Age-Related Decrease in TCR Repertoire Diversity Measured with Deep and Normalized Sequence Profiling

Olga V. Britanova, Ekaterina V. Putintseva, Mikhail Shugay, Ekaterina M. Merzlyak, Maria A. Turchaninova, Dmitriy B. Staroverov, Dmitriy A. Bolotin, Sergey Lukyanov, Ekaterina A. Bogdanova, Ilgar Z. Mamedov, Yuryi B. Lebedev and Dmitriy M. Chudakov

*J Immunol* 2014; 192:2689-2698; Prepublished online 7 February 2014; doi: 10.4049/jimmunol.1302064
http://www.jimmunol.org/content/192/6/2689

**Supplementary Material**
http://www.jimmunol.org/content/suppl/2014/02/07/jimmunol.1302064.DCSupplemental

**References**
This article cites 58 articles, 26 of which you can access for free at: http://www.jimmunol.org/content/192/6/2689.full#ref-list-1

**Subscription**
Information about subscribing to *The Journal of Immunology* is online at: http://jimmunol.org/subscription

**Permissions**
Submit copyright permission requests at: http://www.aai.org/About/Publications/JI/copyright.html

**Email Alerts**
Receive free email-alerts when new articles cite this article. Sign up at: http://jimmunol.org/alerts
Age-Related Decrease in TCR Repertoire Diversity Measured with Deep and Normalized Sequence Profiling

Olga V. Britanova,*† Ekaterina V. Putintseva,*† Mikhail Shugay,*†,‡ Ekaterina A. Merzlyak,* Maria A. Turchaninova,* Dmitriy B. Staroverov,* Dmitriy A. Bolotin,* Sergey Lukyanov,*† Ekaterina A. Bogdanova,* Ilgar Z. Mamedov,* Yuriy B. Lebedev,* and Dmitriy M. Chudakov*†

The decrease of TCR diversity with aging has never been studied by direct methods. In this study, we combined high-throughput Illumina sequencing with unique cDNA molecular identifier technology to achieve deep and precisely normalized profiling of TCR β repertoires in 39 healthy donors aged 6–90 y. We demonstrate that TCR β diversity per 10⁶ T cells decreases roughly linearly with age, with significant reduction already apparent by age 40. The percentage of naive T cells showed a strong correlation with measured TCR diversity and decreased linearly up to age 70. Remarkably, the oldest group (average age 82 y) was characterized by a higher percentage of naive CD4+ T cells, lower abundance of expanded clones, and increased TCR diversity compared with the previous age group (average age 62 y), suggesting the influence of age selection and association of these three related parameters with longevity. Interestingly, cross-analysis of individual TCR repertoires revealed a set >10,000 of the most representative public TCR β clonotypes, whose abundance among the top 100,000 clones correlated with TCR diversity and decreased with aging. The Journal of Immunology, 2014, 192: 2689–2698.

In humans, aging is associated with prominent decline of the adaptive immune response, caused by both altered functionality of aged T cells (1, 2) and overall decrease in naive T cell abundance and TCR diversity. The latter results from thymus involution (3, 4); stochasticity, selectivity, and exhaustion of peripheral proliferation of naive T cells (5–9); and expansion of effector memory clones that take over the homeostatic space (6, 10–13).

It is well acknowledged that high TCR repertoire diversity is a necessary prerequisite for an effective adaptive immune response against new Ags (6). Loss of naive T cells and constriction of the TCR repertoire result in impaired immunity to viral and bacterial infections (6, 14–16), poor response to vaccination (17, 18), poor immune system recovery after chemotherapy and hematopoietic stem cell transplantation (19–21), and poor control of autoreactive T clones and autoimmunity (22–24).

The phenomenon of aging-associated TCR repertoire constriction, although widely accepted, has never been studied by direct methods. Oligonucleotide hybridization assays (25) and spectratyping analysis (26) have suggested that early and midadulthood methods. Oligonucleotide hybridization assays (25) and spectratyping analysis (26) have suggested that early and midadulthood are not associated with significant contraction of the TCR repertoire and that this occurs only after the age of 75 (25, 27).

In this study, we have used an advanced deep TCR β sequencing approach coupled with a sample normalization strategy that allowed us to directly quantify and compare T cell repertoire diversity across samples obtained from numerous individuals of different ages. We achieved deep and accurate profiling of individual TCR β CDR3 repertoires by performing unique barcoding of cDNA molecules (28), followed by paired-end Illumina sequencing and rational bioinformatic analysis of the output data for precise sample normalization (29, 30). This approach enabled us to track age-related changes in the human TCR repertoire with unprecedented accuracy, revealing key parameters of adaptive immunity aging and providing high-quality reference data for future studies of adaptive immunity in health and disease.

Materials and Methods

Sample collection

This study was approved by the local ethics committee and conducted in accordance with the Declaration of Helsinki. All donors were informed of the final use of their blood and signed an informed consent. Ten milliliters of peripheral blood was obtained from 39 systemically healthy Caucasian donors from Russia, aged 6–90 y old (Table I). Peripheral blood was collected into EDTA-treated Vacutainer tubes (BD Biosciences, Franklin Lakes, NJ). PBMCs (at least 7 × 10⁶/sample) were isolated by Ficoll-Paque (Paneco, Russia) density-gradient centrifugation. Total RNA (at least 6 µg/sample) was isolated using Trizol reagent (Invitrogen, Carlsbad, CA), according to the manufacturer’s protocol.

cDNA synthesis

First-strand cDNA was synthesized using the Mint cDNA synthesis kit (Evrogen, Russia), according to the manufacturer’s protocol (Supplemental Fig. 1). RNA (1.5 µg) was collected per 15-µl reaction volume, with a total of four to six tubes per sample, such that all RNA was used for each
blood sample. Synthesis was primed with the BC_R4_short oligonucleotide (5’-GTATCGGATTGCACTG-3’), which is specific to both gene variants of the human TRBC segment. To denature RNA and anneal the priming oligonucleotide, the RNA was incubated at 70°C for 2 min and then at 42°C for 2 min. The SmartMMNNNu adapter (5’-AACGAGUG-GTAUACAGCGAGAUGNUNUNUNUNUNUNUTTGrGrGrG-3’), which carries a molecular identifier (12 random “N” nucleotides) and dU nucleotides (U), was added for the template switch. This reaction was carried out at 42°C for 130 min, with 5 μl IP solution added after the first 40 min. Products of cDNA synthesis were treated with fresh uracil-DNA glycosylase (New England Biolabs, Ipswich, MA) to degrade the SmartMMNNNu adapter (5 U/1 μl reaction) at 37°C for 30 min. To capture the maximum number of input cDNA molecules, the whole volume of the treated cDNA synthesis reaction was used for the first PCR.

**First PCR**

The first PCR amplification was performed using Encyclo PCR mix (Evoget) with a pair of universal primers: M1SS (5’-AAGCGATGTTG-ATCAACGGCA-3’) and BC2_uni_R (5’-TGTTCTTGTAGGTCTACAAAC-3’). The PCR mixture contained 1× Encyclo polymerase buffer, 0.125 mM of each 2’-deoxyxynucleotide 5’-phosphate, 10 pmol of each primer, 1 μl Encyclo polymerase mix, and 3 μl of undiluted first-strand cDNA/50 μl reaction volume. The reaction was performed on an ABI 9700 Thermal Cycler with a gold block in multiple 0.2 ml PCR tubes, such that all cDNA was used (up to 32 tubes/sample), for 18 cycles with the following temperature regimen: 94°C for 10 s, 62°C for 20 s, and 72°C for 30 s.

**Second PCR**

Sample products from the first PCR amplification were combined into one tube and mixed. A 100-μl aliquot of undiluted first PCR product was purified by QIAquick PCR purification kit (Qiagen) and eluted in 20 μl TE buffer. For the second PCR amplification, we used the M1S primer ((N) 2–4((NT) 9(GAGTGGTATCAACGCAGAG); 50 μl reaction) at 37°C for 30 min. To capture the maximum number of input cDNA molecules, the whole volume of the treated cDNA synthesis reaction was used for the first PCR.

**Sequencing**

The PCR product concentration in each library was determined using a QuBit fluorometer (Invitrogen). PCR products of 9–20 samples from different donors were mixed together in an equal ratio. To avoid experimental bias, donors of different age were randomly distributed across the three lanes. Illumina adapters were ligated according to the manufacturer’s protocol. The libraries were analyzed using three Illumina HiSeq2000 lanes, with 100 + 100-nt paired-end sequencing with Illumina primers.

**Raw sequencing data analysis**

Preliminary analysis showed that at least 3 × 10^6 CDR3-containing paired-end sequencing reads were obtained for each sample. From those, we randomly selected 1 × 10^6 high-quality sequencing reads so that only one read was present with a given unique molecular identifier. No binning of CDR3 sequences with identical molecular identifiers was performed, and reads with duplicate molecular identifiers were filtered out. Thus, for each donor we obtained sequences of 1 × 10^6 unique TCR β cDNA molecules. To check for possible effects of this random selection, we performed a repetitive sampling of molecular identifiers that yielded no change in our overall results, including sample diversity (data not shown). Further analysis (final CDR3 identification, clonotype clustering, and correction for reverse transcription, PCR, and sequencing errors) was performed as previously described (29) using our MITCR software (30) (http://mitcr.milaboratory.com/). The threshold on sequencing quality for each nucleotide within the CDR3 region was set as > Q25 (phred). To eliminate a maximal number of accumulated errors, we employed the strictest “eliminate these errors” correction algorithm, which eliminates 98% of artificial CDR3 clonotypes, including variants that arise from rare errors in minor clonotypes. At the same time, this algorithm loses <2% of input TCR β diversity, according to control experiments and in silico modeling.

**Lower-bound estimate of total TCR β CDR3 diversity in blood**

The total number of unique TCR β clonotypes was estimated as in Ref 32, using a nonparametric unseen species model as described previously (33). Briefly, the Poisson process assumption, which relates the expected number of times a species is going to appear in the sample (x) and the sampling depth (t), leads to an estimator for total number of species in the form of an infinite series: where x is the number of species found exactly x times. Unfortunately, these series begin to severely oscillate with increasing sampling depth, which renders it impossible to estimate total diversity. This problem can be dealt with using Euler transformation, which forces series to converge rapidly. We therefore calculated the first x0 terms in modified series, where x0 is the largest number that yields a coefficient of variation for estimating S(t) that is < 0.1. This provides a tradeoff between bias and variance, as suggested by Efron et al. (33).

**Abs and flow cytometry**

An aliquot of PBMC from each sample was stained with Abs for flow cytometry analysis. The following anti-human Abs were used: CD3-Pe7 (clone UCHT1; eBioscience), CD27-PE5 (clone 1A4CD27; Invitrogen), CD4-PE (clone 13B8.2; Beckman Coulter), CD5RA FITC (clone JS-B3; eBioscience). Cells were incubated with Abs for 20 min at room temperature and washed twice with PBS. Flow cytometry was performed with the Cytomix FC 500 (Beckman Coulter), with data analysis carried out using the Cytomix RXP Analysis program (Beckman Coulter).

**Results**

**Normalization of samples using unique molecular identifiers**

Blood samples from different donors—even sequential blood samples from the same donor—may contain different numbers of T cells. Furthermore, none of the TCR gene library preparation and sequencing procedures can be made absolutely uniform, regardless of whether genomic DNA-based (32, 34) or cDNA-based (29, 31, 35) technique is used. Therefore, one can never be sure of the quantity of T cells that are efficiently covered by the output of sequencing analysis of large TCR libraries. This makes it impossible to perform accurate and reproducible deep sequencing-based comparison of relative TCR diversity in blood samples obtained from different donors or at different times.

To overcome this basic limitation, we combined a cDNA-based protocol for preparing quantitative TCR β libraries (31, 35) with a unique molecular barcoding technique similar to that reported in Ref 28. In this technique, the template-switching effect (36, 37) is used to introduce a 5’-adaptor that carries 12 random nucleotides. As a result, each synthesized cDNA molecule is specifically labeled with one of 4^12 (>16.7 million) unique identifier variants. Such molecular identifiers allow robust estimation of the number of cDNA templates in a deeply sequenced library, which cannot be accurately deduced from read data alone (38). The whole TCR β library was further amplified using universal primers (see Materials and Methods and Refs. 29 and 31) and analyzed using paired-end Illumina sequencing, which provided information about both the TCR β CDR3 sequence and the unique molecular identifier of the starting cDNA molecule (Supplemental Fig. 1).

This method allows normalization of any two or more libraries, even if they were obtained from different numbers of T cells and with different sequencing coverage, by analyzing equal numbers of unique cDNA molecules labeled with different molecular identifiers. Because each read represents a distinct cDNA molecule, such normalization eliminates systematic library preparation biases, resulting in a prominent reduction of variance between samples and a notable increase in detectable TCR β diversity per fixed number of analyzed sequencing reads (Fig. 1). This normalization also reduces biases within the sample, because of the stochastic and biased nature of PCR and sequencing, cDNA variants with the same abundance may have sequencing read counts that differ by severalfold. Sampling equal numbers of reads will therefore lead to some variants being overrepresented and some lost, especially the low-abundance variants that primarily contribute to sample diversity. In contrast, extraction of sequences with unique cDNA identifiers allows uniform sampling and accurate comparison of the samples.
Model experiments demonstrated that the average efficiency of the template-switch technique that we used is about one molecular event per two T cells. For example, deep sequencing analysis of libraries prepared from \(2 \times 10^9\) or \(1 \times 10^9\) T cells produced output data comprising \(1 \times 10^6\) or \(5 \times 10^5\) uniquely barcoded variants, respectively. Therefore, each analyzed TCR \(\beta\) cDNA labeled with a unique molecular identifier is equivalent to a single randomly captured T cell, and 1 million such molecular events is equivalent to 1 million T cells randomly captured from a 10 ml peripheral blood sample.

This powerful tool for normalizing analyzed sequencing datasets enabled us to perform deep and accurate comparison of TCR \(\beta\) CDR3 diversity between samples with sufficient but not necessarily equal numbers of T cells or levels of sequencing coverage. Furthermore, this approach minimized the impact of individual donor blood characteristics or unavoidable bottlenecks and biases during blood sampling, library preparation, and sequencing.

**Observed TCR \(\beta\) CDR3 diversity declines throughout life**

Our 39 systemically healthy human donors were divided into four age groups representing young (average 16 y), middle-aged (average 39 y), aged (average 62 y), and long-lived (average 82 y) donors, respectively (Table I).

For each donor, a molecular-barcoded cDNA-based TCR \(\beta\) library was generated and analyzed by deep paired-end Illumina sequencing. We processed the output for normalization via unique molecular identifiers, with further clonotype clustering and error correction using MiTCR software (30) as described in **Materials and Methods**. The resulting normalized datasets, which are available at http://mitcr.milaboratory.com/datasets/, contained full quantitative information on the clonal composition and diversity of TCR \(\beta\) CDR3 repertoires per \(10^6\) T cells from each donor.

The observed diversity per \(10^6\) T cells decreased almost linearly with age \((R = -0.75, p = 4 \times 10^{-3})\), ranging from \(7.8 \times 10^5\) clones for the youngest down to \(1.3 \times 10^5\) clones for one of the aged donors. We observed a significant decrease in the number of unique TCR \(\beta\) clonotypes between young and middle-age \((p < 0.01, 1.3\text{-fold})\) and between middle-age and aged \((p < 0.01, 1.6\text{-fold})\) groups. However, a nonsignificant increase in TCR diversity between the aged and long-lived groups (Fig. 2A, Table I) was observed.

Notably, analysis of \(1 \times 10^6\) randomly chosen CDR3-containing sequencing reads without using molecular identifiers distinguished poorly between different age cohorts \((p > 0.05, \text{two-tailed} t\text{-test} \text{for age groups 1 and 2})\). This is most likely because group 1 is characterized by the most diverse repertoire, whereas read-based data generally underestimate sample diversity (see Fig. 1, Supplemental Fig. 2A and below).

We analyzed absolute T cell counts in blood for selected donors, and observed no significant age-related changes (data not shown), in agreement with previous data showing either no significant change or only a marginal decrease in either absolute count or proportion of CD3+ cells with age (39–41). Thus, we believe that the observed changes in the TCR \(\beta\) repertoire generally refer both to the diversity per 1 million T cells and to the absolute T cell counts, albeit with a possible bias in the latter due to individual variance in T cell counts in the blood.

### Table I. Characteristics of donors by age group

<table>
<thead>
<tr>
<th>Group No.</th>
<th>Average Age (Range), y</th>
<th>No. of Patients (Female/Male)</th>
<th>Observed TRBV CDR3 Diversity per (10^6) T Cells, (\times 10^3)</th>
<th>Lower Bound for Total TRBV CDR3 Diversity, (\times 10^3)</th>
<th>% CD3 Naive, Average, <strong>&quot;</strong></th>
<th>% CD4 Naive, Average, <strong>&quot;</strong></th>
<th>% CD8 Naive, Average, <strong>&quot;</strong></th>
<th>CD4/CD8 T Cells Ratio, Average, <strong>&quot;</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Group 1, young</td>
<td>16 (6–25)</td>
<td>11 (4/7)</td>
<td>6.4 ± 1.0</td>
<td>5.0 ± 1.3</td>
<td>54.3 ± 10.3</td>
<td>49.8 ± 9.6</td>
<td>59.7 ± 14.0</td>
<td>1.38 ± 0.49</td>
</tr>
<tr>
<td>Group 2, middle-age</td>
<td>39 (34–43)</td>
<td>10 (6/4)</td>
<td>5.0 ± 1.3</td>
<td>4.1 ± 1.0</td>
<td>35.2 ± 9.4</td>
<td>33.7 ± 12.4</td>
<td>37.5 ± 8.1</td>
<td>1.97 ± 0.56</td>
</tr>
<tr>
<td>Group 3, aged</td>
<td>62 (61–66)</td>
<td>7 (4/3)</td>
<td>3.1 ± 1.0</td>
<td>2.4 ± 0.5</td>
<td>17.3 ± 8.1</td>
<td>19.9 ± 7.9</td>
<td>18.8 ± 12.6</td>
<td>2.45 ± 0.95</td>
</tr>
<tr>
<td>Group 4, long-lived</td>
<td>82 (71–90)</td>
<td>11 (5/6)</td>
<td>3.6 ± 0.9</td>
<td>2.6 ± 0.2</td>
<td>27.5 ± 14.8</td>
<td>34.6 ± 20.9</td>
<td>11.2 ± 6.1</td>
<td>2.54 ± 1.08</td>
</tr>
</tbody>
</table>
Estimating the lower bound for total individual TCR β CDR3 diversity

Analysis of species accumulation curves (Fig. 2B) showed that the number of TCR β clonotypes detected in each donor sample increased proportionally with the number of cDNA molecules analyzed, with an almost direct relation. The species accumulation curves were similar within age-groups but differed between them. For young individuals, each additional 100,000 cDNA molecules analyzed consistently yielded \( \sim 60,000 \) additional TCR β CDR3 clonotype variants. In general, most curves remained far from saturation, confounding estimation of total TCR β diversity.

To estimate the lower bound of total TCR β diversity for the studied donors, we applied an unseen species model (see Materials and Methods). For young cohort (group 1), the estimated lower bound on total TCR β diversity was \( \sim 7 \times 10^6 \) different clonotypes, and this estimate declined to \( \sim 4 \times 10^6 \) for middle-aged individuals. In group 3, the lower bound was \( \sim 2.4 \times 10^6 \), significantly lower than that of group 2 (1.6-fold, \( p = 0.001 \)) (Fig. 2C).

It should be noted that this is only the lower bound estimate, which would be predicted to increase with greater sampling depth. To verify this, we further analyzed two independent replicas consisting of \( 1 \times 10^6 \) unique cDNA reads each from two donors from our study, aged 25 and 87, whose blood samples contained 44 and 15% naive T cells, respectively (starting from the same blood draw, cells separated at the level of purified PBMC). These additional replicas, which were added to the full analysis pipeline starting from joint reads, increased the directly observed diversity 1.91- and 1.74-fold, and increased the lower-bound estimate for total TCR β diversity 2.3 \( \times \) 0.4- and 2.4 \( \pm \) 0.1-fold for young and long-lived donors, respectively, confirming that determination

**FIGURE 2.** Age-dependent trends in TCR repertoire diversity. (A) The number of unique TCR β CDR3 clonotypes per \( 10^6 \) T cells is shown as a function of donor age. Observed diversity declines roughly linearly with age (\( R = -0.75 \)). The trend is significant (\( p < 0.0001 \); Kruskal–Wallis test). **\( \times \)** \( p < 0.01 \); two-tailed \( t \) test. (B) Species accumulation curves for TCR β clonotypes for all four age groups. We analyzed subsets of a sample of \( 10^6 \) cDNA molecules in increments of \( 1 \times 10^5 \). Plots show the average number of unique clonotypes obtained from random samplings (\( n = 5 \)) of cDNA molecules from each donor. (C) Estimates of the lower bound of total TCR β diversity in human donors. See Materials and Methods for details.
of the lower bound of total TCR β diversity in an individual based on a sample of $1 \times 10^6$ sequenced T cells is still essentially an underestimate.

**Expanding clones occupy homeostatic space**

Expanded TCR β clonotypes—defined as those that constituted $>0.001$, $>0.01$, or $>1\%$ of $10^6$ analyzed T cells—consistently occupied significantly greater homeostatic space in the peripheral blood in an age-dependent fashion when we compared young, middle-aged and aged donors (groups 1, 2, and 3). The percentage of T cells represented by single-cell clonotypes decreased from $\sim 55\%$ down to $24\%$ in these cohorts, with significant inverse correlation with age. Notably, these trends did not extend to group 4, which was characterized by a lower abundance of expanded clonotypes and higher prevalence of low-frequency clonotypes compared with group 3 (Fig. 3).

**Observed TCR β diversity correlates with the percentage of naive T cells**

To estimate the correlation between relative abundance of naive T cells, donor age, and observed diversity of the TCR β repertoire, we measured percentages of naive CD45RA$^{\text{high/CD27high}}$T cells (42) in blood samples from the same 39 donors using flow cytometry (Fig. 4A).

As expected, the percentage of naive T cells generally decreased with age. Performing linear regression over all four age groups yielded a correlation coefficient of $R = -0.68$. This, however, was greatly improved (to $R = -0.85$) by dropping group 4, which was characterized by a markedly higher percentage of naive T cells compared with group 3. Group 3 was also characterized by a significantly lower percentage of naive T cells compared with group 2 ($p = 0.0012$), whereas the difference between group 2 and group 4 was insignificant ($p = 0.13$) (Fig. 4B).

The percentage of naive T cells showed strong correlation with observed TCR β diversity ($R = 0.90$, $p = 1.4 \times 10^{-14}$; Fig. 5), confirming both the high accuracy of unique molecular identifier-mediated profiling of TCR repertoires and the validity of using CD45RA$^{\text{high}}$/CD27$^{\text{high}}$ gating to determine percentages of naive T cells. As mentioned above, unique identifier-based data were more reliable than nonnormalized, read-based data. The latter demonstrated poorer agreement with experimental measurements of naive T cell counts (Supplemental Fig. 2B).

**FIGURE 3.** Clonal space homeostasis. The percent of clonal space occupied by clones of a given type (classified by size) is provided for the four age groups. The correlation of the percentage occupied by clones of a given type with age is indicated. ***$p < 0.0001$, **$p = 0.001$.**

<table>
<thead>
<tr>
<th>Clone Type</th>
<th>Clone Size</th>
<th>Correlation (R) with age (up to 70)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rare</td>
<td>Single-cell events</td>
<td>-0.8, ***</td>
</tr>
<tr>
<td>Small</td>
<td>More than single event, but less than 0.001%</td>
<td>-0.2, n.s</td>
</tr>
<tr>
<td>Medium</td>
<td>From 0.001% to 0.01%</td>
<td>0.6, ***</td>
</tr>
<tr>
<td>Large</td>
<td>From 0.01% to 1%</td>
<td>0.6, ***</td>
</tr>
<tr>
<td>Hyperexpanded</td>
<td>More than 1%</td>
<td>0.5, **</td>
</tr>
</tbody>
</table>

Notably, the observed TCR β diversity per $1 \times 10^6$ T cells reflects age-dependent changes better than the percentage of naive T cells. When we computed the partial correlation (43) between age and observed sample diversity while controlling for naive T cell percentage, we obtained a significant $R = -0.47$ ($p = 0.003$). In contrast, we obtained a nonsignificant $R = -0.03$ ($p = 0.86$) for the partial correlation between age and naive T cell percentage while controlling for observed sample diversity.

**Long-lived donors have high percentages of naive T cells in their CD4$^+$ subset**

Separate flow cytometry analysis of CD4$^+$ and CD8$^+$ subpopulations to determine percentages of naive CD45RA$^{\text{high}}$/CD27$^{\text{high}}$ T cells (gating identical to Fig. 4A) revealed that the percentage of naive CD8$^+$ T cells decreased linearly ($R = -0.88$, $p < 1 \times 10^{-8}$) with age. The percentages of naive CD4$^+$ T cells also decreased linearly ($R = -0.75$) and with similar kinetics up to the age of 70 yr. However, the long-lived donors of group 4 exhibited a 1.7-fold increased percentage of naive T cells in the CD4$^+$ subset compared with group 3, a difference that was close to significant ($p = 0.054$). Moreover, in contrast to other age groups, we observed a significantly higher percentage of naive T cells in the CD4$^+$ subset compared with the CD8$^+$ subset in Group 4 (3.6-fold, $p = 0.004$) (Fig. 4C). Higher percentages of naive CD4$^+$ T cells correlate with the prevalence of low-frequency clonotypes (Fig. 3) and relatively high TCR β diversity (Fig. 2A, Table I) in group 4.

**Age-dependent changes in relative TRBV and TRBJ usage frequency**

We observed only subtle age-related trends for usage of most V and J gene segments. Still, there were several gene segments for which the frequency among clonotypes showed clear correlation with age (Supplemental Fig. 3). This tendency could be partially explained by the changes in the CD4/CD8 T cell ratio. In agreement with previous studies (44), we observed that the CD4/CD8 ratio gradually increases with age (data not shown). Because TRBV7-9 and TRBV13 were recently shown to be more characteristic of CD8$^+$ T cells (45), a decrease in relative abundance of CD8$^+$ T cells should result in lower usage of these TCR V β gene segments, and this is in accordance with our observations. Conversely, TRBV18 is more characteristic of CD4$^+$ T cells (45), and we likewise observed increasing TRBV18 usage frequency with age (Supplemental Fig. 3).
We did not detect differences between age groups in any other bulk parameters, such as average CDR3 length or number of added nucleotides.

**Public clonotypes and aging**

In recent years, it has become clear that public TCR sequences that are identical or near-identical across different individuals are prominently represented in the blood, as a result of convergent recombination and recombinatorial biases (46–51). In agreement with these observations, cross-analysis of normalized TCR datasets for any two donors from our study revealed essential overlaps that constituted between 8,000 and 55,000 CDR3 variants with identical amino acid sequences and between 500 and 7,000 CDR3 variants with identical nucleotide sequences.

To obtain a list of the most common public TCR clonotypes (representative public clonotypes, repPC), we cross-analyzed the top 100,000 clonotypes from each donor sample and selected those amino acid CDR3 variants that were detected in at least 6 of 39 donors (15% of donors). This yielded a repPC list of 10,000 clonotypes, which is available as Supplemental Table I. As expected (47, 52), repPC featured extremely short CDR3 sequences with a median length of 39 nt, compared with the median of 45 nt from all analyzed clonotypes from all of our donors, as well as a smaller number of added nucleotides (Fig. 6A, 6B). Although the J segment usage of repPC was indistinguishable from the average, there were several notable changes in V segment usage; TRBV12-4, TRBV5-1, TRBV7-2, and TRBV7-9 segments were overrepresented, whereas TRBV20-1 and TRBV29-1 segments were less represented in repPCs compared with the average usage among the top 100,000 clonotypes from our 39 donors (Fig. 6C). Interestingly, the number of repPC clonotypes observed among the top 100,000 clonotypes decreased with age across groups 1–3 (R = 0.58, p = 1 × 10^{-12}; Fig. 6D) and correlated with the observed diversity of TCR β repertoire (Fig. 6E).
Discussion

To our knowledge, our study provides the first high-quality reference data on TCR β repertoires obtained from donors of different ages (http://mitcr.milaboratory.com/datasets/). By using unique cDNA molecular identifiers for TCR profiling (see Materials and Methods and Supplemental Fig. 1), we were able to normalize our samples (Fig. 1) and thereby measure and analyze receptor diversity for fixed numbers of T cells directly and accurately (Figs. 2, 3, Supplemental Fig. 2A, 2B), providing an efficient methodology for comparative studies of adaptive immunity.

In silico CDR3 spectratyping vividly demonstrates that expanded clonotypes can skew the spectratype to various extents, such that they are in some cases essentially hidden (Fig. 7). As such, extrapolating spectratype data to whole repertoires, although a powerful and widely used approach (27, 53, 54), can lead to inaccurate estimations. In contrast, the direct identification of hundreds of thousands or millions of unique TCR β clonotypes using deep normalized profiling followed by strict correction of PCR and sequencing errors provides full information on clonal composition and relative sample diversity. This combination of technological advances takes our ability to analyze and understand adaptive immunity to a new level.

We summarize and interpret our results as follows:

1. In human peripheral blood, the percentage of naive T cells in both CD4+ and CD8+ subsets decreases linearly with age, starting from ~50 to 80% in childhood, with kinetics approximately equal.
to 0.75% of total homeostatic space lost per year. The percentage of naive T cells out of all T cells is decreased 4-fold by the age of 70 (Fig. 4B, 4C).

2. Observed TCR β diversity per 10^6 randomly-captured human peripheral blood T cells correlates with the percentage of naive T cells (Fig. 5) and decreases roughly linearly with age, starting from \(\sim 6 \times 10^4\) clonotypes detected per 10^6 T cells in childhood, with kinetics of \(\sim 5 \times 10^3\) fewer observed clonotypes per year. The decrease in observed TCR β diversity is significant by the age of 40 (Fig. 2A), and generally reflects age-dependent changes even better than the percentage of naive T cells. Given that the blood count of T cells remains generally stable with age (39–41), the observed diversity of TCR β variants per 10^6 T cells can be considered as a measure of “efficient T cell diversity” of an individual, i.e., the probability of a successful meeting between a (naive) T cell and its target Ag.

3. It is very challenging to estimate or measure total TCR β diversity correctly due to the high heterogeneity in the frequency of naive T cells. During homeostatic proliferation, some naive T cells divide more efficiently, and some are gradually lost (5, 7, 8). Thus, at a given moment, a naive T cell clone can be ultimately represented by a single cell. Formally, such single-cell clones still add to the total TCR diversity, although this is probably physiologically nonsignificant. Here we have achieved a record depth of individual TCR β sequencing, but still it only represents up to two million T cells from a pool of almost 10^{12} T cells. Thus it is mathematically impossible to extrapolate sequencing data to accurately measure true total TCR β diversity, and here we have only discussed the lower-bound estimate at the achieved depth of profiling, obtained using the method by Efron et al. (33). At a depth of 1 million analyzed T cells (unique cDNA), this estimate constitutes at least 7 \times 10^5 TCR β variants in childhood, decreasing linearly to an average of \(\sim 2.5 \times 10^5\) variants by old age (Fig. 2C). However, this essentially remains an underestimation, as the analysis of an additional one million T cell replicates from two donors of different age increased the lower-bound estimate for total TCR β diversity \(\sim 2.3\) - to 2.4-fold, reaching at least 9 million variants in the 25-yr-old donor. This estimate will probably increase further with greater sequencing depth. The degree to which the lower-bound estimate increased was consistent between the two donors of different ages, suggesting that such estimate could be potentially employed for a rough comparative analysis of relative T cell diversity in individuals. However, only extradeep sequencing of sorted naive T cells will help to accurately answer the question of total naive TCR β diversity and its changes over the course of the aging process, since bulk approaches imply that different numbers of naive T cells are being analyzed from donors of different ages, which largely determines the observed diversity.

4. Expanded clones occupy increasing amounts of homeostatic space with age (Fig. 3), as described previously (10). This is detrimental to the adaptive immune response, as less space remains for naive T cells, given that the total blood count of T cells is generally stable. Overall diversity is mostly determined by the number of rare clonotypes, and we did not observe significant age-related changes in the diversity of non-rare clonotypes (those with more than one T cell per million, \(R = 0.22, p = 0.2\)). Cumulatively, these data may indicate that naive T cell clones are being lost with age, and that existing Ag-experienced T cell clones become expanded to a greater extent, whereas relatively minor numbers of new Ag-experienced clones emerge. However, in the absence of analysis of purified, phenotypically defined T cell subsets, it is difficult to conclusively compare homeostatic changes within and between compartments. It should be also noted that the CMV infection status of the individuals in our study is not known, which could be important due to the known influence of CMV upon absolute numbers of highly-differentiated T cells (55, 56). Moreover, because Abs to CMV are detected in a majority of donors over the age of 40 (57), we expect that we would not be able to perform accurate comparative analysis even with this data, as only a small number of elderly donors would be CMV-negative. Further studies using deep profiling of TCR repertoires with sorted naive and Ag-experienced populations of T cells from sufficient cohorts of donors of different ages with known CMV status should reveal the exact changes in the Ag-experienced and naive subsets associated with
aging. We believe that normalization approach described in this study, based on the use of unique molecular identifiers, should greatly facilitate such studies.

5. Increased percentages of naive T cells in the CD4+ T cell subset, higher prevalence of low frequency clonotypes, and slightly higher TCR β diversity are observed in donors of group 4 (average age 82 y) compared with group 3 (Figs. 2–4, Table I). A plausible explanation is that the “long-lived” individuals in group 4 have passed a certain threshold of age selection, their physiological parameters might differ noticeably from the general population. This implies that we are observing the effects of selection bias and a possible association of increased percentages of naive T cells in the CD4+ subset with longevity. Such an association can be explained by the improved immunoregulatory function of a more diverse CD4+ T cell subset in these individuals, which restrains the generalized inflammation that typically increases with age (58), and provides a better-balanced immune response. Our explanation differs from that of Ferrando-Martinez and colleagues (59), who suggested that the dysfunctional aged thymus could be generally biased toward naive CD4 T cell production. It is possible that the increased percentages of naive T cells in the CD4+ subset with age (58), and provides a better-balanced immune response. Miller, R. A. 1996. The aging immune system: primer and prospectus. Science 273: 70–74.

6. Interindividual TCR sharing in humans is a significant phenomenon and is thought to play an important role in the efficacy of Ag responses (49–51, 60). In this study, we have extracted a list of the 10,692 most representative public TCR β clonotypes that are widely shared (present among the top 100,000 expanded clonotypes) among at least 6 of our 39 donors (i.e., ±15%). These repPC clonotypes display CDR3 properties suggesting closeness to germline: short CDR3 length (median length of 39 nt, compared with 45 nt for all clonotypes) and low numbers of added nucleotides (Fig. 6A, 6B, Supplemental Table I). Notably, the number of repPC clonotypes decreased with age, displaying a pattern similar to that observed for percentage of naive T cells and sample diversity (Fig. 6E, 6F). One possible explanation is that a sizeable portion of high-clonotypes decreased with age, displaying a pattern similar to that with 45 nt for all clonotypes) and low numbers of added nucleotides.

Disclosures
The authors have no financial conflicts of interest.

References