# Diversity and clonal selection in the human T-cell repertoire

Qian Qi[a,1], Yi Liu[b,c,1], Yong Cheng[d], Jacob Glanville[c,e], David Zhang[a], Ji-Yeun Lee[c], Richard A. Olshen[f], Cornelia M. Weyand[a,g], Scott D. Boyd[c,2], and Jörg J. Goronzy[a,g,2,3]

Divisions of [a]Immunology and Rheumatology, Department of Medicine and [f]Biostatistics, [b]Stanford Center for Biomedical Informatics Research, and Departments of [c]Pathology, [d]Genetics, and [e]Immunology, Stanford University, Stanford, CA 94305; and [g]Department of Medicine, VA Palo Alto Health Care System, Palo Alto, CA 94306

Edited by Rino Rappuoli, Novartis Vaccines, Siena, Italy, and approved August 1, 2014 (received for review May 19, 2014)

**T-cell receptor (TCR) diversity, a prerequisite for immune system recognition of the universe of foreign antigens, is generated in the first two decades of life in the thymus and then persists to an unknown extent through life via homeostatic proliferation of naïve T cells. We have used next-generation sequencing and nonparametric statistical analysis to estimate a lower bound for the total number of different TCR beta (TCRB) sequences in human repertoires. We arrived at surprisingly high minimal estimates of 100 million unique TCRB sequences in naïve CD4 and CD8 T-cell repertoires of young adults. Naïve repertoire richness modestly declined two- to fivefold in healthy elderly. Repertoire richness contraction with age was even less pronounced for memory CD4 and CD8 T cells. In contrast, age had a major impact on the inequality of clonal sizes, as estimated by a modified Gini–Simpson index clonality score. In particular, large naïve T-cell clones that were distinct from memory clones were found in the repertoires of elderly individuals, indicating uneven homeostatic proliferation without development of a memory cell phenotype. Our results suggest that a highly diverse repertoire is maintained despite thymic involution; however, peripheral fitness selection of T cells leads to repertoire perturbations that can influence the immune response in the elderly.**

adaptive immune responses | aging | immunosenescence | T-cell homeostasis

The ability of the adaptive immune system to respond to a wide variety of pathogens depends on a large repertoire of unique T-cell receptors (TCRs). TCR diversity is generated by the random and imprecise rearrangements of the V and J segments of the TCR alpha (TCRA) and V, D, and J segments of the TCR beta (TCRB) genes in the thymus. Thymic production of T cells is the sole mechanism to generate TCR diversity. With the involution of the thymus with age, the generation of new naïve T cells with new TCRs dwindles (1, 2). In the mouse, residual thymic activity provides an ongoing source of naïve T cells, although not sufficient to maintain the compartment size. In contrast, homeostatic proliferation of peripheral T cells is the predominant mechanism of T-cell generation in the human adult (3).

Given the dramatic decline in thymic T-cell production, thymic involution has been implicated in defective T-cell responses with aging (4). Elderly individuals are more prone to develop complications from infections; in particular, they are susceptible to newly arising infectious organisms such as West Nile fever or severe acute respiratory syndrome (5, 6). Moreover, immune responses to vaccination are dramatically reduced (7). Shrinkage in the size of the naïve T-cell compartment and holes in the repertoire have been discussed as potential causative factors in age-related impaired adaptive immunity (8). Indeed, holes have been identified in the murine repertoire to contribute to the increased morbidity from influenza infection (9, 10).

Pioneering early work on estimates of diversity of the human TCR repertoire was based on sequencing a few hundred sequences and extrapolating to the scale of the entire repertoire, yielding an estimated lower limit estimate of fewer than 1 million different TCRB genes (11). Studies in more recent years have used high-throughput sequencing to base estimates on greater sequencing depth. Warren et al. (12) measured 1 million different TCRB sequences in a peripheral blood sample. Robins et al. (13) inferred the number of unseen TCRB genes from deep sequencing data by applying a Poisson process model and estimated a diversity of 3–4 million TCRB sequences in the total T-cell populations of two healthy donors.

Although high-throughput DNA sequencing provides the tool to gather extensive datasets, estimates of TCR richness remain a challenge because clinical samples represent a small fraction of the $10^{12}$ cells in the overall T-cell compartment. Richness, the actual number of unique TCR sequences in a T-cell population, is dominated by the infrequent species that are observed rarely or not at all. The major factors affecting accurate identification of rare species in TCR repertoires are the number of cells examined, the variable efficiency of PCR amplification, the expression levels of TCR transcripts in different T cells, and sequencing error rates. In addition, peripheral blood T cells consist of a mixed population of naïve, memory, and effector CD4 and CD8 cells that differ markedly in diversity, with memory populations being less complex than naïve populations.

## Significance

**A decline in the diversity of the T-cell receptor repertoire owing to thymic involution has been implicated as causing defective immune responses in the elderly. By applying next-generation sequencing of replicate TCRB libraries from highly purified T-cell subsets, and using nonparametric statistical analysis, we obtain estimates of repertoire richness in the young adult that are higher than previously reported. Although contracting with age, the repertoire remains highly diverse. These data challenge the paradigm that thymic rejuvenation is needed to maintain diversity and prevent immune incompetence in the elderly. However, we observe an increasing inequality of clonal sizes with age even among naïve T cells. This clonal selection could result in biased and possibly autoreactive immune responses.**

Author contributions: Q.Q., Y.L., Y.C., J.G., R.A.O., C.M.W., S.D.B., and J.J.G. designed research; Q.Q., Y.L., D.Z., and J.-Y.L. performed research; Q.Q., Y.L., R.A.O., S.D.B., and J.J.G. analyzed data; and Q.Q., Y.L., R.A.O., C.M.W., S.D.B., and J.J.G. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Data deposition: The sequences reported in this paper have been deposited in the Sequence Read Archive and can be accessed via the database of Genotypes and Phenotypes (dbGaP), www.ncbi.nlm.nih.gov/gap (accession no. phs000787.v1.p1).

[1]Q.Q. and Y.L. contributed equally to this work.

[2]S.D.B. and J.J.G. contributed equally to this work.

[3]To whom correspondence should be addressed. Email: jgoronzy@stanford.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1409155111/-/DCSupplemental.

IMMUNOLOGY AND INFLAMMATION

The current study was designed explicitly to improve estimation of the TCR repertoire richness of stringently defined naïve and memory T cells. By sequencing multiple replicate TCRB libraries of cells from each T-cell subset and applying nonparametric statistical analysis, we find that human TCR repertoires are an order of magnitude more diverse than previously estimated. Despite significant age-related decreases in richness, humans maintain high diversity during healthy aging. Strikingly, we find age-associated changes in the distribution of clonal sizes, in particular in the naïve compartment, which may reflect unevenness of homeostatic expansion with clonal expansion of some naïve T cells that equal or exceed clonal sizes of memory T cells. The inequalities in clonal sizes could compromise the immune response to the majority of antigenic epitopes while causing an increased responsiveness to selected few epitopes.

## Results

**Gene Segment Use and CDR3 Features in Young and Elderly TCRB Rearrangements.** In initially evaluating TCRB repertoires in young and elderly subjects we compared the composition of TCRB gene rearrangements at the level of TCRBV and TCRBJ gene segment use and the features of the CDR3-encoding junctional nucleotides. Apheresis lymphocytes were obtained from four 20- to 35- and five 70- to 85-y-old healthy adults who were regular platelet donors. Naïve CD4 and CD8 T cells were purified by cell sorting. We used a stringent definition of naïve cells ($CD3^+CD4^+$ or $CD8^+CCR7^+CD45RA^{high}CD28^+$) and very restrictive gating to ensure purity. Approximately 1.5–3 million sequence reads were obtained for each T-cell subset of each donor (Table S1). TCRBV and TCRBJ gene segments were used at comparable frequencies in the repertoires of naïve CD4 and CD8 T cells in young and old individuals (Fig. S1*A*), whereas the memory repertoires of CD4 and, particularly, CD8 T cells show variable and individual-specific gene segment use frequencies (Fig. S1*B*). CDR3 sequences in the young and elderly were comparable in length and did not show any definitive age-related features (Fig. S1 *C* and *D*).

**High Richness of Naïve CD4 and CD8 TCRB Repertoires in Young and Elderly Adults.** To obtain sequence data to estimate global TCRB repertoire richness we used the experimental design of analyzing a series of replicate libraries from independent cell aliquots from each T-cell subset in each individual. These replicates allowed us to calculate repertoire richness by applying the "Chao2" estimator, a nonparametric estimator of unseen species (14). The approach allows estimation of the extent to which the full repertoire is covered and use of this information to determine a lower bound of the total number of species in the repertoire. Because the Chao2 estimator requires only a binary characterization of presence or absence of each clone in each replicate library, it circumvents the challenges that arise in experimental designs using only a single library and avoids confusing the effects of PCR amplification with the presence of expanded T-cell clones. To not count possible sequencing errors as independent sequences, we rejected single sequences as erroneous if a highly similar clone of greater frequency was identified in the same library (see *Materials and Methods* for the definition of similarity).

The lower bounds on TCRB gene richness obtained with this approach yielded higher estimates than previous studies (Fig. 1). Young adults carried an estimated 60–120 million different TCRB genes, both in the CD4 and CD8 naïve T-cell repertoires. This high diversity in nucleotide sequences was reflected in a large functional repertoire of TCR β chains with a lower boundary of ∼20 million different amino acid sequences. To determine the robustness of our estimates, we used two approaches to estimate confidence intervals. We applied the $BC_a$ variant of bootstrapping that is designed for obtaining confidence intervals when the underlying bootstrap distribution is not symmetric about its center

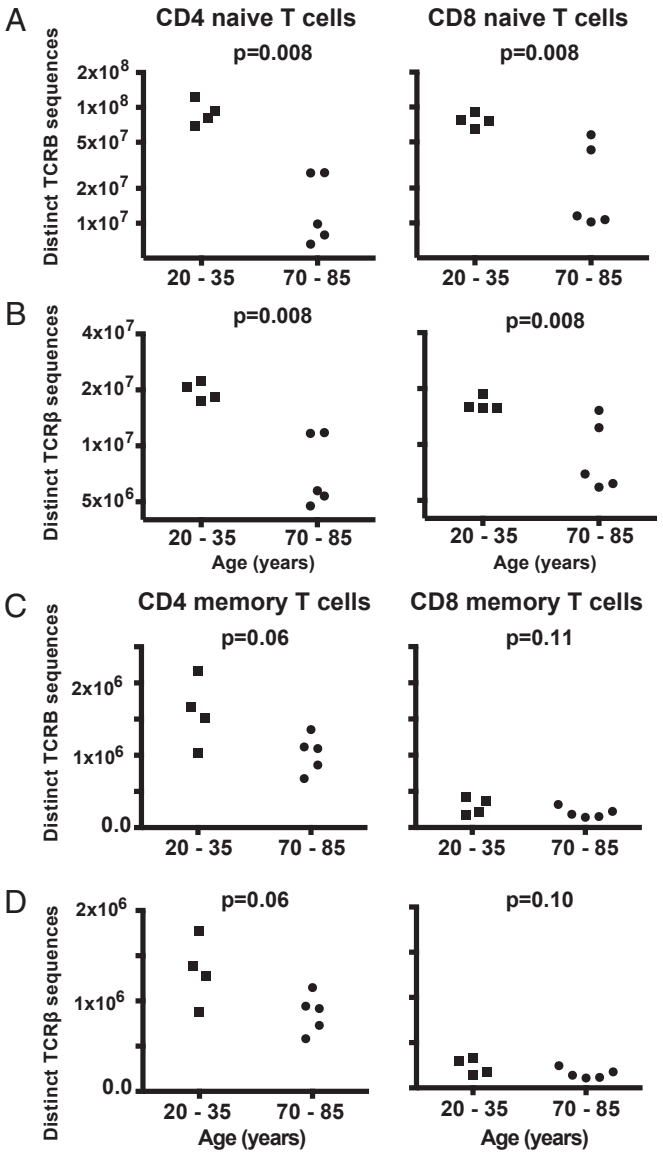


**Fig. 1.** Age is associated with a modest decrease in diversity of the TCRB repertoire. TCRB sequences were obtained from replicate samples of naïve (*A* and *B*) and memory (*C* and *D*) CD4 and CD8 T cells. A lower bound of TCRB richness was estimated by applying nonparametric statistics using the Chao2 estimator. Results are shown for nucleotide (*A* and *C*) and derived amino acid sequences (*B* and *D*). Estimates were compared by Wilcoxon–Mann–Whitney test. Increase in age is associated with a decline in richness of naïve CD4 and CD8 T cells; however, the repertoire in the elderly remains highly diverse. Richness in CD4 and CD8 memory T cells markedly differed, whereas the impact of age was negligibly small.

(15). Second, we estimated the confidence intervals using the approach originally developed by Chao (16). The 95% confidence intervals with both methods were very narrow (Table S2).

Naïve TCRB repertoire richness declined significantly in the 70- to 85-y-old adults to a lower bound richness of 8–57 million different nucleotide sequences encoding ∼5–15 million TCR β-chain amino acid sequences ($P = 0.008$, Fig. 1 *A* and *B*). Interestingly, the estimates in elderly CD4 and CD8 naïve T cells were similar despite the greater decline in CD8 compared with CD4 naïve T-cell numbers with aging (17, 18).

**CD4 and CD8 Memory T Cells Differ in TCRB Richness Independent of Age.** Memory cells have been selected from the naïve repertoire and clonally expanded in response to antigen and are therefore

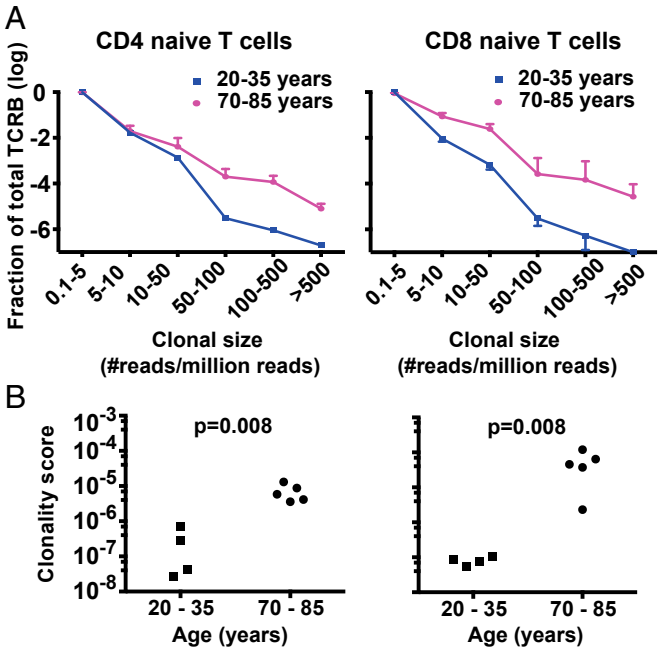


**Fig. 2.** Increase in clonal expansions within naïve CD4 and CD8 T-cell compartments with age. (*A*) The mean number of replicate TCRB sequences was used as an estimate of approximate clonal sizes. The frequency distribution of clonal size bins is shown as (log mean) ± SD of the four young and the five elderly adults. Nondetectable clonal sizes were set at a frequency of 1 in $10^7$. (*B*) Replicate samples of naïve CD4 or naïve CD8 T cells were analyzed for the shared occurrence of TCRB sequences to estimate a clonality score, defined as the probability of two independently identified sequences originating from the same clone. Clonality scores were compared by Wilcoxon–Mann–Whitney test. The observed increase indicate inequality in clonal sizes in the elderly naïve T-cell repertoire with age with increasing number of large clones.

expected to have a lower richness. With advancing age, the number of memory T cells increases and end-differentiated effector CD8 T cells responsive to latent viruses, in particular CMV, accumulate (19, 20). In initial experiments, we compared richness in CD8 central, effector memory, and end-differentiated effector T cells in elderly individuals who did or did not have antibodies to CMV. Richness of effector cell populations was lower than that of central memory T cells and was further compromised in CMV-positive individuals (Fig. S2). To examine the effect of age independent of CMV-induced repertoire changes, we excluded individuals who had positive CMV serology and excluded terminally differentiated effector cells that are known to include clonally expanded effector T cells to latent viruses. Richness in the CD4 memory compartment was about 50-fold lower than in the naïve repertoire, with a lower bound of about 1 million different TCRB genes in each individual. Surprisingly, richness showed negligible contraction with age (Fig. 1 *C* and *D*). CD8 memory T cells were 5- to 10-fold less diverse than CD4 memory T cells, irrespective of age.

**Age-Dependent Clonal Expansion Within the Naïve T-Cell Compartment.** In addition to richness, distributions of clonal size could affect the functionality of the T-cell repertoire. Owing to limitations in sampling ($5 \times 10^6$ naïve T cells out of a total repertoire of up to $10^{12}$ cells per individual), estimates of clonal size distributions are only reliable for clones above a clonal size threshold. In the young as well as the old, most TCR sequences are derived from relatively small clones. In Fig. 2*A*, the frequencies of clones are plotted against clonal sizes (defined as the number of occurrences of identical sequences in the five combined samples of each subset). The frequency distributions for clones above the detection threshold in the repertoires of young and elderly individuals clearly differed, with a small number of naïve T-cell clones showing a striking expansion with age. To compare the contributions of expanded clones to the repertoires, we used a "clonality score" summary metric that is independent of sequencing depth. This metric can be thought of as the probability that two sequence reads selected at random from different replicate library pools will be members of the same clone. Clonality scores in the elderly are approximately >100-fold higher for naïve CD8 T cells and >10-fold higher in CD4 T cells compared with younger individuals (Fig. 2*B*).

Recent studies have identified public TCR sequences that are shared by unrelated individuals (21, 22). Typically, public TCR sequences are those with simple junctions, such as those that lack N-region nontemplated nucleotides (22). To exclude that sequences detected more than once represented coincidentally identical TCR rearrangements rather than clonally expanded T cells, we analyzed sequences present at frequencies of greater than 0.01% in elderly individuals. A TCR sequence was called as public when found in more than one of the nine individuals studied. Twenty-three of the 483 naïve CD4 and 81 of the 878 naïve CD8 frequent TCRB sequences in old individuals were found in more than one individual (Table 1). In contrast, a higher number of 92 of the 483 most frequent CD4 ($P < 0.001$) and 129 of 878 most frequent CD8 TCRB sequences ($P < 0.05$) in the naïve repertoire of young individuals are shared in different individuals. The data indicate that the observed increased clonality in the elderly individuals represent true clonal expansions, whereas many of the apparently clonally expanded sequences in the young repertoires may reflect the presence of simple and public TCR rearrangements.

**Effect of Age on Clonality in the Memory T-Cell Compartment.** In contrast to the age-associated increases in clonality that we observed in naïve CD4 and CD8 cell compartments, there was little effect of age on the clonality of memory CD4 or CD8 T-cell populations. By virtue of memory T cells being present in larger clonal sizes than naïve T cells, clonal size distributions could be examined for a larger fraction of the repertoire. Distributions of clonal sizes were mostly identical for young and elderly individuals and an age-related increase in clonal sizes was seen only for the largest clones (Fig. 3*A*). Accordingly, clonality scores only slightly increased with age (Fig. 3*B*). Notably, CD8 exceeded CD4 memory T cells in clonal sizes.

**Clonally Expanded Naïve T Cells Express TCRB Sequences Distinct from Memory T Cells.** Effector T cells can revert to a phenotype sharing cell surface markers with naïve cells (23). To examine the possibility that the clonal expansions within the naïve compartment represented a contamination with effector T cells masquerading as naïve T cells, we analyzed all TCRB sequences that were found at more than 0.01% frequency in the naïve T-cell population for their representation in the memory population. The reverse analysis was done for clonally expanded sequences in the memory compartments. Results suggested that clonal expansions in the naïve compartment of elderly individuals were not contaminating memory cells. For example, of 121 TCRB sequences

**Table 1. Increased clonality in the naïve T-cell compartment is not caused by an enrichment in public TCRB sequences**

| | CD4 naïve T-cell clones ($n$ = 483) | | CD8 naïve T-cell clones ($n$ = 878) | |
|---|---|---|---|---|
| Age, y | Public | Unique | Public | Unique |
| 20–35 | 92 (19%) | 391 (80.9%) | 129 (14.7%) | 749 (85.3%) |
| 70–85 | 23 (4.8%) | 460 (95.2%) | 81 (9.2%) | 797 (90.8%) |

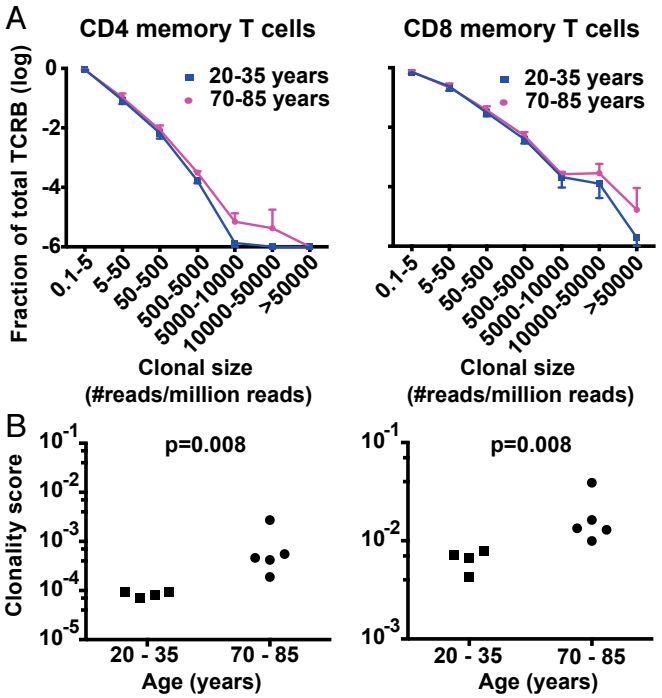


**Fig. 3.** Greater clonality in the repertoire of CD8 than CD4 memory T cells. Frequency distributions of categorized clonal size bins (*A*) and clonality scores (*B*) for memory CD4 and CD8 T cells were determined as described in Fig. 2 for naïve T cells. Nondetectable clonal sizes were set at the minimal frequency of 1 in $10^6$ (*A*). Clonality scores were compared by Wilcoxon–Mann–Whitney test. Large clonal expansions were more frequent in CD8 than in CD4 T cells with only minor influence of age on clonality in both compartments.

clonally expanded in the CD4 naïve T-cell compartment of one individual, 34 were unique for naïve T cells. Conversely, of 920 CD4 memory TCRB sequences, 449 were unique for memory cells. A similar distribution was seen for CD8 T cells: 215 of 379 clones originally identified in the CD8 naïve compartment were uniquely found in that compartment, whereas 349 of 846 memory sequences were not seen at all in naïve CD8 T cells. To further examine whether even the shared clonally expanded populations can be predominantly assigned to one compartment, the frequency of each of these shared sequences in each of the five naïve and five memory CD4 and CD8 T-cell replicate samples was determined. The heat plots in Fig. 4 show the frequency distributions of these clonally expanded CD4 and CD8 TCRB sequences for the different naïve and memory cell samples from two elderly adults. Each row represents one sequence and each lane an independent library of naïve or memory T cells. Although sequences were selected to be present in naïve and memory T cells, very few sequences were found at similar frequencies in both compartments. Importantly, TCRB sequences originally identified in the naïve compartment were virtually absent in memory cell pools, suggesting that these clonal expansions originated from naïve cells and maintained their naïve phenotype.

**Mechanisms of Clonal Expansion Within the Naïve T-Cell Compartment.** A likely interpretation of the increasing inequality of clonal sizes with age is that the homeostatic proliferation of naïve T cells is associated with fitness selection. Clonally expanded T cells might have a competitive growth advantage that in part could be conferred by increased responsiveness to homeostatic cytokines. To test this hypothesis, naïve CD8 T cells labeled with carboxyfluorescein diacetate succinimidyl ester were cultured in the presence of IL-7 and IL-15. Cells that had undergone one or more or two or more divisions after 7 d were purified and sequenced. TCRB sequences from cultured T cells were analyzed for their presence in the replicate libraries of unstimulated naïve T cells. T cells that divided in culture in response to cytokine stimulation show an increased proportion of clones that were expanded in the peripheral blood (Fig. 5). The enrichment was more pronounced for T cells that divided more rapidly. In the population that had divided two or more times, large clones (defined as being found in four replicates of the original blood sample) are enriched compared with small clones (defined as present in one replicate) ($P < 0.001$). In contrast, enrichment was not obvious when T cells purified for high expression of the IL-7 and IL-15 cytokine receptors CD127 and CD215, respectively, were analyzed.

## Discussion

In this study we combined next-generation sequencing with a nonparametric statistical approach using the Chao2 estimator to estimate a lower bound for TCR richness. We found a higher richness in CD4 and CD8 naïve T cells than previous studies. Even though diversity contracts with age, we find that elderly individuals still possess a diverse T-cell repertoire. However, we observe robust clonal expansion with age in the naïve compartments, suggesting that homeostatic proliferation is associated with fitness selection. Finally, we found lower richness in CD8 than in CD4 memory cells, a difference that was preserved during aging.

Thymic involution is the most dramatic age-related change in the human immune system. Understanding whether the T-cell repertoire can be maintained in the absence of thymic activity and whether repertoire contraction contributes to the immune defects in the elderly is critical for designing possible interventions. Conclusions for the human repertoire from animal models are unreliable because the size of the T-cell compartment and mechanisms and kinetics of T-cell homeostasis are fundamentally different in humans and mice (3).

Whether thymic T-cell generation in humans is of any quantitative importance for the steady state of T-cell populations

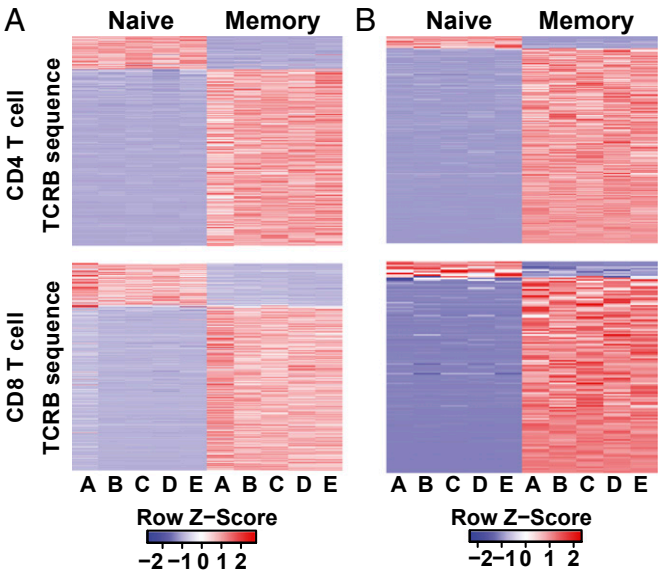


**Fig. 4.** Characterization of clonally expanded naive T cells. TCRB sequences that had frequencies of ≥0.01% and were found in the naïve as well as the memory compartment were analyzed for their frequencies in five independent naïve and memory replicate samples (samples A–E). Representative results of two elderly individuals (one in *A* and the other in *B*) are shown as heat maps using the color scheme from infrequent (blue) to frequent (red). The data indicate that T cells expanded in the naïve and memory compartments are different.

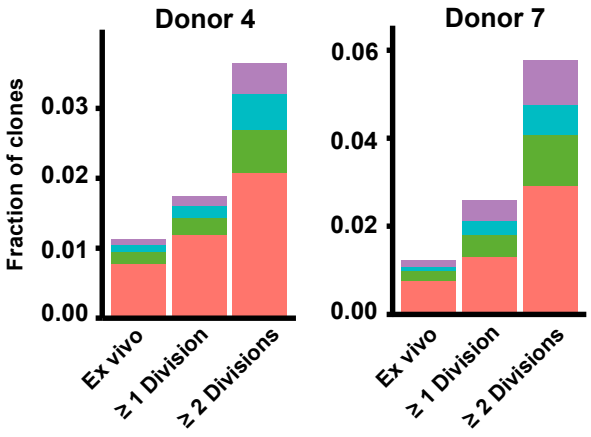


**Fig. 5.** Increased responsiveness of in vivo expanded naïve CD8 T cells to cytokine-induced proliferation. Naïve CD8 T cells were cultured with IL-7 and IL-15. TCRB sequences from naïve CD8 T cells that had divided equal to or more than once or twice were compared with sequences present in the peripheral blood repertoire of the individual. Results from two individuals are shown. The proportions of sequences from the cultured cells that were members of clones detected in four (purple), three (blue), two (green), or one (red) replicates of the original non-cultured T-cell libraries from the peripheral blood are represented as cumulative bar graphs. The fastest-proliferating cultured T cells (right column) show enrichment of large clones found in four replicate libraries from the blood ($P < 0.001$).

after the end of the growth period has been controversial. Some residual thymic tissue persists in elderly humans (24); however, even in the lymphopenic host after chemotherapy or bone marrow transplantation or in HIV patients after initiation of highly active antiretroviral therapy, resurgence of thymic activity does not occur in the majority of individuals older than 40–50 y (25). In the healthy adult, TCR excision circle (TREC) by-products of TCR rearrangement that do not replicate are the best surrogate marker for thymic activity. Mathematical modeling has suggested that the age-related decline in TRECs is best explained by cell loss compensated by homeostatic naïve T-cell turnover rather than by declining thymic efflux. Thymic activity may not play a significant role in the maintenance of TCR diversity in adult life (26).

High-throughput DNA sequencing now enables extensive measurement of TCR populations, but estimation of TCR repertoire richness from sequencing of blood samples has remained a challenge. Here, we combined next-generation sequencing with a nonparametric statistical approach to estimate richness. The use of multiple replicate sequencing libraries generated from independent samples of T cells for each subset enabled us to identify the rare sequences that are critical for estimates of the number of unobserved species and overall TCR richness in the repertoire. We corrected for sequencing errors by excluding sequences that were related to detected peaks in sequence space. In contrast to previously used corrections, our approach does not eliminate essentially all rare clones, but it does still give a conservative appraisal of diversity.

Based on our estimate of 100 million TCRB gene rearrangements and considering that each TCR β-chain in the naïve T-cell repertoire combines with about 25 TCR α-chains (11), the average clonal size of a naïve cell would be about 100–200 cells. This approximation is of the same magnitude as predicted from the dilution of TREC-bearing T cells in the naïve repertoire. In the neonate, naïve T cells proliferate to fill the compartment (27). Recent thymic emigrants in the newborn undergo three to four divisions in the periphery after TCRA gene rearrangement, thereby establishing a minimal clonal size of 10 cells (28). TREC frequencies decline further in subsequent years, suggesting that naïve T cells undergo approximately three more divisions during the growth period in the presence of thymic activity, to give a final average naïve T-cell clonal size of about 100 cells.

Our analysis does not address directly whether a decline from about 20 million to 10 million TCR β-chain sequences leads to immune defects or holes in the repertoire in the elderly. Tetramer studies in noninfected individuals have suggested that the frequency of T cells specific for many exogenous viral peptides, or self-peptides, is on the order of 1 in 1 million (29), suggesting that the estimated elderly repertoires of more than 100 million TCR α-β dimers could be diverse enough to recognize most peptides bound to a given MHC molecule and avoid frank "holes" in the repertoire.

A striking age-related change in our studies was an increase in T-cell clonality in the naïve repertoire. The contribution of clonally expanded T cells to the observed repertoire increased by a factor of >100 for naïve CD8 and of >10 for naïve CD4 T cells compared with young adults. Analysis of these expanded clones showed that the clones in the naïve subsets were distinct from those in the memory compartments, although many of them were present in clonal sizes that were comparable to those of memory cells. If these naïve clones derive from uneven homeostatic proliferation, they maintain a naïve phenotype, in contrast to findings in mice, in which naïve T cells undergoing homeostatic proliferation tend to acquire a memory-like phenotype (30, 31).

Based on the limited number of cells that can be sampled in humans, our clonality index only allows conclusions on the impact of clones that are above a certain size. However, for naïve T cells age-associated increases were observed for the entire range of clonal sizes that could be assessed. This observation is consistent with the interpretation that homeostatic proliferation during aging is associated with increasing unevenness in clonal size distributions of the entire repertoire owing to peripheral selection. In support of this interpretation, we have recently performed an agent-based stochastic in silico simulation of repertoire complexity under homeostatic proliferation and have found only a minimal contraction in diversity over a human lifetime, as long as peripheral fitness selection during homeostatic proliferation can be ignored; even with no thymic activity after age 20 y and 90% shrinkage of the naïve T-cell compartment size (32). Depending on the initial clonal size, increasing variance in clonal abundances will not necessarily lead to clonal extinction. In our data, naïve CD8 T cells had a much higher clonality score than naïve CD4 T cells; however, richness in both compartments was very similar, suggesting a higher variance in clonal sizes in CD8 T cells (Figs. 1 and 2). A large number of homeostatic T-cell divisions or, as explored in our in silico simulation, a rather dramatic co-occurrence of selective forces is required before homeostatic proliferation leads to contraction in richness. However, the inequalities in clonal sizes with some large and many very small clones may generate a less complex T-cell response to peptide antigens, which may be more important than the fewer TCRs available in the repertoires of older compared with younger individuals. Such an unevenness cannot be improved by restoring thymic T-cell generation (32).

Age-associated skewing in clonal size distributions explains why some prior studies have observed a higher decline in diversity with aging. Because sequencing depths in most previous studies were not sufficient and/or the analytical design did not take into account increasing inequality in clonal size distributions, estimates of richness were not reliable. Based on this model, our previous report of a stable TCR repertoire up to the seventh decade of life with a sudden change thereafter indicates an occurrence of increasing T-cell clonal expansions late in life (33). This interpretation is also consistent with observations in a non-human primate model (34).

Homeostatic proliferation and survival of naïve T cells are thought to require the recognition of self-antigen in the presence

of homeostatic cytokines (30, 35). Inequalities in clonal sizes could therefore reflect a peripheral selection based on cytokine responsiveness and/or self-recognition. In support of this interpretation, in vitro culture of naïve T cells in the presence of homeostatic cytokines, but absence of foreign antigens, selected for TCR sequences clonally expanded in vivo. Homeostatic proliferation associated with a peripheral selection may therefore result in a more autoreactive repertoire (36). This may explain why lymphopenia-induced homeostatic proliferation confers an increased risk for autoimmunity (37, 38) and supports the model that increased homeostatic proliferation and associated changes in the repertoire found in patients with rheumatoid arthritis may predispose for the disease (39–41).

Surprisingly, and in contrast to naïve T cells, the impact of age on the memory T-cell repertoire was minimal. Although the size of the memory compartment is known to increase with age, richness did not change significantly, and clonality increased only for the very largest clones. This observation is particularly surprising for CD8 T cells where the terminally differentiated effector T-cell population tends to increase at the expense of central and effector memory T cells (17). It should be noted, however, that we have excluded individuals who were CMV-positive to analyze age effects without the confounding factor of CMV-induced repertoire changes.

Our data highlight striking differences in the repertoires of CD4 and CD8 T memory cells, independent of age. Richness was higher in CD4 memory than in CD8 T cells, whereas clonality was higher in CD8 memory T cells. The prominent clonality in CD8 memory T cells is consistent with previous spectratyping studies demonstrating clonal peaks in the CD8 memory repertoire (42). Our studies now show that the difference between human CD4 and CD8 memory T cells is not limited to clonally expanded CD8 T cells, but also includes a globally decreased richness of the entire CD8 memory repertoire. Given that richnesses in naïve CD4 and CD8 T cells are approximately equal, our data suggest that formation or maintenance of the memory repertoire is more constrained for CD8 than for CD4 T cells. Future studies of the repertoires of antigen-specific T cells after vaccination or infection in humans will be required to further explore the consequences of T-cell subset-specific repertoire contractions and the contribution of clonal expansions to the increased vulnerability of the elderly to common pathogens.

## Materials and Methods

TCRB cDNA libraries were generated from five replicate samples of FACS-sorted naïve and memory CD4 and CD8 T cells from apheresis samples of young and elderly healthy adults and sequenced with an Illumina Miseq sequencer. TCRB repertoire richness was determined by applying the Chao2 nonparametric estimator of the lower bound of species richness in a population after correcting for possible sequencing errors and eliminating TCRB sequences that are close to peaks in sequence space. A clonality score, adapted from the Gini–Simpson index, was determined using the lymphclon inference algorithm. A detailed description of the experimental design and procedures and the statistical analysis is given in *SI Materials and Methods*. Primers are described in Table S3.

**ACKNOWLEDGMENTS.** We thank Dr. A. Chao, National Tsing Hua University, for statistical advice and helpful discussions. This work was supported by National Institutes of Health (NIH) Grants U19 AI090019, R01 AG015043, U19 AI057266, and R01 AI108891 (to J.J.G.); NIH Grant U19 AI090019 and a New Scholar in Aging grant from the Ellison Medical Foundation (to S.D.B.); NIH Grants R01 AR042527, R01 AI044142, and P01 HL058000 (to C.M.W.); and an American Federation for Aging Research Postdoctoral Fellowship in Aging Research (to Q.Q.).

1. Palmer DB (2013) The effect of age on thymic function. *Front Immunol* 4:316.
2. Taub DD, Longo DL (2005) Insights into thymic aging and regeneration. *Immunol Rev* 205:72–93.
3. den Braber I, et al. (2012) Maintenance of peripheral naive T cells is sustained by thymus output in mice but not humans. *Immunity* 36(2):288–297.
4. Nikolich-Zugich J (2008) Ageing and life-long maintenance of T-cell subsets in the face of latent persistent infections. *Nat Rev Immunol* 8(7):512–522.
5. Jean CM, Honarmand S, Louie JK, Glaser CA (2007) Risk factors for West Nile virus neuroinvasive disease, California, 2005. *Emerg Infect Dis* 13(12):1918–1920.
6. Peiris JS, et al.; HKU/UCH SARS Study Group (2003) Clinical progression and viral load in a community outbreak of coronavirus-associated SARS pneumonia: A prospective study. *Lancet* 361(9371):1767–1772.
7. Goodwin K, Viboud C, Simonsen L (2006) Antibody response to influenza vaccination in the elderly: A quantitative review. *Vaccine* 24(8):1159–1169.
8. Cicin-Sain L, et al. (2010) Loss of naive T cells and repertoire constriction predict poor response to vaccination in old primates. *J Immunol* 184(12):6739–6745.
9. Ahmed M, et al. (2009) Clonal expansions and loss of receptor diversity in the naive CD8 T cell repertoire of aged mice. *J Immunol* 182(2):784–792.
10. Yager EJ, et al. (2008) Age-associated decline in T cell repertoire diversity leads to holes in the repertoire and impaired immunity to influenza virus. *J Exp Med* 205(3):711–723.
11. Arstila TP, et al. (1999) A direct estimate of the human alphabeta T cell receptor diversity. *Science* 286(5441):958–961.
12. Warren RL, et al. (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(5):790–797.
13. Robins HS, et al. (2009) Comprehensive assessment of T-cell receptor beta-chain diversity in alphabeta T cells. *Blood* 114(19):4099–4107.
14. Chao A, Bunge J (2002) Estimating the number of species in a stochastic abundance model. *Biometrics* 58(3):531–539.
15. DiCiccio TJ, Efron B (1996) Bootstrap confidence intervals. *Stat Sci* 11(3):189–228.
16. Chao A (1987) Estimating the population size for capture-recapture data with unequal catchability. *Biometrics* 43(4):783–791.
17. Czesnikiewicz-Guzik M, et al. (2008) T cell subset-specific susceptibility to aging. *Clin Immunol* 127(1):107–118.
18. Wertheimer AM, et al. (2014) Aging and cytomegalovirus infection differentially and jointly affect distinct circulating T cell subsets in humans. *J Immunol* 192(5):2143–2155.
19. Weekes MP, Carmichael AJ, Wills MR, Mynard K, Sissons JG (1999) Human CD28-CD8+ T cells contain greatly expanded functional virus-specific memory CTL clones. *J Immunol* 162(12):7569–7577.
20. Newell EW, Sigal N, Bendall SC, Nolan GP, Davis MM (2012) Cytometry by time-of-flight shows combinatorial cytokine expression and virus-specific cell niches within a continuum of CD8+ T cell phenotypes. *Immunity* 36(1):142–152.
21. Robins HS, et al. (2010) Overlap and effective size of the human CD8+ T cell receptor repertoire. *Sci Transl Med* 2(47):47ra64.
22. 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.
23. Berard M, Tough DF (2002) Qualitative differences between naïve and memory T cells. *Immunology* 106(2):127–138.
24. Poulin JF, et al. (1999) Direct evidence for thymic function in adult humans. *J Exp Med* 190(4):479–486.
25. Hakim FT, et al. (2005) Age-dependent incidence, time course, and consequences of thymic renewal in adults. *J Clin Invest* 115(4):930–939.
26. Hazenberg MD, Borghans JA, de Boer RJ, Miedema F (2003) Thymic output: A bad TREC record. *Nat Immunol* 4(2):97–99.
27. Min B, et al. (2003) Neonates support lymphopenia-induced proliferation. *Immunity* 18(1):131–140.
28. Schönland SO, et al. (2003) Homeostatic control of T-cell generation in neonates. *Blood* 102(4):1428–1434.
29. Su LF, Kidd BA, Han A, Kotzin JJ, Davis MM (2013) Virus-specific CD4(+) memory-phenotype T cells are abundant in unexposed adults. *Immunity* 38(2):373–383.
30. Boyman O, Krieg C, Homann D, Sprent J (2012) Homeostatic maintenance of T cells and natural killer cells. *Cell Mol Life Sci* 69(10):1597–1608.
31. Akue AD, Lee JY, Jameson SC (2012) Derivation and maintenance of virtual memory CD8 T cells. *J Immunol* 188(6):2516–2523.
32. Johnson PL, Yates AJ, Goronzy JJ, Antia R (2012) Peripheral selection rather than thymic involution explains sudden contraction in naive CD4 T-cell diversity with age. *Proc Natl Acad Sci USA* 109(52):21432–21437.
33. Naylor K, et al. (2005) The influence of age on T cell generation and TCR diversity. *J Immunol* 174(11):7446–7452.
34. Cicin-Sain L, et al. (2007) Dramatic increase in naive T cell turnover is linked to loss of naive T cells from old primates. *Proc Natl Acad Sci USA* 104(50):19960–19965.
35. Rudd BD, et al. (2011) Nonrandom attrition of the naive CD8+ T-cell pool with aging governed by T-cell receptor:pMHC interactions. *Proc Natl Acad Sci USA* 108(33):13694–13699.
36. Goronzy JJ, Weyand CM (2012) Immune aging and autoimmunity. *Cell Mol Life Sci* 69(10):1615–1623.
37. Datta S, Sarvetnick N (2009) Lymphocyte proliferation in immune-mediated diseases. *Trends Immunol* 30(9):430–438.
38. Jones JL, et al. (2013) Human autoimmunity after lymphocyte depletion is caused by homeostatic T-cell proliferation. *Proc Natl Acad Sci USA* 110(50):20200–20205.
39. Wagner UG, Koetz K, Weyand CM, Goronzy JJ (1998) Perturbation of the T cell repertoire in rheumatoid arthritis. *Proc Natl Acad Sci USA* 95(24):14447–14452.
40. Goronzy JJ, Weyand CM (2001) Thymic function and peripheral T-cell homeostasis in rheumatoid arthritis. *Trends Immunol* 22(5):251–255.
41. Koetz K, et al. (2000) T cell homeostasis in patients with rheumatoid arthritis. *Proc Natl Acad Sci USA* 97(16):9203–9208.
42. Nikolich-Zugich J, Slifka MK, Messaoudi I (2004) The many important facets of T-cell repertoire diversity. *Nat Rev Immunol* 4(2):123–132.