A Novel Approach to Tracking Antigen-Experienced CD4 T Cells into Functional Compartments via Tandem Deep and Shallow TCR Clonotyping

Megan Estorninho, Vivienne B. Gibson, Deborah Kronenberg-Versteeg, Yuk-Fun Liu, Chester Ni, Karen Cerosaletti and Mark Peakman

*J Immunol* 2013; 191:5430-5440; Prepublished online 25 October 2013; doi: 10.4049/jimmunol.1300622
http://www.jimmunol.org/content/191/11/5430

**Supplementary Material**

http://www.jimmunol.org/content/suppl/2013/10/25/jimmunol.1300622.DC1

**References**

This article cites 40 articles, 11 of which you can access for free at: http://www.jimmunol.org/content/191/11/5430.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
A Novel Approach to Tracking Antigen-Experienced CD4 T Cells into Functional Compartments via Tandem Deep and Shallow TCR Clonotyping

Megan Estorninho,* Vivienne B. Gibson,* Deborah Kronenberg-Versteeg,* Yuk-Fun Liu,* Chester Ni,† Karen Cerosaletti,† and Mark Peakman*

Extensive diversity in the human repertoire of TCRs for Ag is both a cornerstone of effective adaptive immunity that enables host protection against a multiplicity of pathogens and a weakness that gives rise to potential pathological self-reactivity. The complexity arising from diversity makes detection and tracking of single Ag-specific CD4 T cells (ASTs) involved in these immune responses challenging. We report a tandem, multistep process to quantify rare TCRβ-chain variable sequences of ASTs in large polyclonal populations. The approach combines deep high-throughput sequencing (HTS) within functional CD4 T cell compartments, such as naïve/memory cells, with shallow, multiple identifier–based HTS of ASTs identified by activation marker upregulation after short-term Ag stimulation in vitro. We find that clonotypes recognizing HLA class II–restricted epitopes of both pathogen-derived Ags and self-Ags are oligoclonal and typically private. Clonotype tracking within an individual reveals private AST clonotypes resident in the memory population, as would be expected, representing clonal expansions (identical nucleotide sequence; “ultraprivate”). Other AST clonotypes share CDR3β amino acid sequences through convergent recombination and are found in memory populations of multiple individuals. Tandem HTS-based clonotyping will facilitate studying AST dynamics, epitope spreading, and repertoire changes that arise postvaccination and following Ag-specific immunotherapies for cancer and autoimmune disease. The Journal of Immunology, 2013, 191: 5430–5440.

A daptive immune responses mediated by T lymphocytes recognizing peptide epitope–HLA glycoprotein complexes via their surface TCR for Ag underlie major protective functions of the immune system, including antimicrobial and vaccine-induced immunity (1). The same molecular interactions are responsible for the recognition of self that results in autoimmune inflammatory diseases, which affect as many as 4% of the population of developed countries (2). Understandably, considerable research effort over recent years has focused on gaining insight into the nature of the trimolecular interaction between the peptide–HLA (pHLA) complex and the TCR on the surface of Ag-specific T cells (ASTs) (3, 4). Of the component parts, the identification and tracking of epitope-specific T cells presents a considerable challenge. The advent of pHLA multimer technology addresses this technology gap to a large extent; however, in human immunology it has been much more successful for the investigation of CD8, rather than CD4, T cell responses (5). This reflects the easier assembly and stability of class I molecules in vitro; the key role of CD8, as opposed to CD4, in pHLA–TCR stabilization; and the observation that, in many contexts, the frequency of CD4 T cell effectors is lower than that for CD8 T cells and, therefore, is at the limit of multimer detection (5). Moreover, some immune responses, such as tumor surveillance and autoimmunity, are characterized by low-avidity interactions that further limit the use of multimer technology (6). Alternative approaches to TCR analysis typically focused upon the generation of Ag-specific CD4 T cell clones, which introduces a functional bias, or used technologies, such as spectratyping or flow cytometry, which are of low resolution and do not provide sequence signatures. This leaves a technology gap in our ability to track epitope- and Ag-specific CD4 T cells without bias for immune-monitoring programs that are critical to the success of diverse fields, such as tumor immunology, vaccine development, autoimmunity, and Ag-specific immunotherapeutics.

The diverse epitope specificity of αβ T cells that equips the adaptive immune system to respond to the vast array of extrinsic and intrinsic Ags to which it is exposed is attributable to possession of highly polymorphic TCRs (7). Potentially, this key attribute provides the solution to CD4 T cell tracking, because the fine specificity of TCR recognition of peptide epitope arises from signature diversity in
the amino acid sequence of the CDR3 loops of the α- and β-chain variable domains. The CDR3 regions are formed by recombination among noncontiguous variable (Vβ), diversity (Dβ), and joining (Jβ) gene segments in the β-chain locus, as well as between analogous Vα and Jα gene segments in the α-chain locus (1). The existence of multiple such gene segments in the α- and β-chain loci allows for a multiplicity of distinct CDR3 sequences to be encoded, with further diversity realized by template-independent addition and deletion of nucleotides at the Vβ-Dβ-Jβ-Jβ, and Vα-Jα junctions during gene rearrangement. Recent advances in DNA-sequencing technology have enabled significantly deeper and higher throughput sequencing than was previously possible (8), potentially enabling interrogation of CD4 T cell populations for disease or immune response signatures. However, such approaches are unlikely to have sufficient resolution to detect and track the rare cohorts of ASTs. To address this technological gap, we developed a tandem, multistep process to identify and track epitope-specific CD4 T cells. We applied the technology in two settings that resolve anamnestic pathogen- and autoantigen-specific responses. In both cases, TCR clonotypes are oligoclonal, rarely shared across individuals, and can be tracked via CDR3 amino acid- and nucleotide-coding signatures as low-frequency cohorts present in the memory population in peripheral blood. Therefore, the approach has the potential to offer insights at multiple levels, including tracking of responders, clonotype frequency, and identification of the functional differentiation of epitope-specific T cells.

Materials and Methods

Subjects
A total of six patients with type 1 diabetes were recruited into this study, which was approved by the Institutional Ethics Review Board. Key inclusion criteria were the presence of the HLA class II susceptibility genotype DRBI*0401; the presence of at least one islet cell autoantibody; and the presence of IFN-γ-producing CD4 T cells directed against naturally processed and HLA-DR4–presented immunodominant islet β-cell autoantigenic epitopes (Table I), as detected by sensitive cytokine ELISPOT, which was performed as described previously (9).

Sorting CD154+CD69+CD4+ T cells to obtain Ag-specific CD4 T cell clones

Fresh PBMCs (2 × 10^7) were incubated at 10^6 cells/ml in 48-well plates (37°C, 5% CO₂) in RPMI 1640/10% human AB serum supplemented with 1 µg/ml each anti-CD28 (clone CD28.2) and anti-CD40 (41/CD40, both BD Biosciences) (200 ng/ml each of the naturally processed and HLA-DR4–presented immunodominant peptide of influenza hemagglutinin (HA306-314), all peptides from Thermo Scientific and 95% purity). Nonadherent cells from these cultures were harvested after 16 h, washed, and stained with the following fluorochrome-conjugated MAbs: anti-CD14 and anti-CD19 (both Pacific blue; Invitrogen); anti-CD3 (allophycocyanin-H7), anti-CD4 (allophycocyanin), anti-CD154 (PE), anti-CD69 (FITC; all from BD Biosciences), and ViViD. Total lymphocytes were gated on forward and side scatter and doublets, and dead cells and CD14+ and CD19+ events were excluded. CD3+CD4+CD154+CD69+ T cells, sorted (Supplemental Fig. 1) as described above, were stained and live single-cell sorted using a FACSARia (BD Biosciences) into 96-well plates containing irradiated mixed-donor PBMCs as feeders. T cells were restimulated with matched feeders (10^5 cells/well) in X-vivo (Lonza/10% AB. Testing to examine the enrichment of Ag-specific T cell clones was performed by [3H]thymidine incorporation using peptide-pulsed (10 µg/ml) PBMCs (2.5 × 10^5/well at 1:1 ratio with clone) for Ag presentation and anti-CD3/anti-CD28 MAb stimulation as control; culture supernatants were taken to examine IFN-γ secretion by ELISA.

Sorting CD154+CD69+ Ag-specific CD4 T cells for multiplex TCR Vβ sequencing

PBMCs (2 × 10^7) were incubated, as described above, with either β-cell autoantigenic peptide (10 µg/ml; Table I) or HAg306-314, harvested and stained, as described above, and sorted directly into RLT lysis buffer (QIAGEN). An additional 5 × 10^5 PBMCs were incubated with the polyclonal stimulus staphylococcal enterotoxin B (1 µg/ml) before sorting into RLT buffer.

Generation of polyclonal T cell line and clone-spiking experiments

A polyclonal CD4 T cell line (termed poly CD4-TCL) was generated from purified CD4 T cells (Isolation Kit II; Miltenyi Biotec, Surrey, UK.) by stimulation with 4 µg/ml PHA (Biostat Diagnostic Systems) in the presence of mixed irradiated feeders at a ratio of 5:1 in X-vivo/5% AB/5% Cellkine (ZepoMetrix). Medium was exchanged every 2 d with reducing concentrations of Cellkine through day 12, when cells were harvested. To examine the sensitivity and specificity of shallow TCR Vβ sequencing of Ag-stimulated CD3+CD4+CD154+CD69+ CD4 T cells, we devised a spike-in experiment in which 4C6 clone cells (10) were sorted in triplicate, at a frequency of 0.2, 0.4, 1, 2, or 5%, directly into RLT buffer into which 500 poly-CD4 TCL cells had been sorted using a FACSAria.

TCR Vβ amplification and sequencing of polyclonal Ag-specific CD4 T cells

Total RNA was extracted using the RNeasy Micro Kit (QIAGEN), and template-switch anchored RT-PCR was performed to provide unbiased and representative TCR Vβ amplification (11), as described (12), with the modification of Phusion High-Fidelity enzyme (Finnzymes) for amplification. Briefly, cDNA was prepared using a SMARTer RACE cDNA Amplification Kit (Clontech), according to the manufacturer’s instructions with the following modifications: 3.25 µl total RNA was added to 1.0 µl 5’-CD5 Primer A; reverse transcription was extended to 120 min at 42°C; three replicade cDNA reactions were pooled and resuspended at a 1:1 ratio with Tricine-EDTA buffer; and 5 µl cDNA was amplified using Phusion and gene-specific TRBC primer (5’-CACAGCGACCTCGGGTGTTGGAACAC-3’), including a 5’ CA-linker and followed by one of 24 10-bp unique MID tag sequences (Supplemental Table I), at 1 µM, paired with the forward 5’ RACE primer (1 µM) nested within the first-round Universal Primer Mix (5’-AAGCAGTGGTAGATACCAAGCAGATG-3’). Touchdown PCR cycling conditions were 98°C for 30 s; 5 cycles of 98°C for 10 s, 65°C for 30 s, 72°C for 45 s; 12–30 cycles at 98°C for 10 s, 68°C for 30 s, 72°C for 40 s; a final extension at 72°C for 5 min. cDNA libraries from MID-labeled samples were gel extracted, equimolar pooled, and sequenced on the Roche 454 platform by Eurofins MWG Operon. V-J-D gene usage and CD3R regions were determined using the IMGT/V-QUEST tool (13, 14). Sequencing data obtained (from Ag-specific CD4 T cells, CD8 spike-in experiments, and the poly CD4-TCL) were scrutinized for sequencing error: nonproductive clonotypes (not mapping to the IMGT/GENE-DB database) were excluded, and only clonotypes at a frequency >0.1% (the mean observed lower limit of detection for 4C6 clone cells spiked at 0.2%, see Results) were considered for further analysis. AST samples are prepared using 5’RACE technology and do not require normalization for PCR bias. In combination with the low error rate of Phusion High-Fidelity DNA polymerase [estimated at 0.11% (15)], these quality control measures limit low-abundance TCR clonotypes introduced by sequencing and PCR error. AST clonotype frequencies were then calculated as a function of the total number of productive clonotypes (using three-point match of V and J gene family and CD3R sequence) that matched to the International ImMunoGeneTics collaboration (IMGT/GENE-DB) database and were above the 0.17% quality control cut-off described above.

Tandem deep-TCR clonotype analysis of naive and memory CD4 T cells

From the same blood sample from which AST clonotypes were examined, ~3 × 10^7 PBMCs were stained with anti-CD3 mAb (allophycocyanin–eFluor 780, PerCP–eFluor 710), anti-CD45RA (PE-Cy7), anti-CD45RO (FITC), anti-CD62L (allophycocyanin; all from eBioscience), and ViViD (Molecular Probes). Approximately 10^6 CD3+CD45RA– (naive) and CD3+CD45RA+CD62L–CD4 T cells were sorted by flow cytometry directly into RLT lysis buffer following gating on total lymphocytes using forward and side-scatter, with the exclusion of doublets and dead cells (Supplemental Fig. 2). Total RNA from naive and memory-enriched CD4+ T cell populations was extracted and reverse transcribed using the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems), and the expressed TRβ genes were sequenced on the Illumina GAIIx analyzer at Adaptive Biotechnologies (Seattle, WA) using
HTS, as described (16). Briefly, analysis of TCRB CDR3 regions was performed using PCR amplification with the forward and reverse primer pools (IDT, Coralville, IA), each at a final concentration of 1 μM, in QIAGEN Multiplex PCR Master Mix (cat. no. 206145) with 10% Q-solution (both from QIAGEN). The reaction was cycled on a Bio-Rad thermocycler (Hercules, CA), using the following cycling conditions: 95˚C for 15 min; followed by 35 cycles at 94˚C for 30 s, 59˚C for 30 s, 72˚C for 1 min; followed by 1 cycle at 72˚C for 10 min. The forward and reverse primers included a tail, with Illumina adapters enabling direct sequencing of the PCR libraries on the Illumina HiSeq platform (Illumina, San Diego, CA). The sequences for both of the TCRB CDR3 regions were delineated according to the definition established by the IMGT collaboration (17). Sequences that did not match CDR3 sequences were removed from the analysis. A standard algorithm was used to identify which V, D, and J segments contributed to each TCRB CDR3 sequence (18). Sequencing data were accessed using the Adaptive Biotechnologies immunoSEQ viewer. All immunoSEQ data are subject to quality control processing with the aim to eliminate sequence reads due to intrinsic sequencing error and PCR (18), excludes all single unique reads and nonproductive clonotypes (reading out-of-frame or with stop codons), and uses an algorithm to normalize for PCR bias. Normalized files, detailing clonotype frequencies; V, J, and D gene usage; and CDR3 sequences were uploaded to the Adaptive Biotechnologies immunoSEQ algorithm to normalize for PCR bias. Normalized files, detailing clonotype frequencies, distributions and distributions were ranked and normalized, z-scores were determined, and Ag-specific and polyclonal CD4+ T cell frequencies and distributions were compared using the unpaired t test or Wilcoxon rank-sum test, as appropriate, according to the normality of the data distribution. Statistical analysis was performed using GraphPad Prism 5 software, and p values < 0.05 were considered statistically significant.

Results

Sorting CD154+CD69+CD4+ T cells post-Ag stimulation in vitro enriches Ag-specific CD4+ T cells

We first validated previous reports that short-term culture in vitro with Ag, coupled with CD40 blockade and sorting on CD154/CD69 coexpression, enables efficient enrichment and sorting of CD4+ ASTs (23, 24) (Fig. 1). For example, using a sensitive IFN-γ ELISPOT assay, in an index donor ~0.0012% PBMCs secrete IFN-γ in response to short-term (16 h) stimulation with the immunodominant peptide of influenza hemagglutinin, HA306–318 (Fig. 2A). From similar cultures, CD154+CD69+ CD4+ T cells were single-cell sorted and expanded nonspecifically before verification of specificity by retesting against HA306–318 (Fig. 2B, 2C). We found that 24.5% of clones (of 49 tested) were specific for HA306–318 when using IFN-γ production as the read-out, thus indicating ≥20,000-fold enrichment (yet likely to be an underestimate of the true number of Ag-specific cells because CD154+CD69+ cells enriched by this approach are known to be polyclonal).

Multiplex TCR Vβ sequencing for ASTs

We next developed a shallow HTS-based Ag-specific clonotyping technology, similar to that described by Mamedov et al. (25), and assessed its performance characteristics by spiking a T cell clone (designated 4C6; clonotyped as TRBV7-9+3 with CDR3β sequence CASSLHHEQYF (10)) at frequencies between 1/500 and 1/20 into 500 cells of a polyclonally expanded short-term CD4+ cell population. In this context, the number of samples and, therefore, requires normalization. As an example of such a context, when comparing the diversity of species in two habitats it is important to normalize for the number of animals studied in each. Therefore, entropy measures were normalized for comparing multiple datasets with different numbers of cells sorted using the following equation: Normalized entropy (nH) = entropy/log2 (number of cells sorted in that sample). Dividing by this number (which is the maximum value of H for that sample) provides a number in the range of 0 to 1, giving information on how oligoclonal or evenly distributed the sample is. The unpaired t test was used to compare mean nH values between clonotyped samples. Clonotype frequencies and distributions were ranked and normalized, z-scores were determined, and Ag-specific and polyclonal CD4+ T cell frequencies and distributions were compared using the unpaired t test or Wilcoxon rank-sum test, as appropriate, according to the normality of the data distribution. Statistical analysis was performed using GraphPad Prism 5 software, and p values < 0.05 were considered statistically significant.

**FIGURE 1.** Flow diagram of experimental strategy. The strategy is a two-step approach combining direct ex vivo, multiplex, shallow, high throughput (HT) TCR clonotyping of Ag-specific CD4+ T cells (top and right arm of flow diagram) with deep HT TCR clonotyping in functional subsets (memory and naive pools; left arm of flow diagram). Fresh PBMCs are partitioned for immediate flow sorting of 10^6 naive and memory CD4+ T cells (A) or simultaneous short-term (16 h) direct ex vivo stimulation with selected Ag (B) (in this case either β-cell Ag IA-2ß,375–377 or influenza HA306–318 in the presence of blocking anti-CD40 mAb to maintain CD40L [CD154] expression on activated CD4+ T cells). Staining examples include staphylococcal enterotoxin B as a positive control and IA-2ß,375–377. Activated (CD154+CD69+) Ag-specific CD4+ T cells are flow sorted (C), subjected to 5’RACE and MID labeled (D), and pooled for multiplex HT TRB sequencing (Roche 454) (E). Following RNA extraction and cDNA synthesis, TRB chains from naive and memory CD4+ T cells are HT deep sequenced (Illumina platform) (F) and analyzed using a combined bioinformatic platform (G).
mAbs as control; black bars, Ag (HA306–318 peptide, 10 μg/ml); dark gray bars, clone and feeder cells alone; light gray bars, clone alone. The restricted oligo-clonotypy observed in response to foreign genetic peptide stimulation and the number derived after polyclonal stimulation. The number of Ag-driven clonotypes was not distributed normally: the median following HA306–318 stimulation was 29 (range, 8–149; n = 6 subjects), whereas the median was 10 (range, 6–94; n = 5; data not shown) following culture with IA-2752–775. The number of unique clonotypes was similar for the two antigenic stimuli but, in both cases, was significantly lower than that observed for poly CD4-TCL (p < 0.002 for both, Fig. 3D). Overall, these data indicate that there is marked oligoclonality in the Ag-specific responder populations, providing further confirmation that CD154+CD69+ cells are Ag-driven, epitope-specific, discrete expanded cohorts of effector CD4 T cells. As a further indication of limited TRB clonality, we calculated entropy (H) as a measure of diversity [uniformity of raw frequency distribution (22)], in this case normalized (nH) to account for differences in cell number in each sample. Clonotype diversity was highly similar in the HA306–318-specific and IA-2752–775-specific responder populations but differed significantly for both in comparison with the poly CD4-TCL (p = 0.002 and 0.0002, respectively, Fig. 3E). These data are statistical indications that Ag-specific populations have low normalized entropy (nH) scores typical of oligoclonal repertoires, whereas the poly CD4-TCL has a score close to unity, consistent with the expected highly diverse TRB repertoire of a polyclonally expanded cohort of cells.

**Examining TRBV gene family, allele, and CDR3β sequence**

The restricted oligo-clonotypy observed in response to foreign and self-antigenic epitopes implies Ag-driven selection. Further restriction of the response could result from selected TRBV gene family and allele usage, without residue conservation across CDR3, which is known as type I bias (27). Examining the frequency of TRBV genes and alleles in Ag-specific responses, we find little...
evidence for type I bias in this small dataset for either stimulus (Fig. 4). Type II bias (predictability in TRBV or TRAV gene usage and clear residue motifs within the CDR3 loop) also was not seen and neither was type III bias (complete TCR \( \alpha \) and/or TCR \( \beta \) sequence predictability, also known as public clonotypy). There is some evidence for TRBV20-1, TRBV5-1, and TRBV6-5 skewing within the HA clonotypes (Fig. 5); however, these TRBV gene families also predominate in the naive repertoire (see below) and may arise more commonly as a result of linkage disequilibrium and recombination of these genes at higher frequency because of their chromosomal location (28).

Cocrystallization of TCR–pHLA complexes indicated that the CDR3b region has a large footprint of interaction with peptide epitope bound in the HLA-binding groove of all of the CDRs.

### Table I. Clinical and mechanistic data from study subjects

<table>
<thead>
<tr>
<th>Patient ID</th>
<th>Age (y)</th>
<th>Sex</th>
<th>Islet Cell Autoantibody Status</th>
<th>ELISPOT Response to β-Cell Peptide&lt;sup&gt;a&lt;/sup&gt; (Stimulation Index&lt;sup&gt;b&lt;/sup&gt;)</th>
<th>Number of CD154⁺CD69⁺ CD4 T Cells Sorted&lt;sup&gt;c&lt;/sup&gt;</th>
<th>β-Cell Specific&lt;sup&gt;d&lt;/sup&gt;</th>
<th>HA Specific&lt;sup&gt;e&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1D1</td>
<td>25</td>
<td>F</td>
<td>GAD⁺</td>
<td>4.4</td>
<td>413</td>
<td>413</td>
<td>904</td>
</tr>
<tr>
<td>T1D2</td>
<td>37</td>
<td>F</td>
<td>GAD⁺</td>
<td>22.0</td>
<td>499</td>
<td>499</td>
<td>727</td>
</tr>
<tr>
<td>T1D3</td>
<td>27</td>
<td>F</td>
<td>GAD⁺, IA-2⁺</td>
<td>4.5</td>
<td>382</td>
<td>382</td>
<td>373</td>
</tr>
<tr>
<td>T1D4</td>
<td>25</td>
<td>M</td>
<td>IA-2⁺</td>
<td>2.0</td>
<td>248</td>
<td>248</td>
<td>264</td>
</tr>
<tr>
<td>T1D5</td>
<td>24</td>
<td>F</td>
<td>GAD⁺, IA-2⁺</td>
<td>4.7</td>
<td>1610</td>
<td>1610</td>
<td>1476</td>
</tr>
<tr>
<td>T1D6</td>
<td>44</td>
<td>M</td>
<td>GAD⁺</td>
<td>3.0</td>
<td>683</td>
<td>683</td>
<td>714</td>
</tr>
</tbody>
</table>

<sup>a</sup>Stimulation Index is the ratio between spot number in the presence/absence of test peptide.

<sup>b</sup>All subjects were tested against IA-2<sub>282-285</sub>, with the exception of T1D1, who responded to PI C19-A3.

<sup>c</sup>All subjects were studied within 3 mo of diagnosis, with the exception of T1D6, who had type 1 diabetes for 33 y.

<sup>d</sup>Female; M, male.
Therefore, we examined evidence of amino acid bias in CDR3β sequences for each antigenic stimulus, focusing on the four most commonly observed sequence lengths (13–16 residues) (data not shown). We observe a trend for positive-to-neutral-to-negative charge shift across the CDR3β sequence viewed from N to C terminus [i.e., upstream and downstream of diversity, N(D)N gene region, respectively]. Using motif (MEME) analysis, we compared similar-length CDR3β sequences for all clonotypes from HA responder T cells (n = 247 sequences) and IA-2 responder T cells (n = 119 sequences; data not shown). No distinct patterns of amino acid usage were seen, with the exception of a trend toward the presence of C-terminal motifs PQHF (encoded by TRBJ1-5*01) and ETQYF (encoded by TRBJ2-5*01); these TRBJ genes were used by >32% of IA-2 responder T cells compared with a frequency of 16% in naive cells.

To summarize, the simultaneous analysis of multiple epitope-specific clonotypes in multiple subjects using our shallow HTS clonotyping approach is able to inform on key characteristics, including clonal frequency, clonality, diversity, gene and family use, and motif expression. In the context of the anamnestic, HLA class
II–restricted CD4 T cell responses that we studied, and with the caveat that the number of subjects is small, we observe a strong trend toward the use of private TCR clonotypes.

**Tandem deep-TCR clonotype analysis of naive and memory CD4 T cells**

In tandem with the AST TCR clonotype analysis, we sorted naive and memory peripheral CD4+ T cells from the same blood sample and performed deep TRBV-D-J sequencing, as described (16, 29) (Table II). The total number of sequence reads achieved for each sample is between 4.47 \( \times 10^5 \) and 3.04 \( \times 10^7 \); the number of resulting productive sequences is 4.33 \( \times 10^5 \)–3.06 \( \times 10^7 \), and the percentage of productive sequences lies between 92.19 and 97.28%. Unproductive sequences are attributable to an intrinsic level of sequencing error introducing frame-shift mutations and/or stop codons. A small proportion of T cells may also produce out-of-frame TRBV transcripts that escape nonsense-mediated decay and accumulate in the nucleus (30). The number of unique sequences for each sample is between 3.14 \( \times 10^4 \) and 6.86 \( \times 10^6 \) and is higher in the naive compartment for each patient, with the exception of T1D3. The diversity of each compartment is calculated by normalizing for the differences in the number of cells sorted between samples and is represented as the normalized entropy (nH) for all samples (Fig. 6A). As reported by other investigators (16, 31), diversity in naive and memory cell populations is not as vastly dissimilar as previously thought, despite the increased number of unique sequences in naive repertoires. Nonetheless, the difference in mean diversity between naive and memory populations was significant (\( p = 0.03 \)).

**Detection of Ag-specific clonotypes in naive and memory populations**

Ag-specific clonotypes were then tracked into the corresponding naive and memory CD4+ T cell populations of the same subject to estimate frequency in these compartments. Searches were conducted based on CDR3β amino acid sequence, V and J gene identity, and by alignment of nucleotide sequences from the shallow (Roche 454) sequencing of Ag-specific TCRs to the deep HTS (Illumina) se-

---

**Table II. Summary of deep sequencing data for naive and memory repertoires**

<table>
<thead>
<tr>
<th>Patient ID</th>
<th>Population</th>
<th>Sorted Cell No.</th>
<th>Total Sequences</th>
<th>Productive Sequences</th>
<th>% of Total Sequences</th>
<th>Unique Sequences</th>
<th>% of Total Unique Sequences</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1D 1</td>
<td>Naive RO&quot;RA&quot;</td>
<td>1.19E+06</td>
<td>2.81E+07</td>
<td>2.67E+07</td>
<td>95.17</td>
<td>6.86E+05</td>
<td>93.67</td>
</tr>
<tr>
<td>T1D 2</td>
<td>Memory RO&quot;RA&quot; CD62L(^{ab})</td>
<td>4.59E+05</td>
<td>3.2E+07</td>
<td>3.06E+07</td>
<td>95.64</td>
<td>1.99E+05</td>
<td>90.89</td>
</tr>
<tr>
<td>T1D 3</td>
<td>Naive RO&quot;RA&quot;</td>
<td>1.00E+06</td>
<td>1.39E+07</td>
<td>1.34E+07</td>
<td>96.12</td>
<td>4.78E+05</td>
<td>95.08</td>
</tr>
<tr>
<td>T1D 4</td>
<td>Memory RO&quot;RA&quot; CD62L(^{ab})</td>
<td>1.00E+06</td>
<td>1.83E+07</td>
<td>1.78E+07</td>
<td>97.28</td>
<td>3.12E+05</td>
<td>94.24</td>
</tr>
<tr>
<td>T1D 5</td>
<td>Naive RO&quot;RA&quot;</td>
<td>1.00E+06</td>
<td>1.78E+07</td>
<td>1.67E+07</td>
<td>93.97</td>
<td>2.15E+05</td>
<td>94.28</td>
</tr>
<tr>
<td>T1D 6</td>
<td>Memory RO&quot;RA&quot; CD62L(^{ab})</td>
<td>8.00E+05</td>
<td>3.04E+07</td>
<td>2.91E+07</td>
<td>95.71</td>
<td>4.85E+05</td>
<td>94.37</td>
</tr>
<tr>
<td>T1D 7</td>
<td>Naive RO&quot;RA&quot;</td>
<td>1.00E+06</td>
<td>6.9E+06</td>
<td>6.36E+06</td>
<td>92.19</td>
<td>6.02E+05</td>
<td>92.63</td>
</tr>
<tr>
<td>T1D 8</td>
<td>Memory RO&quot;RA&quot; CD62L(^{ab})</td>
<td>4.00E+05</td>
<td>7.92E+06</td>
<td>7.34E+06</td>
<td>92.67</td>
<td>3.15E+04</td>
<td>89.93</td>
</tr>
<tr>
<td>T1D 9</td>
<td>Naive RO&quot;RA&quot;</td>
<td>1.00E+06</td>
<td>1.31E+07</td>
<td>1.25E+07</td>
<td>95.56</td>
<td>4.38E+05</td>
<td>94.28</td>
</tr>
<tr>
<td>T1D 10</td>
<td>Memory RO&quot;RA&quot; CD62L(^{ab})</td>
<td>1.00E+06</td>
<td>2.49E+07</td>
<td>2.4E+07</td>
<td>96.68</td>
<td>2.93E+05</td>
<td>92.77</td>
</tr>
</tbody>
</table>

All subjects were tested against IA-2\(^{2752-775}\), with the exception of T1D1, who responded to PI C19-A3.
quencing of naive and memory populations. In the first approach, HA\textsubscript{306-318}-specific and IA-2\textsubscript{752-775}-specific clonotypes were found in CD4 subsets of the majority of subjects and, importantly, when present were more prominent in the memory repertoire (Fig. 6B, 6C), confirming Ag experience. Comparisons at the nucleotide level indicated that these clonotypes represent a mixed pattern of “private” and “public.” The majority of different AST clonotypes in an individual subject are private by virtue of having unique nucleotide sequences that are not found in other patients. In the remainder, the different nucleic acid sequences translate into identical CDR3\textsubscript{B} amino acid sequences (convergent recombination), which may be found in multiple subjects (public; Supplemental Table II). The generation of public CDR3\textsubscript{B} sequences occurring in the naive and memory repertoires of multiple subjects was particularly evident for the HA\textsubscript{306-318}-specific clonotype isolated from T1D2 (present in 6/6 subjects; data not shown) and for the IA-2\textsubscript{752-775}-specific clonotype isolated from T1D6 (present in 5/6 subjects; data not shown). Thus, the potential for public CDR3\textsubscript{B} sequences to be recruited exists, and this further emphasizes the intensely private nature of the Ag-specific response that we observed. Where there were fewer or no Ag-specific clonotypes detected in corresponding naive and memory compartments (T1D5 and T1D4, respectively), this was associated with a substantially lower sequencing depth for these samples (data not shown). This may be important for future studies, indicating a sequencing depth threshold below which Ag-specific clonotypes are difficult to track.

We next combined datasets to examine all Ag-specific clonotypes in naive and memory CD4 T cell populations across the whole patient group, searching for identical V and I gene usage and CDR3\textsubscript{B} protein sequences (Fig. 7). This analysis reiterates several of our findings in individuals, notably the greater diversity of naive over memory clonotypes (2,121,279 versus 1,034,500) and the fact that the number of AST clonotypes found in the memory pool is much higher than is seen in the naive pool (36 versus 13 for HA\textsubscript{306-318} and 27 versus 3 for IA-2\textsubscript{752-775}).

Finally, we examined the frequencies of Ag-specific clonotypes within the memory repertoire. These were observed at low and similar frequencies for each Ag: HA\textsubscript{306-318} clonotypes were observed at a median frequency of $6.6 \times 10^{-6}$ and $2.4 \times 10^{-6}$, whereas IA-2\textsubscript{752-775} clonotypes were present at a median frequency of $8 \times 10^{-6}$ and $3.9 \times 10^{-6}$ in T1D1 and T1D6, respectively (Fig. 8). As a comparator, we also analyzed the frequency of the 10 most and least frequent clonotypes present in the memory repertoire of each subject. The most abundant memory clonotypes are typically present at a frequency $\sim 10^{-3}$, and the least abundant have ultralow frequencies $< 10^{-7}$.

**Discussion**

We identified a technology gap in relation to high throughput detection of the signatures of Ag-specific CD4 T lymphocytes involved in adaptive immune responses and their simultaneous tracking into large, complex cell populations. To address this, we developed a tandem, multistep process in which the first stage identifies rare, Ag-responsive CD4 T cells via sequencing of their signature TCR \textbeta-chain V region. The use of MIDs and multiplex shallow HTS in this component lends it to concerted analysis of multiple response specificities and/or large subject cohorts. This
stage-one analysis can be used to decipher clonality and diversity, as well as the extent to which responses use private or public repertoires. Where public repertoires are not apparent from sequence identity, sequence similarities may suggest a structural basis for publicity [so-called “type IV bias” (27)]. At the second stage, deep HTS of TCR β-chain variable regions within functional CD4 T cell compartments, such as naïve/memory cells, is performed. Analysis of the tandem datasets enables tracking of the signatures of Ag-specific responses into cell populations of interest. The combined analysis can be used to study the potential for responders to be present in T lymphocyte subsets, track their frequency, and provide further insight into repertoire publicity and the extent to which CDR3 β-chain regions make use of convergent recombination.

A tandem sequencing approach was adopted to exploit the advantages of each Next Generation Sequencing technology. The Roche 454 is best suited to sequencing small AST populations because these do not require the read depth that is achieved on the Illumina platform, but they benefit from the longer reads. In addition, the Roche 454 has immense multiplexing capability through the use of MIDs and gaskets to divide the flow cell. This enables up to 24 MID-labeled TCR libraries to be sequenced on a fraction (1/16) of the flow cell without sequencing to unnecessary depth or at excessive cost. In contrast, Illumina technology provides the depth of sequencing required for TCR clonotyping in complex samples to detect rare Ag-specific clonotypes.

Therefore, we used this tandem approach in the context of a longstanding interest in the role of autoreactive CD4 T cells in the autoimmune disease type 1 diabetes (9, 10, 32–34). Previous work suggested that autoreactive T cells are rare in comparison with antipathogen responses (9), but this assumption relied upon identification using effector responses, such as single-cytokine production, as the read-out. In contrast, the upregulation of CD154 and CD69 following short stimulation in vitro with Ag is known to be a feature of polyfunctional T cells (23, 24) and, therefore, provides a less biased estimation of responder frequency. Using this approach, we found very few differences between influenza HA- and β-cell epitope-specific T cell frequencies. Similar numbers were sorted, clonotype frequency and diversity were similar, and estimates of the frequency of TCRs sharing CDR3 β-chains with these Ag-specific clones in corresponding naïve and memory populations across subjects were also similar, with a median frequency among CD4+ memory cells ∼1/100,000/clonotype. Because there were similar numbers of clonotypes identified for each of the recall and self-Ags used, our data suggest that, for this pairing at least, the frequency of responder cells in the memory pool is similar. Thus, our new approach to addressing an old question may prompt revision of the view that the frequency of TCRs with autoreactive potential is ultrarare.

A second hypothesis can also now be addressed with our approach: the degree to which there are clonality and publicity in autoreactive CD4 T cell responses. Early studies in preclinical models suggested that pathogenic T cells might arise from discrete cohorts of clonal populations sharing TCR structural features and, therefore, could be potentially amenable to TCR-specific mAb therapy (35). Subsequent studies in man showed a mixed picture but often relied upon TCRV gene identification or spectratyping of samples from relevant tissues rather than focusing on CDR3 sequence or Ag-specific CD4 T cells (27). Tissue-based studies are limited by the unknown Ag specificity of the T cells present in the sample, making interpretations of clonality difficult. It can be envisaged that, in the future, such tissue samples, which are increasingly available through bioresources, such as the Network for Pancreatic Organ Donors with Diabetes (36), can be clonotyped using our shallow HTS approach. If linked blood, spleen, and lymph node samples are also available, the Ag-specific sequences in responder populations can be compared with clonotypes found in the target tissue infiltrate. As an example, in type 1 diabetes, this would enable the pathogenic potential of peripheral β-cell–specific CD4 T cells to be evaluated through comparison with intraislet sequences; hitherto, few studies of Ag-specific clonotypes have been performed (37). Such studies are typically carried out on limited numbers of clones and subjects and carry a further potential functional bias in relation to the requirement for clones to be expandable from single cells before analysis. Our preliminary study suggests that public clonotypes specific for the exogenous and self recall Ags examined will be unusual but that TCR repertoires frequently include the potential for public responses, because we identified numerous examples of shared CDR3 β amino acid sequence arising via convergent recombination. The fact that these clonotypes are not found in all subjects as Ag responsive could be attributable to several factors, including sampling biases (38) and pairing with TCRVα-chains nonpermissive for the specific pHLA complex. Our findings are in keeping with recent in-depth studies of gluten epitope–specific CD4 T cells, which, despite being focused by repeated exposure to the same exogenous antigenic peptides presented by HLA-DQ8 and bearing high-affinity TCRs, show TRBV9*01 bias but very little conservation of residues in the CDR3 β region (39). The intense privacy of Ag-specific responses that we observed indicates that the development of clonotypic tracking for biomarker purposes may need to be subject specific. In memory clones, the TCRβ sequence alone is likely to be sufficient as a unique identifier for individual memory clones because each β-chain is estimated to combine with a single α-chain (16, 31). This makes it likely that we are detecting expansions of specific, private, Ag-specific T cell clones.

We are aware that there may be limitations to the approach described. An obvious consideration is the limited opportunity to sample the donor repertoire in a single blood draw (38). Sampling issues cannot be overcome using our approach and, therefore, it is expected that not all ASTs within a single patient will be identified. A further potential issue is the detection of AST clone sequences in common across subpopulations, raising the possibility that these arise as part of our cell-isolation protocol or other form of cross-contamination. However, we consider these possibilities unlikely, because studies showed that abundant clones are
observed in different T cell subsets, but this typically is not observed for low-frequency clonotypes (40). We elected to use RNA, rather than DNA, as the preferred starting material. Although RNA is more labile, the presence of multiple copies of mRNA encoding TCR Vβ-chains is known to obviate any potential bottle neck in TCR profiling that may be of particular concern in small populations of T cells (18). In addition, our analysis does not take into account the TCR Vα-chains in clonotype specificity. If the Vβ-chains that we characterized as belonging to ASTs pair with multiple Vα-chains, this will only serve to increase the number of private specificities. There also may be a greater opportunity for Vβ-chain diversity in autoreactive clones because of the presence of two α-chain rearrangements per cell. These further complexities can be addressed in the future using emerging technologies (41) and AST cell lines/clones. Finally, it is known that peptide stability determines the diversity of TCR clone response, with more stable peptides able to recruit many low-affinity clonotypes that are not recruited by less stable peptide–MHC complexes (42). We showed previously that the β-chain epitopes used in the current study have moderate to high affinity for MHC (9, 26) and, therefore, should be able to stimulate ASTs across a range of affinities. Whether these are the ASTs activated in vivo will require clarification in future studies.

Our approach to TCR repertoire analysis made use of Galaxy tools for CDR3 profiling and identifying and searching for AST clonotypes. However, future work in this area is likely to benefit from superior approaches to the bioinformatic challenge of TCR sequencing and clonotype searching in different subpopulations. Bespoke and open access software (such as TCRbase and MiTCR) (25, 43) are emerging that process and analyze data directly from sequencers. These platforms will enable cross-identity and cross–local alignment search tool analysis of multiple datasets of amino acid and nucleotide sequences to identify T cell clones with the same or similar Ag specificity and same or different origin.

In summary, we developed a tandem, multistep process to quantitate rare TCR β-chain variable sequences of Ag-specific CD4 T cells in large polyclonal populations. The approach combines deep and shallow HTS on the same blood samples and could be easily adapted to combined tissue/blood analyses. We find that clonotypes recognizing immunodominant regions of both pathogen-derived and self-Ags are oligoclonal and private and expanded into low-frequency cohorts in the memory population. We show that sequencing should be conducted at a threshold depth to enable detection of these low-frequency clonal expansions. Tandem HTS-based clonotyping will facilitate studying Ag-specific CD4 T cell dynamics, epitope spreading, and repertoire changes that arise post-vaccination and following Ag-specific immunotherapies for cancer and autoimmune disease.

Acknowledgments
We thank Amanda Bishop and Nicola McClintock (both from Southmead Hospital and the University of Bristol, Bristol, U.K.) for assistance in patient recruitment and Pros. David Price (Cardiff University, Wales, U.K.) and Danny Douek (National Institute of Allergy and Infectious Disease, National Institute of Health, Bethesda, MD) for assistance with clonotyping by template-switch anchored RT-PCR.

Disclosures
The authors have no financial conflicts of interest.

References