

2. Material and methods

2.1. Mice

Specific-pathogen-free C57BL/6 (B6) mice expressing the CD45.2 allotype marker, and CD45.1 $Rag2^{-/-}$ P14 (P14) mice expressing a transgenic TCR specific for LCMV epitope GP_{33-41} (GP33) backcrossed onto the $Rag2^{-/-}$ C57BL/6 (B6) background, were obtained from our breeding colonies at the Centre de Distribution, Typage et Archivage (CDTA, Orleans, France). All animal experiments were performed in accordance with National and European Commission guidelines for the care and handling of laboratory animals and were approved by the site ethical review committee.

2.2. The evaluation of T cell numbers in each $CD8^+$ T cell subpopulation.

To eliminate blood-derived T cells, SPF B6 mice ($CD45.2^+$) were exsanguinated before organ removal. From each mouse the spleen (SP), 2 femurs and 40 lymph nodes (LN) were removed from each mouse. The identification of these LN was performed as described previously (Anjuere et al., 1999; Sung et al., 2013; Van den Broeck et al., 2006). Briefly, a control mouse was injected with china ink diluted in PBS. The ink particles are phagocytized by the LN macrophages, in such a way that each LN acquires a black color and becomes easily visualized. This mouse was used as a reference to identify the LNs in experimental

mice, which were not injected with china ink. The five small LN known to be located in the thorax were not identified because they were obscured by blood loss into the mediastinum.

Using an inverted microscope, organs were totally cleaned of fat and other adjoining tissues and distributed in 24-well plates in RPMI medium supplemented with 2% fetal calf serum and HEPES buffer, together with $0.5x10^6$ Monoclonal $CD45.1^+$ P14 cells, as a "reference population" that is crucial for carrying out accurate cell counts. The cells of the reference population undergo the same preparation steps as $CD8^+$ T cells, allowing us to determine non-specific cell loss during preparation steps (average 50%). Cell suspensions were obtained by mechanical disruption with forceps followed by digestion with 0.5 mg/ml collagenase type IV (Worthington Biochemical Corporation, Lakewood, NJ, USA) and 5 µg/ml deoxyribonuclease I (Sigma-Aldrich, St. Louis, MN, USA) for 30 min at 37°C in 5% CO_2 with agitation. We found these steps critical, since cell yields were much higher and the cell suspensions cleaner when compared with those obtained by mechanical disruption alone. Femoral bone marrow was extracted by inserting a syringe equipped with a 26-gauge needle into one end of the bone and flushing with 3 to 4 mL RPMI containing 1% FCS. The total number of BM cells was obtained by multiplying the number of cells recovered from two femurs by 7.9, as previously described (Slifka et al., 1995). We did not study blood cells, for ethical restrictions, since mice must be kept alive during blood removal.

For depletion of non-$CD8^+$ T cells, cell suspensions were labeled with a cocktail of biotin-conjugated monoclonal antibodies (MoAbs) from BD Pharmingen (San Diego, CA, USA) (anti-TER119, CD19, Mac-1, GR1, CD4, B220) and anti-biotin Dynabeads (Dynal AS, Oslo, Norway) following the manufacturer's instructions. All these MoAbs were previously titrated to determine the binding efficiency and the absence of non-specific binding/depletion.

The combination of pMHC dextramers and protein kinase inhibition increases the range of TCR-pMHC interactions and dextramer sensitivity allowing the detection by pMHC

multimers of cognate T cells with low TCR avidity (Lissina et al., 2009). $CD8^+$ T cell enriched suspensions were incubated with 50 nM dasatinib (30 min, 37 °C) (Lissina et al., 2009), and labeled for 20 min at room temperature with APC or PE-labeled dextramers of H-$2D^b$ pMHC I loaded with GP33 peptide or a general negative control dextramer for accessing the unspecific and background staining (Dextramers®, Immudex, Copenhagen, Denmark, previously titrated on P14 Tg cells). This was followed by incubation at 4°C during 30 min with the following MoAbs obtained from BD Pharmingen (San Diego, CA, USA): anti-CD45.1 (A20) PECy7, anti-CD45.2 (104) PerCP, anti-CD3ε (145-2C11) Brilliant Violet 510, anti-CD8β (H35-172) Pacific Blue, anti-CD44 (1M781) APC-eFluor 780, anti-Vb7 (TR310) PE and anti-Vb8 (F23.1) PerCP-Cy5.5. Dead cells were excluded by Sytox Green dead cell stain (Thermo Fisher Scientific, MA USA). For cell analysis and counting, the labeled populations were diluted in 0.5 ml of FACS flow buffer and acquired using the low-speed mode in a FACSCanto II flow cytometer (Becton Dickinson, Franklin Lakes, NJ, USA). The use of low-speed mode was important, since it reduces both the cell loss during acquisition and the background non-specific labeling. Although very time-consuming, in our hands it is the best method to visualize rare cells clearly. Gates identifying dextramer labeled cells were identified in P14 TCR-Tg cells, labeled with dextramers. The data analysis was performed using FlowJo software (TreeStar, Ashland, OR).

The methodology we use to count $GP33^+$ cells is precise, but also laborious and time consuming. In general, we can only study one individual mouse per day, to recover the number of LNs we studied, and to acquire data in diluted samples at the low speed mode. Therefore it is not possible to count cells accurately, and to sort $GP33^+$ populations in the same experiment. For single-cell sorting we used the same method described above, but LN cells were sorted from a smaller pool of LNs (brachial, inguinal, axillary and mesenteric), with the gating strategy shown in Fig.1. Single cells were sorted (FACS-Aria II system,

Becton Dickinson, Franklin Lakes, NJ, USA) into 96-well PCR plates (purity > 99%) that were immediately stored at -80 °C until required.

2.3. Single-cell cDNA synthesis, nested RT-PCR and sequencing

Analysis of *Tcrb* usage was carried out by a single-cell multiplex RT-PCR (Dash et al., 2011) developed in our laboratory (Peixoto et al., 2004), followed by the direct sequencing of the PCR products. For that purpose, we designed 21 TRBV external primers (Supplemental Table 1), which cover the entire repertoire of functional TRBV genes. We validated that each TRBV primer only amplified a single TRBV chain, i.e., all primers were devoid of cross-reactivity, and did not compete with one another, which allowed us to use them in multiplex RT-PCR. The 3' primers were specific of the TRBC region. Briefly, individual cells were incubated at: (i) 42°C for 30 min to synthesize cDNA using specific external primers for *Tcrbv* and *Tcrbc* in 5 µl of reaction (ii) a first 35-cycle amplification step using 21 *Tcrbv* external forward primers; 94°C for 5 min to melt; 5 cycles of 94°C for 30 sec, 68°C for 20 sec and 72 °C for 1 min; 35 cycles of 94°C for 30 sec, 58°C for 20 sec and 72 °C for 1 min; 72°C for 1 min to complete extension in 30 µl of reaction (iii); and then a second, nested, PCR was performed using 1 µl of the first-round product in 10 µl PCR reactions, with another set of specific, validated, *Tcrbv* internal primers (Supplemental Table 1). The following parameters were used: 94°C for 5 min to melt; 5 cycles of 94°C for 30 sec, 68°C for 20 sec and 72 °C for 1 min; 45 cycles of 94°C for 30 sec, 58°C for 20 sec and 72 °C for 1 min; 72°C for 1 min to complete extension. Wells with successful amplifications were identified by migration of a sample of the second PCR reaction on 2% agarose gel. When a band was present, 3 µl of the second PCR product was treated with 1 µl of ExoSAP-IT (usb) in 7 µl reaction mixture at 37 °C for 40 minutes and subsequently at 80°C for 20 minutes. The cDNA-PCR products were sequenced directly in 12 µl reaction mixture of 7 µl purified cDNA, 3 µM specific primers,

and 0,4 µl BigDye™ Terminator v1.1 cycle sequencing kit (Applied Biosystems). The RT-PCR products were purified using PCR purification columns (QIAGEN) according to the manufacturer's instructions and sequenced by the Sanger reaction. Sequences were analyzed according to the ImMunoGeneTics/V-QUEry and STandardization web-based tool (http://imgt.cines.fr). All TCRB nomenclature was according to Bosc and Lefranc (Bosc and Lefranc, 2000). The CDR3 amino acid sequence region begins with the second conserved cysteine encoded by the 3' portion of the TRBV gene segment and ends with the conserved phenylalanine encoded by the 5' portion of the TRBJ gene segment. The number of nucleotides between these codons determines the length and the frame of the CDR3 region. It must be noted that in all experiments, we tested both for possible contaminations and PCR errors influencing our results. To detect possible contaminations, for each 8 tubes containing one cell, we amplified an additional control empty tube, with both the first and the second PCR reaction. To screen for putative PCR errors, 8-16 single cells from Mo TCR-Tg P14 mice were also sequenced in each individual experiment. In all experiments, we sequenced the *Tcrb* of 184 individual P14 Monoclonal TCR-Tg cells. All these 184 single-cells expressed the same *tcrb*, confirming the absence of PCR errors, as expected by the reduced size of the amplicons.

2.4. Statistical analysis

Statistics were performed using Prism 5, GraphPad software (San Diego, USA). Statistical significance of the difference between two groups was evaluated by the Student's t-test. Differences were considered to be significant when $p < 0.05$.

3. Results

3.1. Quantification of the naïve CD8⁺ T cell pools in SPF mice.

First we determined the precise number of $CD8^+$ T cells in different lymphoid organs of specific-pathogen-free 11-12 weeks old C57BL/6 (B6) mice, using a strategy described previously (Sung et al., 2013). Briefly, a known number of naïve P14 $CD8^+CD45.1^+$ T cells were added to the Petrie dish where $CD45.2^+$ cell suspensions from each organ were going to be prepared. This "reference population" undergoes the same preparation steps as the $CD8^+$ T cells from that organ, allowing evaluation of the proportion of cells lost during washes and $CD8^+$ purification steps. We previously showed that loss-rates were identical for different $CD8^+$ "reference" populations, independently of their initial number (Sung et al., 2013). An extra procedure was carried out for bone marrow (BM) $CD8^+$ T lymphocytes: since the BM is highly vascularized leading to the contamination of BM cell suspensions by blood cells, these suspensions were labeled with anti-CXCR4Abs that identify BM resident cells. CXCR4 and its ligand CXCL12 mediate the homing of $CD8^+$ T cells in the BM (Chaix et al., 2014). CXCR4 signalling is essential to T cell retention in BM (Itkin et al., 2016; Petit et al., 2002) and only $CXCR4^+$ cells were considered for BM counts. CD69 was also essential for the persistence of memory T cells in the BM environment (Shinoda et al., 2012) and most of these $CD8^+$ T cells express CD69 (results not shown). After $CD8^+$ T cell enrichment, cells were labeled with CD45.1, CD45.2, CD44, CD8β and CD3ε. The total number of $CD45.2^+$ cells was calculated based on the recovery of the $CD45.1^+$ "reference population", and $CD8^+$ T cell cells failing to express CD44 were considered to be naïve. The separation into $CD44^-$ and $CD44^+$ cells was based in the fluorescence minus one (FMO) of cells labeled with an isotype control (Fig. 1).

Different studies reported that the lymphoid organs of an adult non-infected mouse contain ~ $2\text{-}5x10^7$ $CD8^+$ T cells, but these numbers were not estimated directly (Casrouge et

al., 2000; Jenkins et al., 2010; Pewe et al., 2004). Our direct counts of the $CD8^+$ T cells in SPF B6 mice are closest to the lowest of the previous extrapolations. We counted an average of $2x10^7$ total $CD8^+$ T cells in the total peripheral lymphoid organs i.e, SP, total LN and BM pools (Table 1). The total naïve $CD44^- CD8^+$ T cell pool averaged $7x10^6$ cells, the remaining $13x10^6$ expressing CD44.

3.2. Evaluation of Tcrb expression by individual $CD8^+$ T cells.

For the validation of our single-cell method, we first studied a cohort of $CD8^+$ T cells isolated from the SP and 5 LN (Table 2, Mouse 4-M4). The sequencing efficiency, i.e., the number of the single-cells in which we sequenced an in-frame *Tcrb* chain, was 89.2%. 187 of the 188 single-cells expressed unique *Tcrb* chains (Table 2, Supplemental Table 2). To determine if the single repeat we found in this cohort corresponded to two $CD44^+$ cells, we studied the $CD44^-CD8^+$ populations from three further mice. In these mice, sequencing efficiency ranged from 85-88% (Table 2, M5-M7). We detect all *Tcrbv* and *Tcrbj* genes (Fig. 2-A, B. Supplemental Table 2). The relative distribution of cells expressing each *Tcrbv* was consistent with that described using TCRVB specific Abs (Kato et al., 1994). Those expressing each *Tcrbj* were as described previously (Candeias et al., 1991). As expected in polyclonal repertoires, CDR3 lengths followed a Gaussian distribution, from 7-16 amino acids (aa) (Fig. 2C). This data indicates the efficiency of our methodology: we amplified all *Tcrbv* and *Tcrbj* at the expected frequencies, and our sequencing efficiency was high. To determine if PCR errors could influence our results, in each experiment we also sequenced the *Tcrb* expressed by multiple single-cells from Monoclonal (Mo) P14 TCR-Tg mice. All the 184 MoP14 Tg cells we studied in different experiments expressed the same Tg *Tcrb*, excluding the possibility that PCR errors influenced our results (Supplemental Table 2).

The present data also allowed us to determine other characteristics of the *Tcrb* rearrangements. We found that 2.9% of individual cells expressed two in-frame *Tcrb* chains, (Supplemental Table 2) confirming that *Tcrb* allelic exclusion is not absolute (Stubbington et al., 2016). $CD8^+$ T cells do not express the different TCRVB and the *Tcrbj* genes at the same frequency, but it is not known if this is due to non-random recombination or to selection events (Kreslavsky et al., 2012; Von Boehmer, 2004). We compared *Tcrbv* usage in in-frame and out-of-frame rearrangements to clarify this issue. If higher TCRVB usage is due to preferential rearrangements, *Tcrbv* usage by in-frame (selected) and out-of-frame (non-selected) rearrangements should be similar. Alternatively, selection events could result in different *Tcrbv* usage by out-of-frame and in-frame *Tcrbv* molecules. We found that two thirds of *Tcrbv* molecules were expressed at the same frequency in in-frame and out-of-frame rearrangements (Fig. 3, Supplemental Table 2). However, *Tcrbv* 13-2, 14, 26 and 29 were more abundant in in-frame than out-of-frame rearrangements, indicating that $CD8^+$ T cells expressing these *Tcrbv* were enriched during ontogeny. On the other hand, the frequency of *Tcrbv* 4, 15 and 24 was higher in out-of-frame rearrangements, indicating that cells expressing these *Tcrbv* were counter-selected during ontogeny. In-frame and out-of-frame rearrangements showed the same frequency of *Tcrbj* usage (Supplemental Table 2). We conclude that the differences in the TCRBV distribution found in the peripheral $CD8^+$ T cell pool are mostly due to preferential *Tcrbv* rearrangements. However, the expression of certain TCRBV is modified by selection events. Cells expressing these *Tcrbv* will be referred to as selected (S).

Concerning the composition of the CDR3 region, all sequences we studied had nibbling at the V-D-J junctions and 90% also had N additions, indicating that they were generated after birth (Carlsson and Holmberg, 1990). These results suggest that T cells generated during the fetal/perinatal period only represent a small fraction of the peripheral T cell repertoire. With respect to variability, cells expressing identical *Tcrb* chains were very

rare (Table 2, Supplemental Table 2). In the four mice we studied, we did not find public sequences shared between different mice (Supplemental Table 2). These results contrast with previous reports using pooled cDNA or DNA from $CD8^+$ populations in mice or man, which reported a much lower diversity (Arstila et al., 2000; Qi et al., 2014; Robins et al., 2009).

Since the $CD44^-CD8^+$ samples that we analyzed represented an average of $3.8x10^{-3}$ % of the total $CD44^-CD8^+$ naïve T cell pool, we wished to study a population with a greater coverage. For that purpose, we first studied $CD44^-CD8^+$ T cells expressing either TCRBV13 or TCRBV19, the number of sequences analyzed representing respectively $2.1x10^{-2}$ and $5.7x10^{-2}$ % of total repertoire. However, 99.3% of $VB19^+CD44^-CD8^+$ T cells still expressed unique *Tcrb* sequences (Table 2, Supplemental Table 2). $TCRVB13^+$ cells (which are enriched during selection -S cells- Fig. 3) had a slightly reduced diversity, since 97% of these sequences were unique (Supplemental Table 2).

3.3. Evaluation of Tcrb expression by individual GP33-specific $CD8^+$ T cells.

To further select more representative samples, we studied $CD8^+$ T cells recognizing the immune-dominant GP_{33-43} peptide from the Lymphocytic Choriomeningitis Virus (LCMV) ($GP33^+$ $CD8^+$ T cells). First we determined the precise number of these cells in different lymphoid organs of SPF B6 mice (Fig. 1). We observed that the average total number (SP + LN + BM) of $CD44^-GP33^+$ cells was 2,734 ± 208, $CD44^{int}GP33^+$ was 6,658 ± 677 and $CD44^{high}GP33^+$ was 5,852 ± 401. Therefore, the majority of $GP33^+$ $CD8^+$ T cells did not have the $CD44^-$ naïve phenotype. $CD44^{high}GP33^+$ cells were particularly abundant in the BM of naïve mice (Fig. 1). We studied also a sub-dominant LCMV epitope (GP276) and another unrelated antigen, ovalbumin (OVA). As expected the total number of $GP276^+$ (9,530 ± 1,572) were less abundant than $GP33^+$ (15,244 ± 1,155), and OVA^+ cells were also less abundant (9,097 ± 1,432) (results not shown).

Next we studied the TCR characteristics of the $GP33^+$-specific repertoire. Previous analysis of TCR repertoires specific for the GP33 epitope showed a preferential usage of TRVB13 (Blattman et al., 2000; Lin and Welsh, 1998). We also observed a dominant bias toward TRBV13, up to 39% of $CD44^-GP33^+$ cells expressing TRBV13 (Figure 4A). By contrast, the expression frequency of TRBV29 was reduced from an average of 12% in total naïve cells to 5% in $CD44^-GP33^+$ cells. Small modifications in TRBV12-2 and TRBV3 expression frequency were also found. We found that the distribution of TRBJ genes is identical in both pools (results not shown). We stained $CD8^+$ T cells with specific antibodies for TRBV13 (anti-Vb8) and TRBV29 (anti-Vb7) in additional mice, confirming our results derived from TCRB sequence analysis (Figure 4B, C). Of note, these *Tcrb* distributions were common to $GP33^+$ cells from all lymphoid organs, including the rare $CD44^-GP33^+$ cells present in the BM (results not shown).

Our study of *Tcrb* expression by $CD44^-GP33^+$ cells produced samples that are much more representative, corresponding to 7-21% of the total $CD44^-GP33^+$ pool in each mouse (Table 3). However, as usual in the sorting of rare cells, sequencing efficiency decreased (from an average of 90% to 70%) but was still high. The study of antigen-specific cells could reduce variability, since naïve T cells recognizing the same epitope could be more likely to share identical *Tcrb* chains. However, the sharing of such identical *Tcrb* chains between antigen-specific cells was mostly been reported after pathogen infection and not in naïve T cell repertoires (Kedzierska et al., 2004). Indeed, the variability of $CD44^-GP33^+CD8^+$ T cells we find was 98-100%, i.e., similar to that found in the total naïve T cell pool (Table 3, Supplemental Table 3).

Cells from different individuals may share identical TCRB chains. These public TCRs were previously reported in mouse (Kedzierska et al., 2004). Some public TCRs can be generated from a near-germline V-D-J recombination, with no or minimal random template

nibbling or nucleotide additions. We did not find this type of public *Tcrb* sequences in the CD44^{-}GP33^{+} repertoire. Others can be generated by "convergent recombination", individual cells expressing different *Tcrb* expressing a TCRB protein with the same amino-acid composition. A single CASSDWGRDTLYFG TCRB was shared between mouse 1 and 2, but was absent in other mice. In these samples, from these 6 different mice we found a single sequence equivalent to the P14 TCRB chain in mouse 1 (Supplemental Table 3), indicating that this Tg-TCR is not particularly frequent in the naïve LCMV-specific pool. We conclude that virtually all naïve GP33^{+} TCRB repertoires are private and diverse.

The majority of the GP33^{+} CD8^{+} T cells in SPF young adult mice were not naïve but rather expressed CD44. If CD8^{+} T cells from B6 mice only acquire CD44 expression after antigen stimulation (Hao et al., 2006), and these CD44^{high} cells are known to behave as memory cells (Freitas and Rocha, 2000; Haluszczak et al., 2009), then it must be concluded that the majority of GP33^{high} cells in non-infected SPF mice are antigen-experienced cross-reactive cells (Fig. 1, Table 1). They may be generated by the homeostatic proliferation that occurs immediately after birth, when the first naïve T cells leave the thymus to seed the peripheral "empty" pool (Kieper and Jameson, 1999). However, 80% of CD44^{high}GP33^{+} *Tcrbs* had N additions (Table 4, Supplemental Table 3). This percentage is lower than that in CD44^{-}GP33^{+} T cells (90%) indicating that some degree of perinatal homeostatic proliferation contributes to the CD44^{high}GP33^{+} population. However, the majority (80%) is not generated during the perinatal period. Immune responses to self or environmental antigens (microbiota and food) are conjectured to generate these "mock" memory cells (Freitas and Rocha, 2000; Jameson et al., 2015; Su and Davis, 2013). Surprisingly, the repertoire of CD44^{high} cells was also diverse (Table 4, Supplemental Table 3), indicating that selection pressures for diversity are also occurring in the CD44^{high} pool (Quinn et al., 2016).

3.4. Mathematical and computational determination of Tcrb diversity.

We undertook a mathematical and computational study of the statistical properties of the sequences of samples of cells taken at random, under different hypotheses for the distribution of "clonotype" sizes in the naïve $CD8^+$ T cell repertoire, which are detailed in bellow and in a previous manuscript (Lythe et al., 2016).

We consider, from a general viewpoint, sampling from a repertoire containing a total of *S* cells that are shared among *N Tcrb* "clonotypes". That is, *N* is equal to the total number of distinct *Tcrb* sequences in the repertoire. We use the letter *i* to denote a "clonotype" in the repertoire that consists of n_i cells. Thus, we have $i=1,2,\ldots,N$ and $n_1+n_2+\ldots+n_N=S$. The mean clonal size is denoted by $\acute{n}$. It is equal to S/N, the mean number of cells per "clonotype". Three types of hypothesis we considered are as follows:

i. That each individual "clonotype" has the same number of cells;
ii. That the "clonotype" sizes follow a simple distribution, for example the geometric distribution where (according to the data) there are more "clonotypes" with small size than large clones;
iii. That there are two types of "clonotypes" in the repertoire, the majority represented by one cell and a few made up of only few cells.
iv. That there are two types of "clonotypes" in the repertoire, the majority represented by one cell and a small minority of "clonotypes" that contain many cells.

Suppose that a sample of *m* cells is taken and the *Tcrb* of each of the cells is sequenced. We define *q* to be the probability that one cell, randomly chosen from the total of *S* cells, is part of the sample of size *m*:

$q=\frac{m}{S}$. Let m_0 be the number of distinct sequences in the sample, and let m_1 be the number of sequences found only once in the sample. If $m_2, m_3, \ldots$ is the number of sequences found twice, three times, ... then

$$m_0 = m_1 + m_2 + m_3 + \ldots \quad \text{and} \quad m = m_1 + 2m_2 + 3m_3 + \ldots .$$

Under hypothesis (i), the mean of the ratio m_0/m can be written

$$\frac{\acute{m}_0}{m} = 1 - \frac{1}{2} q(n-1) + O\left((q(n-1))^2\right). \quad [1]$$

3.4.1. Estimating the mean clonal size of the CD44$^-$GP33$^+$ subset

We concentrated on the GP33$^+$ subset, since each of the 9 samples we collected from different mice represent almost 10 percent of the total repertoire of GP33$^+$CD8$^+$ T cells. Here, the value of *S* is the total number of CD44$^-$GP33$^+$ cells, estimated to be 441 (BM) or 2293 (SP+LN). Thus, with sample size *m* between 94 and 271, the value of *q* is between 0.04 and 0.12. Hypothesis (i) is not consistent with the data: if $n = 1$ then *m* is always equal to m_0 ; if $n = 2$ or larger, the predicted values of the ratio $\frac{m}{m_0}$ are larger than those observed. We evaluated if the data was compatible with other hypothesis.

Geometric distribution. We first consider the geometric distribution of values of numbers of cells per clone, n_i . The statement that n_i has a geometric distribution with mean $\acute{n}$ is that

$$P(n_i = k) = \frac{1}{\acute{n}}\left(1 - \frac{1}{\acute{n}}\right)^{k-1}, \qquad k = 1, 2, \ldots$$

Note that $\acute{n} \geq 1$. The fraction of clones that consists of only one cell is

$$P(n_i = 1) = \frac{1}{\acute{n}}.$$

If the distribution of values of n_i is geometric, then the distribution of the number of copies of each *Tcrb* sequence found in a sample of *m* cells is also geometric, with mean equal to

$1+(\acute{n}-1)q$. That is, the mean of the ratio $m/\ m_0$ is $1+(\acute{n}-1)q$. Because the values of *S* and *m* are known, we obtain one estimate of $\acute{n}$ from each measured value of m_0 :

$$\acute{n}=1+S\left(\frac{1}{m_0}-\frac{1}{m}\right).[2]$$

We use **[ii]** to estimate $\acute{n}$ in the CD44$^-$GP33$^+$ repertoire . There are six independent measurements:

- Mouse 5 (SP and LN, S=2293): 271 sequences, 268 unique, so estimate $\acute{n}=1.09$.
- Mouse 6 (SP and LN, S=2293): 188 sequences, 186 unique, so estimate $\acute{n}=1.13$.
- Mouse 7 (SP and LN, S=2293): 128 sequences, 127 unique, so estimate $\acute{n}=1.14$.
- Mouse 10 (SP and LN, S=2293): 244 sequences, 240 unique, so estimate $\acute{n}=1.16$.
- Mouse 11 (SP and LN, S=2293): 165 sequences, 165 unique, so estimate $\acute{n}=1.00$.
- Mouse 12 (BM, S=441): 94 sequences, 93 unique, so estimate $\acute{n}=1.05$.

The mean of the estimated values of $\acute{n}$ is 1.10, with standard deviation 0.06.

Poisson. As a check that our estimate of a mean clonal size not much larger than one is not due to a particularity of the geometric distribution, we now consider the hypothesis that the number of cells per clone, in the repertoire, has a Poisson distribution. The statement that n_i has a positive Poisson distribution with mean $\acute{n}$ is that

$$P(n_i=k)=\frac{1}{e^{\lambda}-1}\frac{\lambda^k}{k!}k=1,2,\ldots,where\,\acute{n}=\frac{\lambda e^{\lambda}}{e^{\lambda}-1}.[3]$$

In this case, the distribution of the number of copies y_i of each *Tcrb* sequence found in a sample of *m* cells is also positive Poisson, with

$$P(y_i=k)=\frac{1}{e^{\lambda q}-1}\frac{(\lambda q)^k}{k!}k=1,2,\ldots$$

The mean value of $\frac{m}{m_0}$ is $\frac{\lambda q e^{\lambda q}}{e^{\lambda q}-1}$ which, because $\lambda q \ll 1,$ we can expand as

$1+\frac{1}{2}\lambda q+\frac{1}{4}(\lambda q)^2+\ldots.$ Retaining up to first order in λq, gives

$$\lambda = 2S\left(\frac{1}{m_0} - \frac{1}{m}\right). [4]$$

For each mouse, we estimate λ using **[6]**, then calculate $\acute{n}$ using **[4]**:

- Mouse 5 (SP and LN, S=2293): 271 sequences, 268 unique, so estimate $\acute{n} = 1.10$.
- Mouse 6 (SP and LN, S=2293): 188 sequences, 186 unique, so estimate $\acute{n} = 1.14$.
- Mouse 7 (SP and LN, S=2293): 128 sequences, 127 unique, so estimate $\acute{n} = 1.15$.
- Mouse 10 (SP and LN, S=2293): 244 sequences, 240 unique, so estimate $\acute{n} = 1.17$.
- Mouse 11 (SP and LN, S=2293): 165 sequences, 165 unique, so estimate $\acute{n} = 1.00$.
- Mouse 12 (BM, S=441): 94 sequences, 93 unique, so estimate $\acute{n} = 1.05$.

The mean of the estimated values of $\acute{n}$ is 1.10, with standard deviation 0.06.

Thus, these two different mathematical approaches give similar estimates of 1.1 as the mean number of cells per "clonotype".

3.4.2. Estimating the mean clonal size of the CD44^{high}GP33^{+} subset

Table IV summarizes data from single-cell *Tcrb* sequencing of CD44^{high}GP33^{+} cells:

- Mouse 14 (BM, S=2178): 99 sequences, 99 unique, so estimate $\acute{n} < 1.10$.
- Mouse 15 (BM, S=2178): 71 sequences, 71 unique, so estimate $\acute{n} < 1.20$.
- Mouse 16 (BM, S=2178): 69 sequences, 69 unique, so estimate $\acute{n} < 1.20$.

Because no repeats were found, the estimates of $\acute{n}$ are upper limits.

3.4.3. Estimating the mean clonal size of the PB1-F262 subset

Quinn *et al.* studied the number and phenotype of influenza A virus-specific CD8^{+} T cells, selected using tetramers, in B6 mice (Quinn et al., 2016). They performed single-cell *Tcra* and *Tcrb* sequencing of one such population, specific for D^{b}PB1-F262 influenza A virus peptide. Consistent with our results, they observed that the naïve repertoire was almost completely diverse, with only a small number of repeated *Tcrb* sequences. Here, we take their published *Tcrb* chain sequencing data (Dataset S1) and obtain estimates of the mean number of cells per clonotype, using **[2]** with S=150.

- Mouse 1: 30 sequences, 28 unique, estimate $\acute{n} = 1.36$.
- Mouse 2: 28 sequences, 28 unique, estimate $\acute{n} = 1.00$.
- Mouse 3: 11 sequences, 11 unique, estimate $\acute{n} = 1.00$.
- Mouse 4: 16 sequences, 16 unique, estimate $\acute{n} = 1.00$.
- Mouse 5: 26 sequences, 24 unique, estimate $\acute{n} = 1.48$.
- Mouse 6: 20 sequences, 20 unique, estimate $\acute{n} = 1.00$.
- Mouse 7: 74 sequences, 72 unique, estimate $\acute{n} = 1.06$.
- Mouse 8: 37 sequences, 37 unique, estimate $\acute{n} = 1.00$.
- Mouse 9: 47 sequences, 44 unique, estimate $\acute{n} = 1.22$.
- Mouse 10: 38 sequences, 37 unique, estimate $\acute{n} = 1.11$.

3.4.4. Clonotypes with two cells

In the GP33^{+} subset, we could ascribe the occurrence of repeated *Tcrb* sequences in our data to the existence of clonotypes with two cells in the repertoire. Can the same explanation hold for the full naïve repertoire?

There are N clonotypes in the repertoire of S cells. We denote the number of clonotypes consisting of one, two, three … cells by N_1, N_2, N_3, …. That is,

$$N = N_1 + N_2 + N_3 + \ldots$$

and

$$S = N_1 + 2N_2 + 3N_3 + \ldots$$

The probability that both cells of a clonotype that has two cells in the repertoire are found in a sample of m cells is

$$r_2 = m(m-1)/S(S-1).$$

Since there are N_2 such clonotypes, the mean number of pairs of cells in the sample that are the only two cells of their type in the repertoire is

$R^2 = N_2 r_2 \sim N_2\, m^2/S^2$.

In order for there to be, on average, one such repeated *Tcrb* per sample of m cells, we need $N_2 \sim (S/m)^2$.

Antigen-specific repertoire: m=100 and S=2300. We need N2=23^2 = 529, which is possible. For example, the data from GP33$^+$ subset could be attributed to 1242 cells with unique *Tcrb* and 529 doublets.

3.4.5. The naïve CD8$^+$ repertoire

Tables II, III and IV summarize single-cell sequence data from the naïve CD8$^+$ repertoire (M4) and a variety of repertoire subsets (M5-M16). The values of S range from 0.27×10^6 (M9) to 17×10^6 (M4), with values of m between 83 and 188 and, hence, values of q are between 0.005 and 0.00001. The mean clonal size in the naïve CD8$^+$ repertoire is therefore the same as that of the GP33$^+$ subset (not shown). However, hypothesis (iii) is not compatible with the overall naïve CD8$^+$ repertoire: m=100 and S=10^6. We need N2=10^{10}, which is impossible. Finding repeated *Tcrb* sequences in average once per sample of 100 cells from a repertoire of 1 million cells cannot be due to "clonotypes" in the naïve repertoire with two cells.

In the case of the naïve CD8$^+$ repertoire, and the subsets restricted to TRBV13 and TRBV19 expression, the detection of one or two repeated sequences per sample is only consistent with the hypothesis (iv) that rare large "clonotytpes" exist in the naïve repertoire. It must be noted that the majority of the repeated sequences found in CD8$^+$ T cell repertoires express the selected *Tcrbv*, i.e., those *Tcrbv* that are more abundant in in-frame than out-of-frame rearrangements.

We conclude that by using different mathematical approaches, and by studying different samples collected from up to 13 mice as well as the data from (Quinn et al., 2016) collected from 10 mice, the average clonotype size of peripheral CD8$^+$ T lymphocytes is 1.1 cell. These results show that taking into consideration only *Tcrbv* diversity, the number of unique TCRB chains approaches the number of naïve T cells.

4. Discussion

When we initiated this work, we aimed to study TCR diversity by co-amplifying both the *Tcrb* and the *Tcra* chains expressed by each individual cell. The present results show that such complex study is not justified, since it would not modify considerably the estimations of the total diversity of the naïve $CD8^+$ T cell pool, based on the current single-cell *Tcrb* chain expression study. Of the 2,269 in-frame *Tcrb* chains we sequenced in individual cells, 2,248 (99%) were unique. This extensive diversity was not a property of a single sample. It was shared by all samples we studied from 13 individual mice; it was present in samples from total $CD8^+$ T cells, from $CD8^+$ T cells expressing a single TCRVB, and even from naïve $CD44^-GP33^+$ and $CD44^+GP33^+$ $CD8^+$ T lymphocytes. Overall, the mathematical analysis of our results show that the average *Tcrb* distribution in the naïve $CD8^+$ T cell pool is 1.1 cells, i.e., the number of different *Tcrb* expressed by naïve $CD8^+$ T cells approaches the total number of the naïve $CD8^+$ T cells. The addition of *Tcra* expression studies could only increase diversity (Cukalac et al., 2015). Therefore, the study of *Tcra* co-expression would not modify the general conclusion of this study: the number of different TCRs expressed by naïve $CD8^+$ T cells approaches the total number of the naïve $CD8^+$ T cells.

Our conclusions differ from some of the previous reports in several aspects. Firstly, our study contradicts previous studies indicating less repertoire diversity in both mouse and man (Arstila et al., 1999, 2000; Casrouge et al., 2000; Qi et al., 2014; Robins et al., 2009). However, these studies had limitations that can only be overcome by studying individual cells. All of them amplified samples of T cells using pools of multiple primers. Therefore, they cannot determine sequencing efficiency, i.e., the number of T cells that had their TCRs actually identified. These amplification conditions (when multiple primers are used simultaneously for amplification) favor primer competition, which increases with the number

of primers present during amplification (Peixoto et al., 2004), and is facilitated by the many similarities among the *Tcrb* genes. In conditions of primer competition, larger clones are preferentially amplified. The cloning of bands with the same CDR3 size after spectrotyping will further select for more abundant amplicons. Exhaustive next-generation sequencing was reported to be insufficient to capture the full repertoire of a subject (Warren et al., 2011; Zarnitsyna et al., 2013), the PCR amplification overestimating the repeated observation of TCR clonotypes in the sample, leading to false saturation (Robins et al., 2009). Next-generation sequencing is also associated with multiple PCR errors (Nguyen et al., 2011). The criterion used to select "true" sequences varies and is rarely described; in studies clearly defining this criterion, up to 50% of sequences were rejected (Nguyen et al., 2011). Rare TCR sequences might be mistaken for error-containing sequences and ignored, while larger clones, generating more abundant amplicons, have an increased probability that at least some of these amplicons are accepted as "true sequences", once again artificially reducing the estimates of repertoire diversity. Importantly, bulk studies cannot estimate clone sizes, and therefore evaluate repertoire diversity. They cannot determine whether identical amplicons derive from the same cell or correspond to multiple cells expressing the same TCR. Our approach overcomes these limitations. We counted precisely the number of cells where the *Tcrb* chain was sequenced. We had no primer competition since, for each individual cell, a single pair of primers was used for *Tcrb* chain amplification. In this single-cell analysis, the frequency of *Tcrbv* gene expression was consistent with that determined by cell surface staining with specific TCRB antibodies. Moreover, we are able to avoid the possibility that PCR errors bias our diversity estimates. In each experiment, we amplified the *Tcrb* from individual monoclonal P14 $CD8^+$ T cells expressing a known transgenic TCRB. All these cells had identical *tcrb* sequences.

As a preliminary step, we established why different TCRVBs were expressed at different frequencies in the peripheral pools. By comparing the frequency of *Tcrbv* families in the out-of-frame (non selected) and in-frame (selected) naïve cells, we showed that the majority of *Tcrbv* chains were expressed at similar frequencies in in-frame and out-of-frame rearrangements, indicating that their expression frequency was determined by the frequency of the respective *Tcrbv* rearrangements. However, some *Tcrbv* genes (e.g. *Tcrbv* 29) were more abundant in in-frame than out-of-frame rearrangements, while others were less abundant. These selections could occur already at the DN3 stage, by a different capacity of each TCRVB to associate with the pre-Tα chain. Indeed, the enrichment in *Tcrbv* 29 we here confirm was described to occur at the DN3-DN4-DP transition (Wei et al., 2006). However, the selection process may differ for different TCRVB chains, occurring at any stage after TCRA expression. It is possible that different TCRVB have different capacities to associate with TCRA chains, or form TCRA/B heterodimers with different avidity to bind to MHC. All these events could modify the efficiency of the thymus positive selection process, or/and of the peripheral T cell survival. To our knowledge, these aspects are yet to be studied with the necessary detail. In the sole study where the TCRB and TCRA chain co-expression was studied in individual cells, no preferential TCRB/TCRA association was detected, but the number of individual cells evaluated was relatively small (Dash et al., 2011).

Considering the reliability of our approach, determination of overall TCR diversity requires that multiple individual mice be evaluated and that the samples studied in each mouse are representative of the total TCR pool. For a rigorous study of the repertoire of the total naïve T cell pool, a representative sample (an average of 10^6cells/per mouse) should have been studied. Since for each cell a specific amplification for each *Tcrbv* is performed, such study would require $7x10^6$ PCR/sequences/cell (7 different PCRs per mouse for sequencing 10^6 cells), a total of $2.1x10^7$ individual sequences for three mice. Therefore, although we also

performed the mathematical analysis of diversity in the total naïve $CD8^+$ T cell pool, we focused our mathematical analysis of repertoire diversity on naïve $CD44^-GP33^+$ cells, where we can study representative samples in each mouse, corresponding to 7-20% of the total $CD44^-GP33^+$ pool.

It must be considered if the pool of $GP33^+$ cells we studied, which were much more abundant than previously described, could be contaminated with non-specific naïve cells which could artificially reduce diversity. Indeed, the number of $GP33^+CD8^+$ in naïve mice varied from 300 cells/mouse (Obar et al., 2008) to 1,000-1,200/spleen (Pewe et al., 2004; Seedhom et al., 2009). The differences in our methodology approach go some distance towards explaining these different results. Firstly, we did not study the same organs: of the 46 to 49 LNs described in the mouse we studied 40-42 while previous studies only studied 5 LNs, so that 41 to 44 LN were simply missing from their counts (Obar et al., 2008). One could envisage that all the larger LNs were studied, but that was not the case: para-aortic and some of the neck LNs are much larger than the axilliary and inguinal LN that they studied. We found that by studying this reduced LN pool, 50-60% of the LN $CD8^+$ T cells are lost (Sung, C. & Rocha, B. unpublished observations). Lastly, we have also evaluated the antigen-specific cells residing in the BM. The total number of $CXCR3^+CD8^+$ T resident BM cells was around $3x10^6$, i.e., the absence of BM counts also reduced the total $CD8^+$ cell yields by 15%. Since $GP33^+$ cells are particularly abundant in the BM, the absence of BM counts reduces the $GP33^+$ pool by 35%.

We also did not use the same methods to isolate Ag-specific cells. In previous studies (Obar et al., 2008) organs were dissociated by mechanical disruption alone, while we removed all fat under an inverted microscope with 10-30 fold amplification and further digested cell suspensions with collagenase (that increases cell yields) and DNAase (that removes debris and dead cells preventing the formation of clumps, thus decreasing cell loss during washing).

In our experience (Rocha, B. unpublished), cell viability is much increased by fully fat removal and the cell yields/organ increased by these further digestion steps. Positive selection of tetramers^{+} cells was reported to lead to the loss of 45-60% of antigen-specific cells (Lee and Lufkin, 2012; Moon et al., 2009), most of low-affinity antigen-specific cells being probably lost during this step (Dolton et al., 2015; Hadrup et al., 2009). By contrast, we did not perform positive selection of antigen-specific cells, but rather depleted cell suspensions of non-CD8^{+} T cells and calculated loss rates, by introducing to the cell suspension a precise known number of a "reference population" which undergoes the same preparation steps. Evaluation of loss rates was not previously performed (Haluszczak et al., 2009; Nunes-Alves et al., 2015; Obar et al., 2008; Seedhom et al., 2009).

As described in our methods section, we used dextramers rather than tetramers to identify antigen-specific cells, and dasatinib pre-treatment which blocks antigen induced TCR down-regulation and enhances peptide-MHC multimer fluorescence (Lissina et al., 2009). New generation peptide-MHC multimers were shown to detect more antigen-specific T cells compared with the equivalent tetramers and the positive selection of antigen-specific cells can be avoided (Huang et al., 2016). By contrast, the use of pMHC tetramers was reported to underestimate cells with lower affinity TCRs while pMHC dextramers are able detect antigen-specific cells with lower TCR avidity (Dolton et al., 2014; Massilamany et al., 2011) increasing the total number of antigen specific CD8^{+} T cells detected (Hataye et al., 2006; Huang et al., 2016; Obar et al., 2008). Moreover, it was claimed that this strategy increases the detection of antigen-specific cells without altering CD8^{+} T cell phenotype and inducing unspecific staining (Dolton et al., 2014; Lissina et al., 2009). Dextramers are specific for CD8^{+} T cells, do not bind to CD4 cells and yield minimal background staining comparable to tetramers over a broad range of concentrations (Massilamany et al. 2011; Dolton et al. 2014; Huang, et al. 2016). Indeed, low affinity antigen-specific cells that "escape" the positive

selection using tetramer$^+$ cell enrichment can be further identified by the use of these multimers (Huang et al., 2016). We showed that antigen-specific cells identified with MHC I multimers differentiate into cytotoxic cells after stimulation with the respective peptides (Munitic et al., 2009; Sung et al., 2013), while other studies demonstrate that low affinity multimer-binding identifies antigen-specific cells since they showed effector cytokine responses comparable to those of high-affinity tetramer$^+$ cells (Huang et al., 2016). Lower affinity CD8$^+$ T cells are also active participants in the immune response (Martinez and Evavold, 2015; Ozga et al., 2016), even extremely low-affinity TCR stimulation induces normal CD8$^+$ T cells activation and memory generation (Oberle et al., 2016; Tungatt et al., 2015). It must be noted, however, that the methodology we used may have limitations. Dextramer$^+$ CD8$^+$ T cells show a continuum of labeling intensities, rather than defined clear-cut populations, well separated from negative cells. The barrier separating GP33$^+$ from GP33$^-$ CD8$^+$ T cells was established by using CD8$^+$ T cells non-labeled with dextramers and MoP14 TCR-Tg cells labeled with dextramers. Using these barriers, many cells show low dextramer labeling. We conclude that our approach has significantly refined the number of steps for positive identification of naïve antigen-specific CD8$^+$ T cells. However, there is still the risk that inclusion of lower affinity TCRs and flow cytometric background events has included cells that are not functionally recruited to the immune response after LCMV infection.

Our mathematical analysis focused on what can be deduced, using our samples about the number of cells expressing unique *Tcrb* or *Tcrb* "clonotytpe" subsets and distribution of number of cells per "clonotype". Computational experiments, in which a repertoire is constructed in silico, based on different hypothesized distribution of clonotype sizes, and then sampled from, are used to validate mathematical conclusions. These mathematical analyses showed that *Tcrb* repertoires were compatible with a Geometric and Poisson distribution, the

probability of finding “clonotypes” with small size being much higher than that of finding large clones. However it was possible that the rare repeated sequences we found reflected the presence of two types of populations- the vast majority of cells would express unique *Tcrbs*, while the repeated *Tcrb* sequences would reflect a relative abundance of “clonotypes” of 2-3 cells sharing the same *Tcrb* or the presence of very rare larger clone sizes. However, the analysis of the naïve $CD8^+$ T cell population is only compatible with the last hypothesis. It must be noted that the majority of repeats found in the naïve $CD8^+$ T cell pool express selected *Tcrbv*, i.e., those that are expressed at higher frequency in in-frame than in out-of-frame rearrangements, suggesting the same mechanisms would be involved in the preferential selection of particular *Tcrbv*, and the generation of relatively large clonotype sizes.

Thus, the small size clones giving an essentially “private” TCR repertoire unique to each mice (Turner et al., 2003). However, some “public” T cell response, in which multiple individuals sharing identical in responding to a same pMHC epitope, giving a much more limited TCRB diversity, has been observed in a variety of immune responses (Cibotti et al., 1994; Kedzierska et al., 2004; Madi et al., 2014). Since these studies only evaluated TCRB expression, they may have failed to reveal a more extensive diversity introduced by the TCRA chain. As shown in a recent study, *Tcra* expression exposing diversity in an immune response, reported as having a narrow “public” *Tcrb* repertoire (Cukalac et al., 2015). The sharing of TCRB in the naïve T cell repertoire among multiple individuals provides the molecular basis for public T cell responses (Cukalac et al., 2015; Kedzierska et al., 2006; La Gruta et al., 2010). These shared TCRs may originate large clones selected to expand due to an optimal TCR recognition during the immune response. Several mechanisms have been proposed to generate naïve public TCRs, including preferences in the usage frequency and pairing of different V-D-J gene segments during initial TCR rearrangement (Turner et al., 2006; Venturi

et al., 2008). Some of these public TCR were generated from a near-germline V-D-J recombination (Fazilleau et al., 2005; Venturi et al., 2011) and other by "convergent recombination", where individual cells expressing different *Tcrb* nucleotide sequences "converge" to encode a TCRB protein with the same amino-acid composition (Turner et al., 2006; Venturi et al., 2008). Thymic selection can also influence TCR bias by both limiting (negative selection) and shaping (positive selection) the public naïve TCR repertoire (Turner et al., 2006; Venturi et al., 2008). It must be noted that in the 13 mice we studied, we did not find public *Tcrb* shared by all mice, and (Quinn et al., 2016) also failed to report such public sequences in anti-influenza virus repertoires from ten different mice.

Using this reliable single-cell approach, our results were striking, and impact in our holistic understanding of the immune system. During thymic differentiation, pre-T cells expressing a TCRB chain undergo 6-8 divisions at the TN3-TN4 transition, and continue to divide at the immature single-positive (ISP) and at least three times at the DP differentiation stages (Kedzierska et al., 2004; Kreslavsky et al., 2012; Von Boehmer, 2004). Because of TCRB allelic exclusion, thymocytes should maintain the expression of the same TCRB chain throughout all these divisions, thus up to 2^{11} cells expressing the same TCRB could be generated. Even if only 3% of these cells survived thymic selection events (Huesmann et al., 1991), one would expect an average clone size of 62 to 102 cells sharing the same TCRB chain. The difference between this putative average clone size, and the average of 1.1 cells we here determined can only be explained by the existence of a previously unknown mechanism determining TCRB repertoires: a selection for virtually complete diversity.

It is interesting to speculate how such extensive diversity is established and maintained in the periphery. The remarkable diversity of the peripheral naïve pools indicate that niches for selection and for optimal cell survival in the thymus or/and in the periphery must be very diverse and of very reduced size, to allow for cellular competition to impose the

survival of unique “perfect fits”. Peripheral intra-clonal and inter-clonal competition is a component of homeostasis that contributes to selection and maintenance of a diverse peripheral T cell repertoire (Freitas and Rocha, 2000; Hao et al., 2006; Hataye et al., 2006; Leitao et al., 2009). The specificity of TCR contact with a restricting MHC molecule is a prerequisite for peripheral inter-clonal T cell competition (Agenes et al., 2008). A positive correlation between the diversity of peptide-MHC complexes expressed in thymic epithelial cells and the diversity of the T cell repertoire has been demonstrated (Barton and Rudensky, 1999; SantAngelo et al., 1997). Moreover, a major selection process also occurs during thymus selection, where intra-clonal competition and a saturable TCR-specific dependent niche for positive selection is also described in the thymus (Bautista et al., 2009; Canelles et al., 2003). That such selection occurs at the TCR/pMHC interaction can be explained by the nature of naïve T cells surviving signals, which require TCR/pMHC interactions (Tanchot et al., 1997).

Another finding of this study is the remarkable number of $CD44^{high}$ T cells recognizing the GP33 LCMV peptide in SPF non-infected mice. These “mock memory cells” are known to be generated by homeostatic proliferation (HP), although the origin of the antigens inducing their switch from a naïve to a memory phenotype were reported to be diverse and are likely multiple. We showed that homeostatic proliferation and conversion from the naïve to the memory $CD44^{high}$ phenotype requires recognition of cross-reactive antigens and that the degree of this HP increases with the degree of cross-reactivity, cells with multiple cross-reactivity out-competing slowly proliferating cells (Freitas and Rocha, 2000; Hao et al., 2006). Such cross-reactivity was attributed to the recognition of self-antigens (Quinn et al., 2016; Surh and Sprent, 2000), but stimulation by food antigens and microbiota were also implicated (Beura et al., 2016; Kim et al., 2016). It was shown that T cells transferred to SCID germ-free (GF) mice proliferated slowly, while the same cells proliferated rapidly when

transferred to SPF SCID mice (Kieper et al., 2005). These experiments propose a major role of microbiota and food antigens in inducing rapid HP (Kieper et al., 2005), known to promote cellular competition (Hao et al., 2006). Moreover, these authors demonstrated the absence of bacteria in the gut of GF mice at the end of the experiment (Kieper et al., 2005). The major difficulty of maintaining the GF condition throughout experiments and the exposure to foreign antigens in the form of food (including microbial antigens present in the autoclaved feed) may explain the different results either supporting a major reduction of $CD44^{high}$ cells in GF mice (Beura et al., 2016; Dobber et al., 1992; Kieper et al., 2005) or indicating that $CD44^{high}$ cells are not modified by the GF condition (Haluszczak et al., 2009; Quinn et al., 2016). Cross-reactivity was also demonstrated in human adults, which also have memory cells that are specific for pathogens that they were not exposed previous (Su et al., 2013). On the other hand, all antigen-specific cells in human newborns have a naïve phenotype (Neller et al., 2015; Su et al., 2013), suggesting a major role of the cross-reactivity to environmental antigens in generation of human "mock memory cells" (Birnbaum et al., 2014; Su et al., 2013).

Our study also shows that cross-reactive $CD44^{high}GP33^{+}$ are particularly abundant in the BM. The BM is known to have niches that promote the long-term persistence of memory cells (Geerman et al., 2016). It is possible that these "antigen-experienced" mock memory cells have an important role in the response to LCMV. Although the BM is not usually considered as a major a target for the LCMV infection, LCMV can be present directly in BM (Slifka et al., 1995). Therefore it will be also interesting to evaluate the ability to BM mount primary immune responses against this virus, thus also functioning as a secondary lymphoid organ.

The present results also support a theoretical study postulating that the peripheral TCR repertoires should be extensively cross-reactive. To mobilize enough naïve cells to respond

efficiently to each individual challenge in primary immune responses, each individual naïve T cell should be able to recognize multiple antigens, and be mobilized in many different responses (Sewell, 2012). It remains to be fully understood how the immune system reconciles these three major characteristics: the extensive diversity and cross-reactivity of the peripheral TCR repertoires with the exquisite specificity of T cell immune responses.

Conflict of Interest: The authors have no financial conflicts of interest.

Acknowledgments: The European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement 317040 (ITN QuanTI) funded this study, and the researchers P. Gonçalves and M. Ferrarini. We thank António Freitas for their critical reading of the manuscript and comments.

Appendix A. Supplementary data: Supplementary data associated with this article can be found, in the online version, at…

References

Agenes, F., Dangy, J.P., and Kirberg, J. 2008. T cell receptor contact to restricting MHC molecules is a prerequisite for peripheral interclonal T cell competition. J Exp Med 205, 2735-U2742.

Anjuere, F., Martin, P., Ferrero, I., Fraga, M.L., del Hoyo, G.M., Wright, N., and Ardavin, C. 1999. Definition of dendritic cell subpopulations present in the spleen, Peyer's patches, lymph nodes, and skin of the mouse. Blood 93, 590-598.

Arstila, T.P., Casrouge, A., Baron, V., Even, J., Kanellopoulos, J., and Kourilsky, P. 1999. A direct estimate of the human alphabeta T cell receptor diversity. Science 286, 958-961.

Arstila, T.P., Casrouge, A., Baron, V., Even, J., Kanellopoulos, J., and Kourilsky, P. 2000. Diversity of human alpha beta T cell receptors. Science 288, 1135.

Barton, G.M., and Rudensky, A.Y. 1999. Requirement for diverse, low-abundance peptides in positive selection of T cells. Science 283, 67-70.

Bautista, J.L., Lio, C.W.J., Lathrop, S.K., Forbush, K., Liang, Y.Q., Luo, J.Q., Rudensky, A.Y., and Hsieh, C.S. 2009. Intraclonal competition limits the fate determination of regulatory T cells in the thymus. Nat Immunol 10, 610-U674.

Beura, L.K., Hamilton, S.E., Bi, K., Schenkel, J.M., Odumade, O.A., Casey, K.A., Thompson, E.A., Fraser, K.A., Rosato, P.C., Filali-Mouhim, A., et al. 2016. Normalizing the environment recapitulates adult human immune traits in laboratory mice. Nature 532, 512-+.

Birnbaum, M.E., Mendoza, J.L., Sethi, D.K., Dong, S., Glanville, J., Dobbins, J., Ozkan, E., Davis, M.M., Wucherpfennig, K.W., and Garcia, K.C. 2014. Deconstructing the Peptide-MHC Specificity of T Cell Recognition. Cell 157, 1073-1087.

Blattman, J.N., Sourdive, D.J.D., Murali-Krishna, K., Ahmed, R., and Altman, J.D. (2000). Evolution of the T cell repertoire during primary, memory, and recall responses to viral infection. J Immunol 165, 6081-6090.

Bosc, N., and Lefranc, M.P. 2000. The mouse (Mus musculus) T cell receptor beta variable (TRBV), diversity (TRBD) and joining (TRBJ) genes. Exp Clin Immunogenet 17, 216-228.

Candeias, S., Waltzinger, C., Benoist, C., and Mathis, D. 1991. The V beta 17+ T cell repertoire: skewed J beta usage after thymic selection; dissimilar CDR3s in CD4+ versus CD8+ cells. J Exp Med 174, 989-1000.

Canelles, M., Park, M.L., Schwartz, O.M., and Fowlkes, B.J. 2003. The influence of the thymic environment on the CD4-versus-CD8 T lineage decision. Nat Immunol 4, 756-764.

Carey, A.J., Gracias, D.T., Thayer, J.L., Boesteanu, A.C., Kumova, O.K., Mueller, Y.M., Hope, J.L., Fraietta, J.A., van Zessen, D.B., and Katsikis, P.D. 2016. Rapid Evolution of the CD8+ TCR Repertoire in Neonatal Mice. J Immunol 196, 2602-2613.

Carlsson, L., and Holmberg, D. 1990. Genetic-Basis of the Neonatal Antibody Repertoire - Germline V-Gene Expression and Limited N-Region Diversity. Int Immunol 2, 639-643.

Casrouge, A., Beaudoing, E., Dalle, S., Pannetier, C., Kanellopoulos, J., and Kourilsky, P. 2000. Size estimate of the alpha beta TCR repertoire of naive mouse splenocytes. J Immunol 164, 5782-5787.

Chaix, J., Nish, S.A., Lin, W.H.W., Rothman, N.J., Ding, L., Wherry, E.J., and Reiner, S.L. (2014). Cutting Edge: CXCR4 Is Critical for CD8(+) Memory T Cell Homeostatic Self-Renewal but Not Rechallenge Self-Renewal. J Immunol 193, 1013-1016.

Cibotti, R., Cabaniols, J.P., Pannetier, C., Delarbre, C., Vergnon, I., Kanellopoulos, J.M., and Kourilsky, P. (1994). Public and Private V-Beta T-Cell Receptor Repertoires against Hen Egg-White Lysozyme (Hel) in Nontransgenic Versus Hel Transgenic Mice. J Exp Med 180, 861-872.

Cukalac, T., Kan, W.T., Dash, P., Guan, J., Quinn, K.M., Gras, S., Thomas, P.G., and La Gruta, N.L. 2015. Paired TCR alpha beta analysis of virus-specific CD8(+) T cells exposes diversity in a previously defined 'narrow' repertoire. Immunol Cell Biol 93, 804-814.

Das, D.K., Feng, Y., Mallis, R.J., Li, X., Keskin, D.B., Hussey, R.E., Brady, S.K., Wang, J.H., Wagner, G., Reinherz, E.L., et al. 2015. Force-dependent transition in the T-cell receptor beta-subunit allosterically regulates peptide discrimination and pMHC bond lifetime. Proc Natl Acad Sci U S A 112, 1517-1522.

Dash, P., McClaren, J.L., Oguin, T.H., Rothwell, W., Todd, B., Morris, M.Y., Becksfort, J., Reynolds, C., Brown, S.A., Doherty, P.C., *et al.* (2011). Paired analysis of TCR alpha and TCR beta chains at the single-cell level in mice. J Clin Invest. 121, 288-295.

Dobber, R., Hertogh-Huijbregts, A., Rozing, J., Bottomly, K., and Nagelkerken, L. 1992. The involvement of the intestinal microflora in the expansion of CD4+ T cells with a naive phenotype in the periphery. Dev Immunol 2, 141-150.

Dolton, G., Lissina, A., Skowera, A., Ladell, K., Tungatt, K., Jones, E., Kronenberg-Versteeg, D., Akpovwa, H., Pentier, J.M., Holland, C.J., et al. 2014. Comparison of peptide-major histocompatibility complex tetramers and dextramers for the identification of antigen-specific T cells. Clin Exp Immunol 177, 47-63.

Dolton, G., Tungatt, K., Lloyd, A., Bianchi, V., Theaker, S.M., Trimby, A., Holland, C.J., Donia, M., Godkin, A.J., Cole, D.K., et al. 2015. More tricks with tetramers: a practical guide to staining T cells with peptide-MHC multimers. Immunology 146, 11-22.

Eltahla, A.A., Rizzetto, S., Pirozyan, M.R., Betz-Stablein, B.D., Venturi, V., Kedzierska, K., Lloyd, A.R., Bull, R.A., and Luciani, F. 2016. Linking the T cell receptor to the single cell transcriptome in antigen-specific human T cells. Immunol Cell Biol 94, 604-611.

Fazilleau, N., Cabaniols, J.P., Lemaitre, F., Motta, I., Kourilsky, P., and Kanellopoulos, J.M. (2005). Valpha and Vbeta public repertoires are highly conserved in terminal deoxynucleotidyl transferase-deficient mice. J Immunol 174, 345-355.

Freitas, A.A., and Rocha, B. 2000. Population biology of lymphocytes: the flight for survival. Annu Rev Immunol 18, 83-111.

Geerman, S., Hickson, S., Brasser, G., Pascutti, M.F., and Nolte, M.A. 2016. Quantitative and Qualitative analysis of Bone Marrow cD8(+) T cells from Different Bones Uncovers a Major contribution of the Bone Marrow in the Vertebrae. Front Immunol 2016, 6. 660.

Genolet, R., Stevenson, B.J., Farinelli, L., Osteras, M., and Luescher, I.F. (2012). Highly diverse TCRalpha chain repertoire of pre-immune CD8(+) T cells reveals new insights in gene recombination. EMBO J. 31, 1666-1678.

Hadrup, S.R., Bakker, A.H., Shu, C.Y.J., Andersen, R.S., van Veluw, J., Hombrink, P., Castermans, E., Straten, P.T., Blank, C., Haanen, J.B., et al. 2009. Parallel detection of antigen-specific T-cell responses by multidimensional encoding of MHC multimers. Nat Methods 6, 520-U579.

Haluszczak, C., Akue, A.D., Hamilton, S.E., Johnson, L.D.S., Pujanauski, L., Teodorovic, L., Jameson, S.C., and Kedl, R.M. 2009. The antigen-specific CD8(+) T cell repertoire in unimmunized mice includes memory phenotype cells bearing markers of homeostatic expansion. J Exp Med 206, 435-448.

Hao, Y., Legrand, N., and Freitas, A.A. 2006. The clone size of peripheral CD8 T cells is regulated by TCR promiscuity. J Exp Med 203, 1643-1649.

Hataye, J., Moon, J.J., Khoruts, A., Reilly, C., and Jenkins, M.K. 2006. Naive and memory CD4(+) T cell survival controlled by clonal abundance. Science 312, 114-116.

Huang, J., Zeng, X., Sigal, N., Lund, P.J., Su, L.F., Huang, H., Chien, Y.H., and Davis, M.M. 2016. Detection, phenotyping, and quantification of antigen-specific T cells using a peptide-MHC dodecamer. P Natl Acad Sci USA 113, E1890-E1897.

Huesmann, M., Scott, B., Kisielow, P., and Vonboehmer, H. 1991. Kinetics and Efficacy of Positive Selection in the Thymus of Normal and T-Cell Receptor Transgenic Mice. Cell 66, 533-540.

Itkin, T., Gur-Cohen, S., Spencer, J.A., Schajnovitz, A., Ramasamy, S.K., Kusumbe, A.P., Ledergor, G., Jung, Y., Milo, I., Poulos, M.G., et al. 2016. Distinct bone marrow blood vessels differentially regulate haematopoiesis. Nature 21, 323-328.

Jameson, S.C., Lee, Y.J., and Hogquist, K.A. 2015. Innate memory T cells. Adv Immunol 126, 173-213.

Jenkins, M.K., Chu, H.H., McLachlan, J.B., and Moon, J.J. 2010. On the Composition of the Preimmune Repertoire of T Cells Specific for Peptide-Major Histocompatibility Complex Ligands. Annu Rev Immunol 28, 275-294.

Kato, T., Suzuki, S., Sasakawa, H., Masuko, K., Ikeda, Y., Nishioka, K., and Yamamoto, K. 1994. Comparison of the J-Beta Gene Usage among Different T-Cell Receptor V-Beta Families in Spleens of C57bl/6 Mice. Eur J Immunol 24, 2410-2414.

Kedzierska, K., Turner, S.J., and Doherty, P.C. 2004. Conserved T cell receptor usage in primary and recall responses to an immunodominant influenza virus nucleoprotein epitope. P Natl Acad Sci USA 101, 4942-4947.

Kedzierska, K., Venturi, V., Field, K., Davenport, M.P., Turner, S.J., and Doherty, P.C. (2006). Early establishment of diverse T cell receptor profiles for influenza-specific CD8(+)CD62L(hi) memory T cells. P Natl Acad Sci USA 103, 9184-9189.

Kieper, W.C., and Jameson, S.C. 1999. Homeostatic expansion and phenotypic conversion of naive T cells in response to self peptide/MHC ligands. P Natl Acad Sci USA 96, 13306-13311.

Kieper, W.C., Troy, A., Burghardt, J.T., Ramsey, C., Lee, J.Y., Jiang, H.Q., Dummer, W., Shen, H., Cebra, J.J., and Surh, C.D. 2005. Cutting edge: Recent immune status determines the source of antigens that drive homeostatic T cell expansion. J Immunol 174, 3158-3163.

Kim, K.S., Hong, S.W., Han, D., Yi, J., Jung, J., Yang, B.G., Lee, J.Y., Lee, M., and Surh, C.D. 2016. Dietary antigens limit mucosal immunity by inducing regulatory T cells in the small intestine. Science. 351, 858-863.

Kreslavsky, T., Gleimer, M., Miyazaki, M., Choi, Y., Gagnon, E., Murre, C., Sicinski, P., and von Boehmer, H. 2012. beta-Selection-Induced Proliferation Is Required for alpha beta T Cell Differentiation. Immunity 37, 840-853.

La Gruta, N.L., Rothwell, W.T., Cukalac, T., Swan, N.G., Valkenburg, S.A., Kedzierska, K., Thomas, P.G., Doherty, P.C., and Turner, S.J. (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, 1885-1894.

Lee, M.Y., and Lufkin, T. 2012. Development of the "Three-step MACS": a novel strategy for isolating rare cell populations in the absence of known cell surface markers from complex animal tissue. J Biomol Tech 23, 69-77.

Lefranc, M.P. (2001). IMGT, the international ImMunoGeneTics database. Nucleic Acids Res 29, 207-209.

Leitao, C., Freitas, A.A., and Garcia, S. 2009. The Role of TCR Specificity and Clonal Competition During Reconstruction of the Peripheral T Cell Pool. J Immunol 182, 5232-5239.

Lin, M.Y., and Welsh, R.M. (1998). Stability and diversity of T cell receptor repertoire usage during lymphocytic choriomeningitis virus infection of mice. J Exp Med 188, 1993-2005.

Lissina, A., Ladell, K., Skowera, A., Clement, M., Edwards, E., Seggewiss, R., van den Berg, H.A., Gostick, E., Gallagher, K., Jones, E., et al. 2009. Protein kinase inhibitors substantially improve the physical detection of T-cells with peptide-MHC tetramers. J Immunol Methods 340, 11-24.

Lythe, G., Callard, R.E., Hoare, R.L., and Molina-Paris, C. 2016. How many TCR clonotypes does a body maintain? J Theor Biol 389, 214-224.

Madi, A., Shifrut, E., Reich-Zeliger, S., Gal, H., Best, K., Ndifon, W., Chain, B., Cohen, R.R., and Friedman, N. (2014). T-cell receptor repertoires share a restricted set of public and abundant CDR3 sequences that are associated with self-related immunity. Genome Res 24, 1603-1612.

Martinez, R.J., and Evavold, B.D. (2015). Lower affinity T cells are critical components and active participants of the immune response. Front Immunol 6, 468.

Massilamany, C., Upadhyaya, B., Gangaplara, A., Kuszynski, C., and Reddy, J. 2011. Detection of autoreactive CD4 T cells using major histocompatibility complex class II dextramers. Bmc Immunol 12.

Moon, J.J., Chu, H.H., Hataye, J., Pagan, A.J., Pepper, M., McLachlan, J.B., Zell, T., and Jenkins, M.K. 2009. Tracking epitope-specific T cells. Nat Protoc 4, 565-581.

Munitic, I., Decaluwe, H., Evaristo, C., Lemos, S., Wlodarczyk, M., Worth, A., Le Bon, A., Selin, L.K., Riviere, Y., Di Santo, J.P., et al. (2009). Epitope specificity and relative clonal abundance do not affect CD8 differentiation patterns during lymphocytic choriomeningitis virus infection. J Virol 83, 11795-11807.

Neller, M.A., Ladell, K., McLaren, J.E., Matthews, K.K., Gostick, E., Pentier, J.M., Dolton, G., Schauenburg, A.J.A., Koning, D., Costa, A.I.C.A.F., et al. 2015. Naive CD8(+) T-cell precursors display structured TCR repertoires and composite antigen-driven selection dynamics. Immunol Cell Biol 93, 625-633.

Nguyen, P., Ma, J., Pei, D., Obert, C., Cheng, C., and Geiger, T.L. 2011. Identification of errors introduced during high throughput sequencing of the T cell receptor repertoire. BMC genomics 12, 106.

Nunes-Alves, C., Booty, M.G., Carpenter, S.M., Rothchild, A.C., Martin, C.J., Desjardins, D., Steblenko, K., Kloverpris, H.N., Madansein, R., Ramsuran, D., et al. 2015. Human

and Murine Clonal CD8+T Cell Expansions Arise during Tuberculosis Because of TCR Selection. Plos Pathog 11(5).

Obar, J.J., Khanna, K.M., and Lefrancois, L. 2008. Endogenous naive CD8+ T cell precursor frequency regulates primary and memory responses to infection. Immunity 28, 859-869.

Oberle, S.G., Hanna-El-Daher, L., Chennupati, V., Enouz, S., Scherer, S., Prlic, M., and Zehn, D. (2016). A Minimum Epitope Overlap between Infections Strongly Narrows the Emerging T Cell Repertoire. Cell Rep 17, 627-635.

Ozga, A.J., Moalli, F., Abe, J., Swoger, J., Sharpe, J., Zehn, D., Kreutzfeldt, M., Merkler, D., Ripoll, J., and Stein, J.V. (2016). pMHC affinity controls duration of CD8+ T cell-DC interactions and imprints timing of effector differentiation versus expansion. J Exp Med. 213, 2811-2829.

Peaudecerf, L., Lemos, S., Galgano, A., Krenn, G., Vasseur, F., Di Santo, J.P., Ezine, S., and Rocha, B. (2012). Thymocytes may persist and differentiate without any input from bone marrow progenitors. J Exp Med 209, 1401-1408.

Peixoto, A., Monteiro, M., Rocha, B., and Veiga-Fernandes, H. 2004. Quantification of multiple gene expression in individual cells. Genome res 14, 1938-1947.

Penit, C., Lucas, B., and Vasseur, F. (1995). Cell expansion and growth arrest phases during the transition from precursor (CD4-8-) to immature (CD4+8+) thymocytes in normal and genetically modified mice. J Immunol 154, 5103-5113.

Penit, C., and Vasseur, F. (1997). Expansion of mature thymocyte subsets before emigration to the periphery. J Immunol 159, 4848-4856.

Petit, I., Szyper-Kravitz, M., Nagler, A., Lahav, M., Peled, A., Habler, L., Ponomaryov, T., Taichman, R.S., Arenzana-Seisdedos, F., Fujii, N., *et al.* (2002). G-CSF induces stem cell

mobilization by decreasing bone marrow SDF-1 and up-regulating CXCR4. Nat Immunol 3, 687-694.

Pewe, L.L., Netland, J.M., Heard, S.B., and Perlman, S. 2004. Very diverse CD8 T cell clonotypic responses after virus infections. J Immunol 172, 3151-3156.

Qi, Q., Liu, Y., Cheng, Y., Glanville, J., Zhang, D., Lee, J.Y., Olshen, R.A., Weyand, C.M., Boyd, S.D., and Goronzy, J.J. 2014. Diversity and clonal selection in the human T-cell repertoire. P Natl Acad Sci USA 111, 13139-13144.

Quigley, M.F., Greenaway, H.Y., Venturi, V., Lindsay, R., Quinn, K.M., Seder, R.A., Douek, D.C., Davenport, M.P., and Price, D.A. 2010. Convergent recombination shapes the clonotypic landscape of the naive T-cell repertoire. P Natl Acad Sci USA 107, 19414-19419.

Quinn, K.M., Zaloumis, S.G., Cukalac, T., Kan, W.T., Sng, X.Y., Mirams, M., Watson, K.A., McCaw, J.M., Doherty, P.C., Thomas, P.G., et al. 2016. Heightened self-reactivity associated with selective survival, but not expansion, of naive virus-specific CD8+ T cells in aged mice. Proc Natl Acad Sci U S A. 113: 1333-1338.

Robins, H.S., Campregher, P.V., Srivastava, S.K., Wacher, A., Turtle, C.J., Kahsai, O., Riddell, S.R., Warren, E.H., and Carlson, C.S. 2009. Comprehensive assessment of T-cell receptor beta-chain diversity in alpha beta T cells. Blood 114, 4099-4107.

SantAngelo, D.B., Waterbury, P.G., Cohen, B.E., Martin, W.D., VanKaer, L., Hayday, A.C., and Janeway, C.A. 1997. The imprint of intrathymic self-peptides on the mature T cell receptor repertoire. Immunity 7, 517-524.

Seedhom, M.O., Jellison, E.R., Daniels, K.A., and Welsh, R.M. 2009. High frequencies of virus-specific CD8+ T-cell precursors. Journal of virology 83, 12907-12916.

Sewell, A.K. 2012. Why must T cells be cross-reactive? Nat Rev Immunol 12, 668-677.

Shortman, K., and Jackson, H. (1974). The differentiation of T lymphocytes. I. Proliferation kinetics and interrelationships of subpopulations of mouse thymus cells. Cell Immunol 12, 230-246.

Slifka, M.K., Matloubian, M., and Ahmed, R. 1995. Bone-Marrow Is a Major Site of Long-Term Antibody-Production after Acute Viral-Infection. J Virol 69, 1895-1902.

Stubbington, M.J.T., Lonnberg, T., Proserpio, V., Clare, S., Speak, A., Dougan, G., and Teichmann, S.A. 2016. T cell fate and clonality inference from single-cell transcriptomes. Nat Methods 13, 329-332.

Su, L.F., and Davis, M.M. 2013. Antiviral memory phenotype T cells in unexposed adults. Immunol Rev 255, 95-109.

Su, L.F., Kidd, B.A., Han, A., Kotzin, J.J., and Davis, M.M. 2013. Virus-specific CD4(+) memory-phenotype T cells are abundant in unexposed adults. Immunity 38, 373-383.

Sung, H.C., Lemos, S., Ribeiro-Santos, P., Kozyrytska, K., Vasseur, F., Legrand, A., Charbit, A., Rocha, B., and Evaristo, C. 2013. Cognate antigen stimulation generates potent CD8(+) inflammatory effector T cells. Front Immunol 4, 452.

Surh, C.D., and Sprent, J. 2000. Homeostatic T cell proliferation: How far can T cells be activated to self-ligands? J Exp Med 192, F9-F14.

Tanchot, C., Lemonnier, F.A., Perarnau, B., Freitas, A.A., and Rocha, B. 1997. Differential requirements for survival and proliferation of CD8 naive or memory T cells. Science 276, 2057-2062.

Tungatt, K., Bianchi, V., Crowther, M.D., Powell, W.E., Schauenburg, A.J., Trimby, A., Donia, M., Miles, J.J., Holland, C.J., Cole, D.K., et al. (2015). Antibody Stabilization of Peptide-MHC Multimers Reveals Functional T Cells Bearing Extremely Low-Affinity TCRs. J Immunol 194, 463-474.

Turner, S.J., Diaz, G., Cross, R., and Doherty, P.C. (2003). Analysis of clonotype distribution and persistence for an influenza virus-specific CD8(+) T cell response. Immunity 18, 549-559.

Turner, S.J., Doherty, P.C., McCluskey, J., and Rossjohn, J. (2006). Structural determinants of T-cell receptor bias in immunity. Nat Rev Immunol 6, 883-894.

Van den Broeck, W., Derore, A., and Simoens, P. 2006. Anatomy and nomenclature of murine lymph nodes: Descriptive study and nomenclatory standardization in BALB/cAnNCrl mice. J Immunol Methods 312, 12-19.

Venturi, V., Price, D.A., Douek, D.C., and Davenport, M.P. (2008). The molecular basis for public T-cell responses? Nat Rev Immunol 8, 231-238.

Venturi, V., Quigley, M.F., Greenaway, H.Y., Ng, P.C., Ende, Z.S., McIntosh, T., Asher, T.E., Almeida, J.R., Levy, S., Price, D.A., *et al.* (2011). A Mechanism for TCR Sharing between T Cell Subsets and Individuals Revealed by Pyrosequencing. J Immunol 186, 4285-4294.

Von Boehmer, H. 2004. Selection of the T-cell repertoire: Receptor-controlled checkpoints in T-cell development. Adv Immunol 84, 201-238.

Warren, R.L., Freeman, D., Zeng, T., Choe, G., Munro, S., Moore, R., Webb, J.R., and Holt, R.A. 2011. Exhaustive T-cell repertoire sequencing of human peripheral blood samples reveals signatures of antigen selection and a directly measured repertoire size of at least 1 million clonotypes. Genome res 21, 790-797.

Wei, D.G., Curran, S.A., Savage, P.B., Teyton, L., and Bendelac, A. (2006). Mechanisms imposing the Vbeta bias of Valpha14 natural killer T cells and consequences for microbial glycolipid recognition. J Exp Med 203, 1197-1207.

Zarnitsyna, V.I., Evavold, B.D., Schoettle, L.N., Blattman, J.N., and Antia, R. 2013. Estimating the diversity, completeness, and cross-reactivity of the T cell repertoire. Front Immunol 4, 485.

Figure legends

Figure 1. The gating strategy used for the identification and single-cell sorting of $CD44^{-}$ (negative), $CD44^{int}$ (intermediate) and $CD44^{high}$ $TCR\alpha\beta^{+}$ $CD8\alpha\beta^{+}$ D^{b}-GP33-specific T cells in lymph nodes, spleen and bone marrow. The detailed methods are described in M&M. Briefly, a known number of monoclonal $CD45.1^{+}$ P14 TCR transgenic cells was added to each petri dish with the different organs from $CD45.2^{+}$ B6 mice. Cell suspensions from the organs were

prepared by mechanic disruption and collagenase and DNAase digestion, depleted of non-$CD8^+$ T cells and labelled with anti-CD45.1, anti-CD3ε, anti-CD8, anti-CD44, Db GP33 dextramers and Sytox to eliminated dead cells. Cells were successively gated (upper graphs from left to right) in $CD8^+$; $CD3^+$; $Sytox^-$; FSC/SSC, and doubles were eliminated. The reference population was identified by CD45.1 expression (middle graphs), and allowed us to establish the relative loss rate and gates for dextramer D^b-$GP33^+$cells (right). These gates were used to identify, sort or quantify $GP33^+$ cells, in different organs. The numbers shown in each quadrant represent the mean+/SEM of three mice, each studied in a separate experiment.

Figure 2. Relative frequency (%) of TRBV (A); TRBJ usage (B); CDR3 amino-acid length (C) in $CD8^+$ T cells from SPF B6 mice. Results represent the mean +/- SEM of four mice. The number of *Tcrb* sequences studied in each mouse is shown in Table 2. Cells expressing TRBV21 are too rare to be visible on this scale.

Figure 3. Relative frequency (%) of TRBV usage by in-frame (n = 579- grey bars) and out-of-frame (n = 165-black bars) TCR rearrangements in $CD44^-CD8^+$ T cells of SPF B6 mice. Results represent the mean +/- SEM.

Figure 4. (A) Relative frequency (%) of TRBV usage in the total $CD44^-$ (white bars; 3 mice; 386 *Tcrb* sequences) versus antigen-specific $CD44^-GP33^+$ (black bars, 5 mice; 1.104 *Tcrb* sequences) from SPF B6 mice. (B) Representative dot plots identifying TRBV13 (Vb8) and TRBV29 (Vb7) measured by FACS. (C) Relative frequency (%) of TRBV13 (Vb8) and TRBV29 (Vb7) usage in the total $CD44^-CD8^+$ and antigen-specific $CD44^-GP33^+$ in spleen (SP) and bone marrow (BM) from SPF B6 mice.