



## Generation of self-reactive, shared T-cell receptor $\alpha$ chains in the human thymus

Nelli Heikkilä<sup>a,\*</sup>, Silja Sormunen<sup>b</sup>, Joonatan Mattila<sup>a</sup>, Taina Härkönen<sup>c</sup>, Mikael Knip<sup>c,d,e,f</sup>, Emmi-Leena Ihantola<sup>g</sup>, Tuure Kinnunen<sup>g,h</sup>, Ilkka P. Mattila<sup>i</sup>, Jari Saramäki<sup>b</sup>, T. Petteri Arstila<sup>a</sup>

<sup>a</sup> Research Programs Unit, Translational Immunology, and Medicum, University of Helsinki, Haartmaninkatu 3, 00290, Helsinki, Finland

<sup>b</sup> Department of Computer Science, Aalto University, Konemiehenkatu 2, 02150, Espoo, Finland

<sup>c</sup> Pediatric Research Center, Children's Hospital, University of Helsinki and Helsinki University Hospital, Stenbäckinkatu 9, 00290, Helsinki, Finland

<sup>d</sup> Research Program for Clinical and Molecular Metabolism, University of Helsinki, Haartmaninkatu 8, 00290, Helsinki, Finland

<sup>e</sup> Folkhälsan Research Center, Topeliuksenkatu 25, 00250, Helsinki, Finland

<sup>f</sup> Department of Pediatrics, Tampere University Hospital, Elämäntie 2, 33520, Tampere, Finland

<sup>g</sup> Department of Clinical Microbiology, Institute of Clinical Medicine, University of Eastern Finland, Puijonlaaksontie 2, 70210, Kuopio, Finland

<sup>h</sup> Eastern Finland Laboratory Centre (ISLAB), Puijonlaaksontie 2, 70210, Kuopio, Finland

<sup>i</sup> Department of Pediatric Cardiac and Transplantation Surgery, Hospital for Children and Adolescents, Helsinki University Central Hospital, Stenbäckinkatu 9, 00290, Helsinki, Finland

### ARTICLE INFO

#### Keywords:

T cell receptor  
T cell recombination  
Thymus  
Thymic selections  
Autoantigen  
Autoimmunity

### ABSTRACT

The T-cell receptor (TCR) repertoire is generated in a semistochastic process of gene recombination and pairing of TCR $\alpha$  to TCR $\beta$  chains with the estimated total TCR diversity of  $>10^8$ . Despite this high diversity, similar or identical TCR chains are found to recur in immune responses. Here, we analyzed the thymic generation of TCR sequences previously associated with recognition of self- and nonself-antigens, represented by sequences associated with autoimmune diabetes and HIV, respectively. Unexpectedly, in the CD4<sup>+</sup> compartment TCR $\alpha$  chains associated with the recognition of self-antigens were generated in significantly higher numbers than TCR $\alpha$  chains associated with the recognition of nonself-antigens. The analysis of the circulating repertoire further showed that these chains are not lost in negative selection nor predominantly converted to the regulatory T-cell lineage. The high abundance of self-reactive TCR $\alpha$  chains in multiple individuals suggests that the human thymus has a predilection to generate self-reactive TCR $\alpha$  chains independently of the HLA-type and that the individual risk of autoimmunity may be modulated by the TCR $\beta$  repertoire associated with these chains.

### 1. Introduction

T-cell functionality is dependent on antigen recognition by a specific T-cell receptor (TCR), a heterodimeric cell surface receptor that binds to a complex formed by a peptide antigen in the groove of an HLA-molecule. The genes encoding the two TCR chains, TCR $\alpha$  and TCR $\beta$ , are assembled from gene segments in somatic gene recombination events during T-cell development, producing a highly diverse repertoire of surface TCR heterodimers [1]. The theoretical upper limit of TCR diversity has been calculated to range from  $10^{15}$  to  $10^{20}$ , greatly exceeding the total number of T cells in the human body, while the actual human TCR repertoire has been estimated to be in excess of  $10^8$  [2–5]. Most of this diversity is concentrated in the complementarity determining region 3 (CDR3), the junctional section generated by

imprecise joining of the gene segments. Moreover, structurally the hypervariable loops encoded by the CDR3 sequences are mainly responsible for binding to the antigenic peptide, and thus likely to be the most important determinants of specificity [6].

Given this high diversity, the default expectation has been that the TCR repertoire in each individual would be largely unique, or private. However, a large and growing body of research has shown that different individuals often use similar or same TCRs in antigen recognition at least for one chain of the heterodimer, whether the specificity be against viral, self- or tumor proteins [7,8]. This phenomenon of shared, or public clones has been especially prominent in chronic or latent viral infections, such as cytomegalovirus or herpes simplex virus [9]. Several studies also suggest that the public responses may be qualitatively different from private responses. For example, in some murine models of

\* Corresponding author. Haartmaninkatu 3, 00290, Helsinki, Finland.

E-mail address: [nelly.heikkila@helsinki.fi](mailto:nelly.heikkila@helsinki.fi) (N. Heikkilä).

<https://doi.org/10.1016/j.jaut.2021.102616>

Received 25 November 2020; Received in revised form 4 February 2021; Accepted 7 February 2021

Available online 27 February 2021

0896-8411/© 2021 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

autoimmunity the destructive autoreactive responses are dominated by public clones [10,11], and studies of long-term HIV-positive non-progressors have identified highly cross-reactive public responses controlling HIV replication [12].

Underlying such public responses is the realization that a surprisingly large fraction of the naive T-cell repertoire can be shared by unrelated individuals. For the TCR $\beta$  chain this overlap has recently been reported to be up to 10%, and for TCR $\alpha$  chain 26.5% in peripheral blood [13], and in the thymus 6.1% and 46.7% for TCR $\beta$  and TCR $\alpha$  chains, respectively [14]. Much of this sharing has been suggested to result from convergent recombination in the thymus, i.e., the biased generation of TCR sequences converging on a number of chains that arise repeatedly, often because they contain little or no junctional modifications [7]. Here, we have studied the thymic generation of TCR chains previously identified to be associated with pathological conditions and show that many of these can also be found in the thymus of unrelated individuals. Surprisingly, our results show that the human thymus is predisposed to generating self-reactive TCR $\alpha$  chains that persist in the periphery in the population of conventional T cells.

## 2. Materials and methods

### 2.1. Samples

The study and all sample collections were approved by the Ethics Committee of the Hospital District of Helsinki and Uusimaa and a written informed consent was obtained from the subjects or their parents. Six thymus resects were obtained from immunologically healthy infants undergoing corrective cardiac surgery for a congenital heart defect (aged 4–8 months, 2/6 were female). Two of the thymus donors (samples A and B) were monozygous twins and the genetic impact on the TCR repertoire has been analyzed previously [15]. Thymocytes were mechanically extracted from the resects and aliquots of 10 million thymocytes were stored as dry pellets in  $-70^{\circ}\text{C}$ .

The peripheral blood samples were derived from the Finnish Pediatric Diabetes Register [16]. The register collects data and biological samples from children with newly diagnosed type 1 diabetes (T1D) and their first-degree relatives. The participants are screened for HLA-conferred susceptibility to T1D and tested for five T1D-associated autoantibodies [islet cell antibodies (ICA), insulin autoantibodies and antibodies to glutamic acid decarboxylase 65 (GADA), islet-antigen 2 (IA2A), and zinc transporter 8 (ZnT8A)]. The current study includes ten siblings (aged 3–14 years, 5/10 female) with a heterozygous HLA-DR3/DR4 genotype. The children tested negative for the T1D-associated autoantibodies and were free of clinical symptoms of diabetes at the time of sampling and also for a period of 2–4 years after the sampling. Peripheral blood mononuclear cells (PBMC) were isolated using Ficoll Paque Plus (GE Healthcare, Chicago, Illinois, USA) gradient centrifugation and an aliquot of five million cells were frozen in  $-140^{\circ}\text{C}$  with FCS-DMSO freezing media.

The blood samples for sorting of regulatory T cells (Treg) and conventional T cells (Tconv) were obtained from three immunologically healthy adults (aged 21–30 years, 1/3 female). PBMCs were isolated with Ficoll Paque Plus gradient centrifugation. Freshly isolated PBMCs were stained for flow cytometric cell sorting for 30 min in  $4^{\circ}\text{C}$  with the following cell surface markers: LIVE/DEAD Fixable Green Dead Cell Stain (DCM) in 1:1000 concentration (Thermo Fisher Scientific, Waltham, Massachusetts, USA), PeCy7-labeled mouse anti-human CD25 in 1:50 concentration (clone BC96, eBioscience, San Diego, California, USA) and BV510-labeled mouse anti-human CD4 in 1:50 concentration (clone SK3, BD Biosciences, San Diego, California, USA). The sorting was performed using BD FACSAria II sorter (BD Biosciences, San Diego, California, USA) and the sorted populations defined as DCM-CD4 $^{+}$ CD25 $^{\text{high}}$  for Tregs and DCM-CD4 $^{+}$ CD25 $^{-}$  for Tconvs. The sorting yielded 83 000–200 000 Tregs and 1.4–2.5 million Tconvs per sample with a purity of  $>95\%$ . Sorted cells were stored as dry pellets in  $-70^{\circ}\text{C}$ .

### 2.2. Sequencing

DNA was extracted from frozen thymocytes and PBMCs using DNeasy or QIAasympphony kits (both from Qiagen, Hilden, Germany) according to the manufacturer's instructions. The sequencing of the TCR $\alpha$  and TCR $\beta$  sequences was performed with ImmunoSEQ sequencing service (Adaptive Biotech, Seattle, Washington USA) as previously described [5]. The sequencing assay uses a standardized quantity of quality-controlled DNA and consists of a multiplex PCR assay that spans recombinant TCR $\alpha$  and TCR $\beta$  genes at a sufficient length to cover the entire CDR3 region and to identify VJ for TCR $\alpha$  gene and VDJ for TCR $\beta$  gene. Amplicon sequencing was performed on the Illumina platform. TCR $\alpha$  and TCR $\beta$  definitions were based on the IMGT database ([www.imgt.org](http://www.imgt.org)). Primer bias was corrected using a synthetic repertoire of TCRs, and barcoded, spiked-in synthetic templates. All sequence datasets analyzed in the current study are available at the immuneACCESS data repository ([clients.adaptivebiotech.com/immuneaccess](http://clients.adaptivebiotech.com/immuneaccess)).

### 2.3. Database

Epitope-specific TCRs were collected in the literature and were principally issued from the previously published databases Mc-PAS-TCR [17] and VDJDdb [18]. Some manually collected TCRs were also included [19–22]. We only accepted TCRs for which the specificity had been established by HLA multimer-staining or verified by peptide stimulation assays. The entire list of epitope-specific TCRs is available in [Supplement 1](#).

### 2.4. TCR analysis

TCR sequences were downloaded in the format of immunoSEQ files with Export Sample (v2) (Adaptive Biotechnologies, Seattle, Washington, USA). The scripts to analyze sequence overlap and to search for the epitope-specific TCRs in the sequenced repertoires were written in programming languages R ([www.r-project.org](http://www.r-project.org)) and Python ([www.python.org](http://www.python.org)). All scripts are available on request. To measure the overlap between two sequence datasets, we used the Jaccard index, which is defined as the intersection of two datasets divided by their union. For random subsampling of the epitope-specific TCRs to a desired sample size we used Python `random.sample` of the `random` package.

Clonal overlap was assessed calculating the Jaccard index, which is defined as the size of the intersection of two data sets (A and B) divided by the size of their union:  $J(A, B) = \frac{|A \cap B|}{|A \cup B|}$ . The averages of the detection rate and of the abundance of TCRs related to self and nonself were compared applying two-sided Wilcoxon signed rank test. Statistical significance was defined as  $p < 0.05$ . The statistics were calculated using SPSS Statistics Software version 24.

## 3. Results

### 3.1. TCR clone distribution in thymus samples

Thymus samples were obtained from six immunologically healthy infants (A–F) during corrective cardiac surgery. From each donor an aliquot of 10 million mechanically extracted thymocytes was sequenced for TCR $\alpha$  and TCR $\beta$  chains. Two of the donors (A and B) were monozygotic twins. As previously reported, their TCR repertoires showed a genetic impact in the V and J gene usage and in the generation of junctional sequences, but not in thymic selection [15]. On average, we obtained 5.9 and 1.1 million unique TCR $\alpha$  and TCR $\beta$  nucleotide sequences, respectively. Of the unique TCR $\alpha$  chains 31.2% and of the unique TCR $\beta$  chains 78.8% were in-frame ([Supplement 2](#)). For unknown reasons, sequencing of sample E produced clearly fewer TCR sequences than the other samples.

In peripheral repertoires, the TCR clone sizes display broad

distributions and are thought to follow a power law where the occurrence of a clone varies as a power of its size [23]. This phenomenon was particularly striking within the thymic TCR $\alpha$  repertoires where the probabilities of clone sizes display a straight line on a double logarithmic scale, which is a hallmark of a power law distribution (Fig. 1A). Consequently, the thymic TCR repertoire contains some remarkably high-abundance TCR $\alpha$  clones. The most abundant clones are represented on the average by 872 genomes (range 106–1249), while the majority of clones had a clone size of only one. The number of non-templated inserts between the TRAV and TRAJ segments was lower in the high-abundance clones than in the entire repertoire suggesting that they are closer to the germline and more easily generated (Fig. 1B). Also, clonotype sharing was more common among the highly abundant clonotypes as measured by the Jaccard index, the intersection of two samples divided by their union (Fig. 1C). Indeed, the most abundant clonotypes were typically shared among all individuals and displayed a large clonal size in all of them (Fig. 1D). A broad distribution of clone sizes and a lower number of non-templated insertions among the most abundant clones were observed in the TCR $\beta$  repertoire, although less clearly than in the TCR $\alpha$  repertoire (Fig. 1A, Supplement 3).

### 3.2. Database of TCRs associated with self- and nonself-epitopes

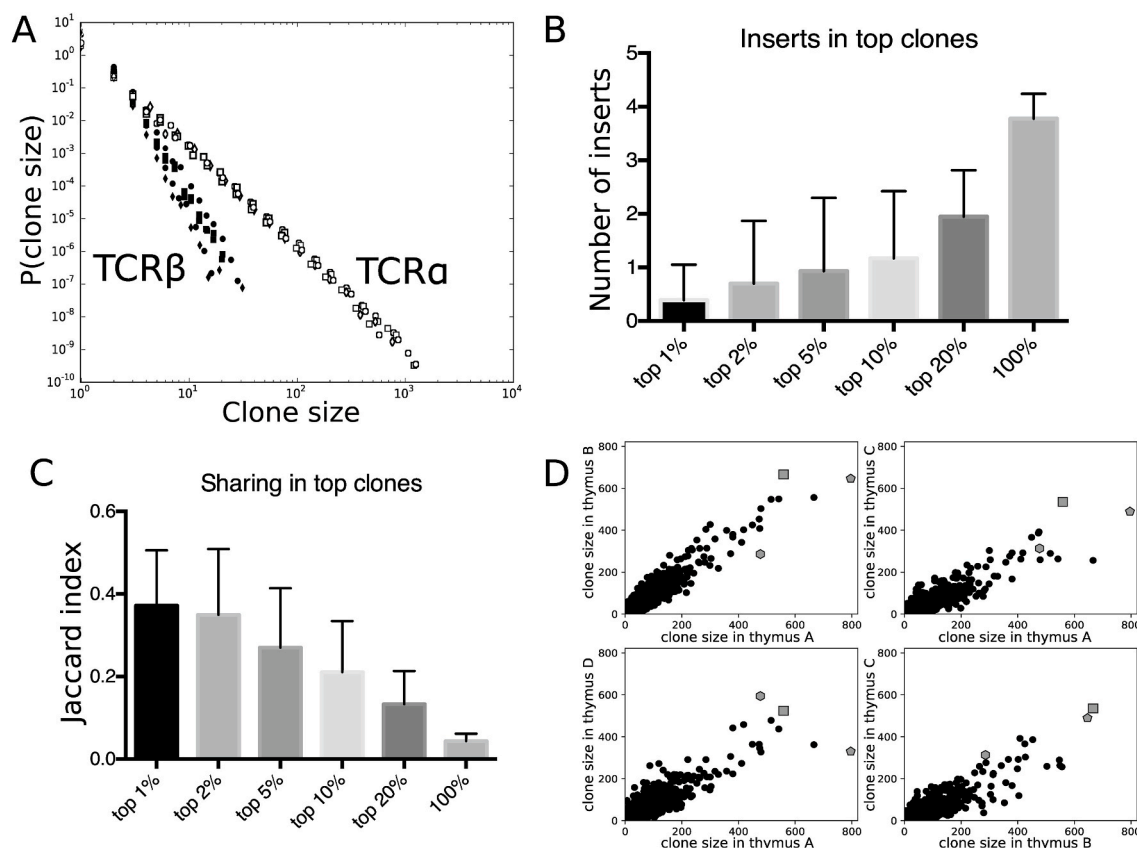
We then performed a literature search to create a database of TCR sequences associated with a known specificity, concentrating on two categories of antigens: First, TCRs identified to be specific for self-epitopes, and second, TCRs specific for nonself-epitopes. The majority of TCRs was derived from two previously published databases Mc-PAS-TCR [17] and VDJdb [18] but manually searched sequences specific to

islet-antigens were also included [19–22]. In the final database the self-reactive TCR chains were limited to those recognizing islet antigens, since TCR $\alpha$  chains with other specificities were particularly scarce in the literature. As a representative nonself-antigen we selected HIV, because unlike herpesviruses or influenza, our HIV-negative donors had not encountered its antigens. Further, we only included TCRs for which the specificity was determined with sufficient rigor. Our final reference database consisted of 546 unique TCR $\alpha$  clones and 2407 unique TCR $\beta$  clones that were specific for T1D- or HIV-related epitopes (Supplement 1). We further divided the TCRs to CD4<sup>+</sup> and CD8<sup>+</sup> subsets depending on the CD4/CD8 phenotype and HLA class I/II restriction of the original T cells from which the TCR was derived (hereafter CD4<sup>+</sup> and CD8<sup>+</sup>; Table 1). The majority of the T1D-associated TCRs were derived from CD4<sup>+</sup> T cells restricted by the HLA-DR4 or HLA-DR3 haplotypes associated with a high T1D risk [24].

**Table 1**

A summary of T1D- and HIV-specific TCR $\alpha$  and TCR $\beta$  sequences in the reference database. The sequences are divided into CD4 and CD8 categories according to the phenotype of the cell from which the TCR clonotype was originally derived.

		T1D	HIV
TCR $\alpha$	CD4	130	184
	CD8	80	152
TCR $\beta$	CD4	71	409
	CD8	108	1819



**Fig. 1.** The features of highly abundant thymic TCR $\alpha$  clonotypes. The distribution of clone sizes in TCR $\alpha$  (open symbols) and TCR $\beta$  (black symbols) repertoires. The TCR $\alpha$  clone size distributions show a regular slope of approximately  $-3$  on the log-log scale for  $n = 5$ , the less deeply sequenced sample E being excluded (A). The number of non-templated insertions (B) and the interindividual clonotype sharing measured with Jaccard index (C) for the most abundant top 1%, 2%, 5%, 10%, 20% TCR $\alpha$  clonotypes and for the entire (100%) repertoire,  $n = 6$ . TCR $\alpha$  clone size correlations between two individuals and the same highly abundant TCR $\alpha$  clonotypes indicated, representative plots are shown (D).

### 3.3. Preferential thymic generation of TCRα clonotypes used by CD4<sup>+</sup> T cells associated with T1D

To compare TCR sequences in our reference database with the thymus samples, an exact match with CDR3 amino acid sequence and matching TCR V- and J-gene segments were required. Using these criteria, on average 24.1% of the 184 HIV-associated CD4<sup>+</sup> TCRα chains in the database were found in the thymus samples. Unexpectedly, a significantly higher fraction of the T1D-associated CD4<sup>+</sup> TCRα chains was detected in the thymus samples (32.8% of 130 sequences in the database, Wilcoxon signed rank test  $p = 0.028$ ; Fig. 2A). On the average, each CD4<sup>+</sup> T1D-associated amino acid chain was encoded by 2.5 unique nucleotide sequences while each CD4<sup>+</sup> HIV-associated amino acid chain was encoded by 2.1 unique nucleotide sequences ( $p = 0.028$ ; Fig. 2B). Moreover, since the sequencing assay is based on genomic DNA, it allows a reasonable estimate of the clonal copy numbers. The average clone size corresponding to T1D-associated TCRα amino acid chains was significantly larger than HIV-associated chains (12.4 vs. 9.9,  $p = 0.028$ ; Fig. 2C).

For TCRα chains derived from CD8<sup>+</sup> T cells, amino acid sequences associated with T1D were found less frequently in the thymic samples than those associated with HIV (25.2% of 80 sequences in the database vs. 37.2% of 152 sequences in the database,  $p = 0.028$ ), and there was no difference in the average number of nucleotide sequences encoding them or in the average clonal size (Fig. 2B&C).

To ascertain that the results were not skewed by the different number of T1D- and HIV-associated sequences in our database, we drew five random subsets of HIV-associated sequences matched to the size of the T1D-sequence sets and used these to search for matching sequences in the thymus samples. Again, CD4<sup>+</sup> T1D-associated TCRαs were detected

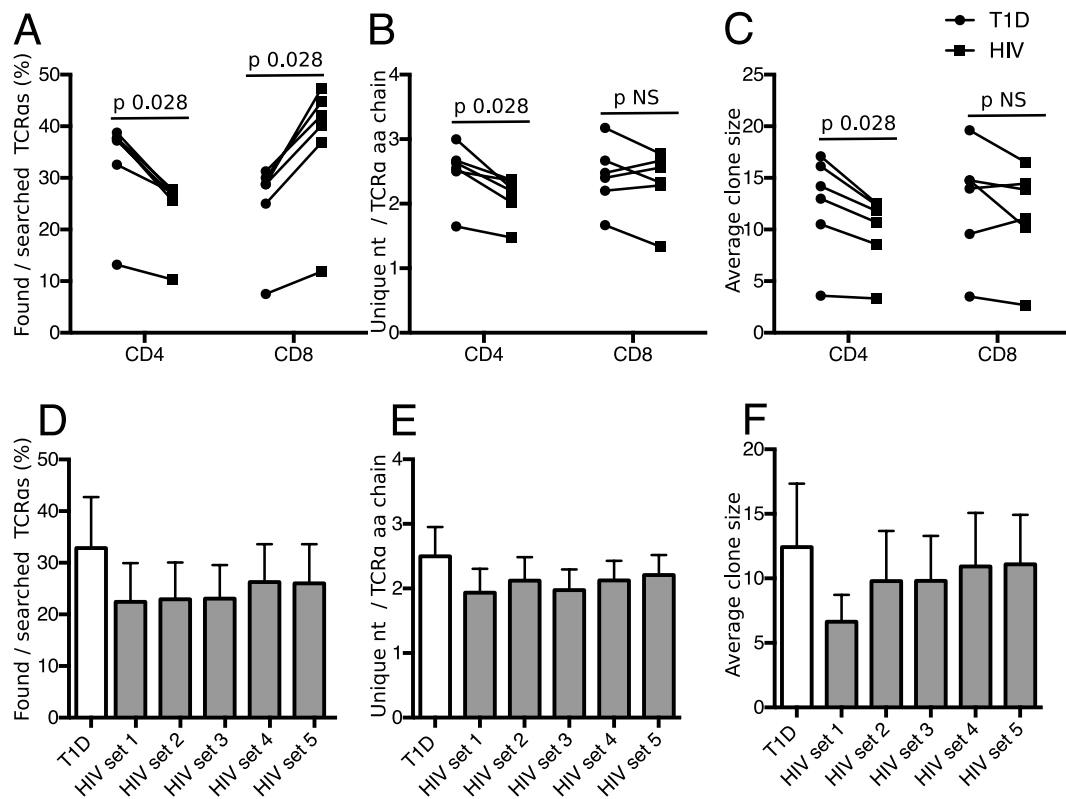
more frequently than the HIV-associated TCRαs (32.8% vs. mean of five resampled HIV-TCR sets 24.2%,  $p = 0.028$ ; Fig. 2D), whereas CD8<sup>+</sup> HIV-associated TCRαs were more frequent than T1D-associated TCRαs (25.2% vs. mean of five resampled HIV-TCR sets 40.3%,  $p = 0.028$ ; data not shown). The number of unique nucleotide sequences was again higher for CD4<sup>+</sup> T1D- than HIV-associated chains (2.5 vs. mean of five HIV resamplings 2.1,  $p = 0.028$ ; Fig. 2E), as was the average clonal size (12.4 vs. mean of five HIV resamplings 9.6,  $p = 0.028$ ; Fig. 2F).

Overall, the average combined clone size of the CD4<sup>+</sup> T1D-associated TCRα chains identified in the thymus samples was 577. The average combined clone size of the HIV-associated TCRα chains, measured from the database size-matched resamplings, was 328 ( $p = 0.028$ ).

Despite the relatively high number of TCRβ chains in the reference database, there were very few matches with the thymus samples (Supplement 4). For both CD4<sup>+</sup> and CD8<sup>+</sup> sequences only 1–2 matches of T1D-associated sequences were found in 4/6 thymus samples, each encoded by a single nucleotide sequence and with a clone size of one. CD4<sup>+</sup> HIV-associated sequences were similarly rare in the thymuses, with 1–3 matches in 4/6 thymus samples, encoded on the average by 1.3 nucleotide sequences and with an average clone size of 1.4. CD8<sup>+</sup> HIV-associated sequences were found slightly more often. On average, 1.8% of 1819 HIV-associated sequences were found in the thymuses, encoded by 1.4 nucleotide sequences and with a clone size of 1.8. Thus, the TCRβ sequences were not analyzed further.

### 3.4. T1D-associated TCRα chains persist in the periphery

Since the increased frequency of T1D-associated TCRα chains was detected in unselected thymocyte samples, it was possible that these sequences were obtained from immature thymocytes before negative



**Fig. 2.** Epitope-specific TCRα sequences in thymus samples. The frequency of CD4<sup>+</sup> and CD8<sup>+</sup> T1D- and HIV-associated TCRα reference database sequences found in the six thymus samples (A), the average number of unique nucleotide sequences encoding each identified amino acid chain (B), and their average clone sizes (C). The T1D- and HIV-associated sequences are indicated with circles and squares, respectively. Wilcoxon signed rank test was applied to compare the means between the two groups. Comparison of the average frequency of CD4<sup>+</sup> T1D-associated sequences (D), the average number of unique nucleotide sequences encoding each identified T1D-associated amino acid chain (E), and their average clone sizes (F) with the HIV-associated reference database resampled five times to the size of the T1D database. The bars show the average and standard deviation. In (A-C) the six individual thymus samples are shown, in (D-F)  $n = 6$ .

selection, so that the clones might be deleted before thymic egress. We therefore studied the presence of the T1D- and HIV-associated TCRα chains in peripheral blood samples from ten healthy children, aged 3–14 years. All donors had the T1D high-risk HLA-DR3 and DR4 haplotypes, but no autoantibodies to islet antigens (GADA, ICA, IA2A, IAA, Znt8A), and no signs of clinical T1D during a follow-up for 2–4 years after sampling. From these samples we obtained an average of 825 000 total TCRα reads, corresponding to 587 000 unique nucleotide sequences (Supplement 2).

Again, both T1D- and HIV-associated TCRαs were readily detected in the peripheral blood samples. Similarly to the thymus, for CD4<sup>+</sup> sequences T1D-associated TCRαs were detected more often than HIV-associated TCRαs (T1D 15.7% of the 130 sequences in the database vs. HIV 8.4% of 184 sequences in the database, *p* 0.005), while the opposite was observed for CD8<sup>+</sup> reference sequences (T1D 10.8% of the 80 sequences in the database vs. HIV 18.2% of the 152 sequences in the database *p* 0.005; Fig. 3A). The number of unique nucleotide sequences encoding each detected amino acid chain was higher for CD4<sup>+</sup> T1D- than HIV-associated TCRαs (1.5 vs. 1.3, *p* 0.013; Fig. 3B), but the average clone size was similar (Fig. 3C). For the reference sequences obtained from CD8<sup>+</sup> T cells, the number of unique nucleotide sequences encoding the detected amino acid chains was similar, but some of the HIV-associated clones appeared to be expanded and thus the average clone size was higher for HIV-associated than T1D-associated sequences (17.7 vs. 3.4, *p* 0.005). The children participating in our study were all HIV-seronegative.

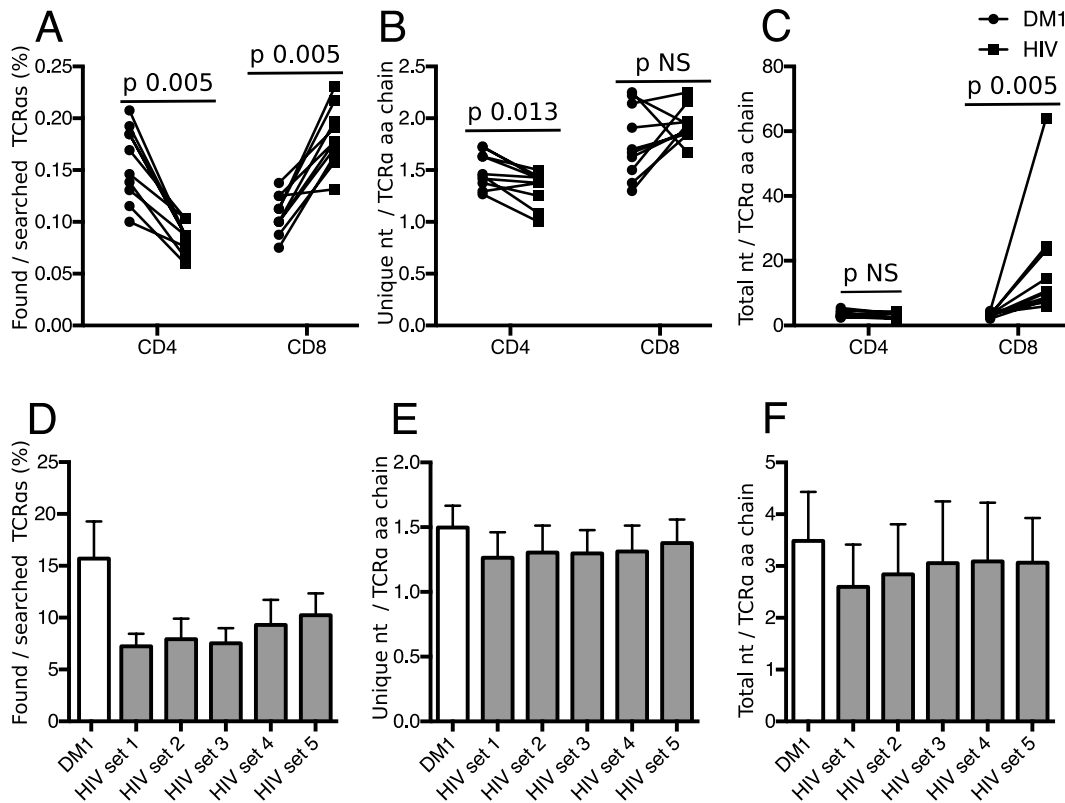
Finally, we repeated the random sampling of the HIV reference database to the same number of sequences contained in the T1D database, as described for the thymus samples. The sampling of HIV-related sequences was again done five times and the average numbers compared

with the T1D data, with essentially similar results to those obtained with the full reference database (Fig. 3D–F). In the peripheral samples the average combined clone size of CD4<sup>+</sup> T1D-associated TCRα chains was 72.9 and that of HIV-associated sequences, measured after matched resampling, was 32.9 (*p* 0.005).

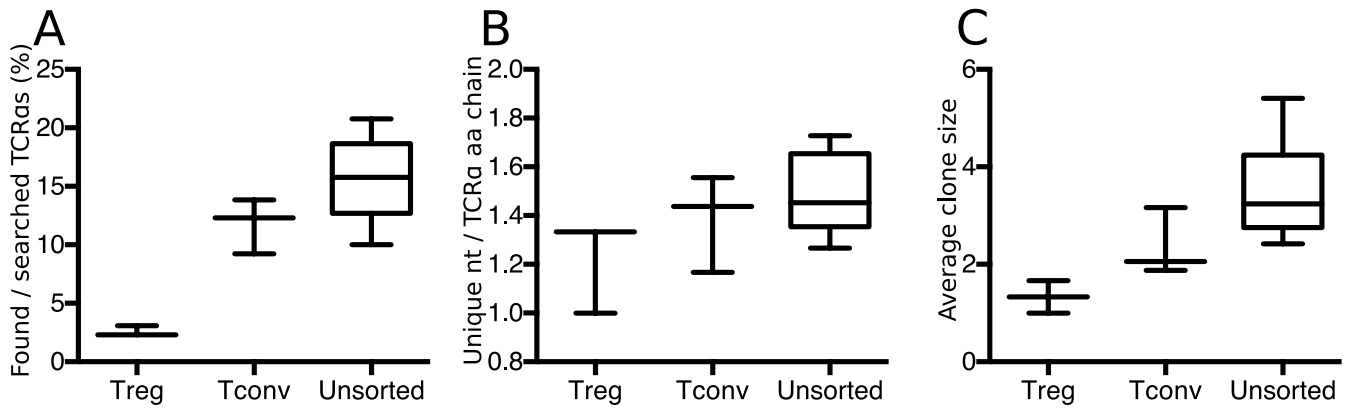
### 3.5. T1D-associated TCRα clonotypes are not enriched within peripheral blood regulatory T cells

Another possibility was that, although originally identified in effector T cells, some of the T1D-associated sequences might be diverted to the regulatory T cell (Treg) lineage. We sorted conventional T (Tconv) cells (CD3<sup>+</sup>CD4<sup>+</sup>CD25<sup>high</sup>) and Tregs (CD3<sup>+</sup>CD4<sup>+</sup>CD25<sup>−</sup>) from peripheral blood samples from three immunologically healthy adults, aged 21–30 years (Supplement 5). TCRα sequencing of the two cell subsets produced on average 308 000 and 53 000 total nucleotide reads corresponding to 247 000 and 39 000 unique sequences for Tconv and Treg populations, respectively (Supplement 2).

Since the TCRα sequences were derived from sorted CD4<sup>+</sup> T cells, we only searched the samples for the CD4<sup>+</sup> sequences in the reference database. In the Treg samples only 3–4 T1D-associated reference sequences were detected (2.6% of the 130 sequences in the database; Fig. 4A), most likely because of the smaller number of unique sequences in the samples. These TCRα chains were represented by few unique clones (average 1.3; Fig. 4B) and their total genome number was low (average 1.2; Fig. 4C). In the Tconv samples, the average detection rate of T1D-associated chains was 11.8% of the 130 sequences, well within the range of the detection rate in the unsorted peripheral blood samples (10.0–20.8%). Also, the number of unique nucleotide sequences encoding the T1D-associated amino acid chains and the average clone



**Fig. 3.** Epitope-specific TCRα sequences in peripheral blood samples. The frequency of CD4<sup>+</sup> and CD8<sup>+</sup> T1D- and HIV-associated TCRα reference database sequences found in the ten peripheral blood samples (A), the average number of unique nucleotide sequences encoding each identified amino acid chain (B), and their average clone sizes (C) The T1D- and HIV-associated sequences are indicated with circles and squares, respectively. Wilcoxon signed rank test was applied to compare the means between the two groups. Comparison of the average frequency of CD4<sup>+</sup> T1D-associated sequences (D), the average number of unique nucleotide sequences encoding each identified T1D-associated amino acid chain (E), and their average clone sizes (F) with the HIV-associated reference database resampled five times to the size of the T1D database. In (A–C) the ten individual donors are shown, in (D–F) *n* = 10.

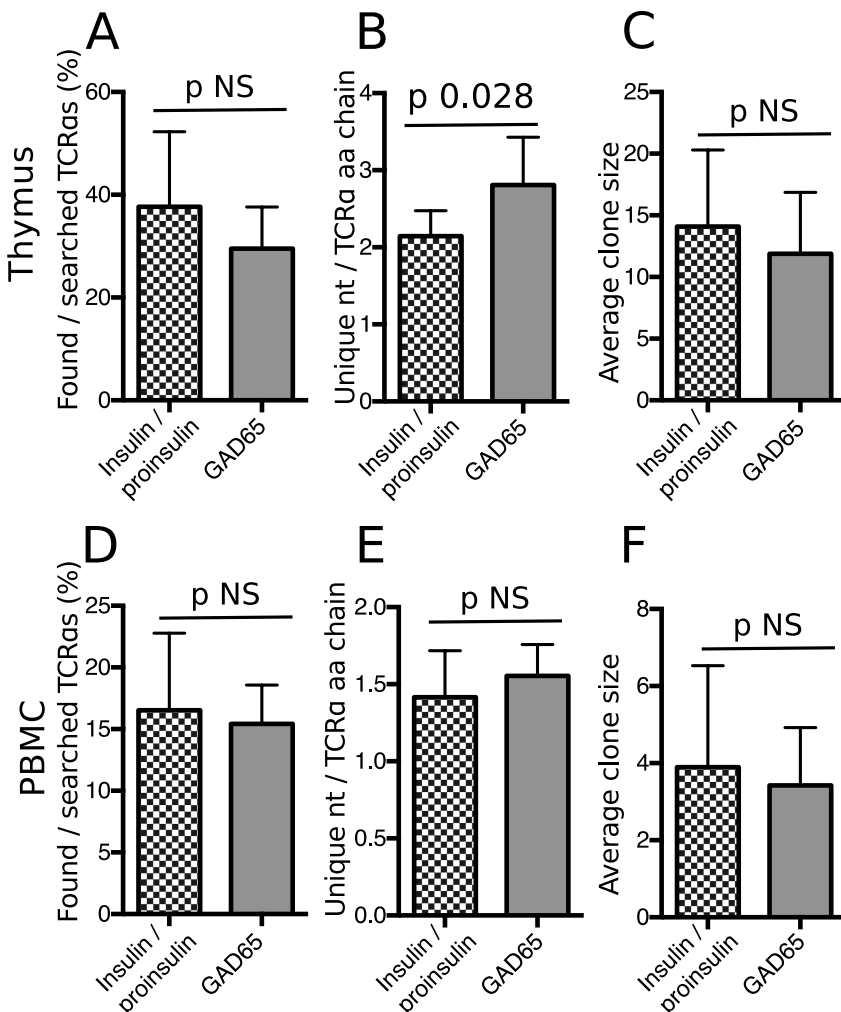


**Fig. 4.** Analysis of the CD4<sup>+</sup> T1D-associated TCRα sequences in peripheral blood Treg and Tconv populations. The frequency of T1D-associated reference database sequences found in Treg, Tconv and unsorted cells (A), the average number of unique nucleotide sequences encoding each identified amino acid chain (B), and their average clone sizes (C). Median values are indicated by the horizontal line, the box indicates 25–75% percentiles and the whiskers minimum and maximum values. For Treg and for Tconv n = 3, for unsorted n = 10.

size was similar among sorted Tconv and unsorted peripheral blood samples (Fig. 4). Thus, the depletion of Treg cells did not result in the disappearance of the T1D-associated sequences.

Also, many of the T1D-associated sequences were found in both subsets. Altogether eight and 26 unique T1D-associated TCRα amino acid chains were detected in the Treg and Tconv samples, respectively. Six of these TCRs were found both in Treg and Tconv population.

Of the CD4<sup>+</sup> HIV-associated TCRα chains only one was detected in the Treg samples. It was encoded by a single unique nucleotide sequence in each sample and the average clone size was 3.7. In the Tconv samples the frequency of CD4<sup>+</sup> HIV-associated TCRα chains as well as their nucleotide sequence numbers, and the average clone sizes were within the range of the values observed for unsorted peripheral blood samples (Supplement 6).



**Fig. 5.** Comparison of CD4<sup>+</sup> insulin/proinsulin-with GAD65-associated TCRα chains. The average frequency of insulin/proinsulin- and GAD65-associated TCRα reference database sequences found in the six thymus samples (A), the average number of unique nucleotide sequences encoding each identified amino acid chain (B), and their average clone sizes (C). Wilcoxon signed rank test was applied to compare the means between the two groups. The average frequency of insulin/proinsulin- and GAD65-associated TCRα reference database sequences found in the ten peripheral blood samples (D), the average number of unique nucleotide sequences encoding each identified amino acid chain (E), and their average clone sizes (F). In (A-C) n = 6, in (D-F) n = 10.

### 3.6. Generation of GAD65-and insulin/proinsulin-associated TCR $\alpha$ chains

For a self-antigen to have an effect on thymocyte selection it has to be present in the thymus [25]. Of the islet antigens insulin has been conclusively demonstrated to be expressed as a tissue-specific antigen in thymic antigen-presenting cells, whereas for GAD65 conflicting reports have been published [26–28]. Our T1D reference database contained 46 CD4<sup>+</sup> TCR $\alpha$  chains associated with insulin/proinsulin and 83 associated with GAD65. One chain was derived from a T cell specific for islet-specific glucose-6-phosphatase catalytic subunit-related protein (IGRP) and was not analyzed separately.

Of the CD4<sup>+</sup> TCR $\alpha$  chains associated with GAD65 on average 29.5% were detected in the thymus samples and of those associated with insulin/proinsulin 37.7%, but variation was considerable and the difference not statistically significant ( $p$  0.12; Fig. 5A). The average number of unique nucleotide sequences encoding the GAD65-associated TCR $\alpha$  amino acid chains was slightly higher than that of insulin/proinsulin (2.8 vs 2.1,  $p$  0.028; Fig. 5B), but the average clone sizes were similar (11.9 vs. 14.1,  $p$  0.17; Fig. 5C). Resampling of the GAD65 reference database five times to match the size of the insulin/proinsulin database produced similar results (Supplement 7).

In the peripheral blood samples, no differences between TCR $\alpha$ s associated with GAD65 and insulin/proinsulin were detected (Fig. 5D–E), and sequences associated with both antigens were found at a similar frequency in sorted Tconv cells.

## 4. Discussion

With the exception of rare inherited forms of autoimmunity, autoimmune diseases require environmental influences [29]. Infections are well established as triggering autoimmunity, most clearly shown in rheumatic fever after streptococcal pharyngitis or spondyloarthropathies after bacterial enteritis in individuals carrying the HLA-B27 haplotype [30–32]. Increasing evidence also links viral infections to autoimmunity. Epstein-Barr virus infection seems to be required, though not alone sufficient, for the development of multiple sclerosis, and it has also been linked to other autoimmune diseases, including systemic lupus erythematosus, rheumatoid arthritis, and Sjögren's syndrome [33,34]. More recently emerged pathogens have also been implicated. Human immunodeficiency virus increases the risk of systemic autoimmunity even when viral replication is suppressed [35,36], while the current outbreak of COVID-19 is also associated with autoimmune manifestations [37,38]. Infections have also been linked to type 1 diabetes and the pancreatic autoantigens analyzed in the present study. In particular, Coxsackie B enterovirus can trigger pancreatic autoimmunity in animal models and in humans it has been linked to cases in which the auto-reactivity first targets GAD65 instead of insulin/proinsulin [39–41]. The evidence supporting the role of infections in the development of autoimmunity is thus strong, and the suggested pathogenetic mechanisms, although yet poorly understood, include molecular mimicry, bystander activation and epitope spreading [42].

However, it is clear that the individual risk of developing autoimmunity is modified by a complex interplay of genetic susceptibility and other environmental factors, as well. The general increase in the prevalence of autoimmune diseases in industrialized countries has been explained by changes in the early microbial exposure. The hygiene hypothesis postulates that in the absence of previously prevalent microbial pathogens, the immune system is prone to react inappropriately to self-antigens [43]. To this background our results add a potential new risk factor for autoimmunity: The thymic generation of shared, autoantigen-associated TCR $\alpha$  chains and their semistochastic pairing to TCR $\beta$  partners.

The possibilities to define the epitope-specificity of TCRs have recently evolved from the experimental characterization of HLA tetramer-binding T cells to computational predictions in large T-cell

repertoires. The TCR repertoires with predicted epitope-specificity typically contain public receptor chains with high generation probabilities and form clusters according to the sharing of amino acid motifs [44]. Here, we have utilized experimentally well validated epitope-specific TCR chains to investigate the thymic generation and selection of public TCR clonotypes. We have previously reported that a remarkably large fraction of preimmune thymic TCR $\alpha$  repertoire is shared between unrelated individuals. These public TCR $\alpha$ s had fewer non-templated nucleotides, higher generation probabilities than the non-shared repertoire and the clone sizes were often large [14]. In the TCR $\beta$  chain the interindividual sharing was much lower, probably because the presence of two gene junctions decreases the likelihood of convergent recombination. These findings are also likely to explain why in the present study the epitope-specific public repertoire is so conspicuous in the TCR $\alpha$  but not in the TCR $\beta$  repertoire.

In our data a substantial fraction of the T1D-associated TCR $\alpha$  amino acid sequences from the reference database were found in the thymus, and for CD4<sup>+</sup> reference sequences they were indeed more common than HIV-associated sequences, were encoded by a higher number of unique nucleotide sequences and displayed a larger average clone size. Although the differences were not large, it is notable that they were highly consistent and found in every one of the thymus samples. This latter fact also shows that the phenomenon is independent of HLA type, since with the exception of one pair of twins, the samples were obtained from unrelated individuals. This may be due to the fact that HLA binding is mainly mediated by hypervariable loops encoded by germline sequences in the V genes, and thus likely to affect mostly V gene usage patterns [6]. Indeed, the aspect of TCR repertoire shown to be most strongly affected by heritable factors is TCR gene segment usage [13,15, 45].

Our data also clearly showed that the higher detection rate of CD4<sup>+</sup> T1D-associated sequences persisted in the periphery and in Treg-depleted cells, indicating that the chains were not lost in negative selection; nor were they diverted to the regulatory lineage. This was also true for TCR $\alpha$  chains associated with insulin/proinsulin, a self-antigen expressed by thymic epithelial cells for the purposes of thymic selection [26,27,46]. Thus, at the very least, thymic tolerance mechanisms appear to have no negative impact on the public component of islet-reactive TCR $\alpha$  chains. Indeed, thymic repertoire generation seems to even favor them, in every thymus sample we analyzed. Because of the scarcity of well-defined autoreactive TCR $\alpha$  chains in the literature, we cannot say whether these findings are unique to islet antigens or a more general feature of TCR $\alpha$  chains specific to other self-antigens, as well. Furthermore, our cohort of pediatric donors was not screened for other than islet-associated autoantibodies, so we cannot address their putative susceptibility to other forms of autoimmunity. Nevertheless, all donors were clinically healthy.

An obvious point regards the TCR $\beta$  partners of the public TCR $\alpha$  repertoire, which remain unknown in our data. Despite advances in determining the sequences of TCR $\alpha$ / $\beta$  pairs, the clone size of our public chains is too small for experimental resolution. In the thymus the average frequency of all the T1D-associated TCR $\alpha$  chains combined was  $6/10^5$  and in the periphery  $9/10^5$ , well below the detection threshold of current techniques. The studies so far published indicate that public heterodimers are rare, and repertoire sharing typically concerns only one of the chains [47]. Since antigen recognition, as well as thymic selection mostly works on the heterodimer level, the impact of either of the partner chains alone is limited. However, substantial experimental evidence shows that in some cases the specificity of the TCR may be determined predominantly by one of the chains, or that a public TCR $\alpha$  or TCR $\beta$  chain can promiscuously pair with multiple different chains and still preserve the antigen specificity. This was already apparent in our database of epitope-specific TCRs, where identical TCR $\alpha$  chains paired with multiple different TCR $\beta$  chains and vice versa (Supplement 1). Furthermore, Nagatsugawa and colleagues showed directly that a given TCR $\alpha$  chain can be enough to confer specificity, while the TCR $\beta$  partner

modifies the TCR affinity over a range of two orders of magnitude [48]. In another interesting example, TCRs isolated from patients with beryllium disease expressed either of two conserved TCR $\beta$  chains, differing by a single amino acid. One of these TCR $\beta$  chains was enough to confer high affinity, allowing promiscuous pairing with a variety of TCR $\alpha$  chains, while the other required an exact TCR $\alpha$  partner [49]. Finally, some TCRs have been shown to possess considerable structural plasticity and are thus able to adjust and bind to a range of slightly different antigens [50,51], suggesting that they may be amenable to promiscuous pairing without loss of specificity.

Together with these earlier reports our data suggest a scenario in which the human thymus readily and repeatedly generates public TCR $\alpha$  chains associated with the recognition of islet antigens. These chains are not deleted in negative selection, but their final specificity and affinity is modulated by their TCR $\beta$  partner. Since TCR  $\alpha$  to  $\beta$  pairing is essentially random, we suggest that the actual risk posed by these autoantigen-associated TCR $\alpha$  chains and part of the individual susceptibility to T1D is determined by the TCR $\beta$  microrepertoire paired with them. The threshold for developing clinical autoimmunity is further modified by the genetic risk profile, early microbial environment, and the life-long history of encounters with infectious agents.

The existing data on public T cell clones from T1D patients, although few studies so far have addressed TCR $\alpha$ , is consistent with the scenario outlined above. A specific role for public TCR $\alpha$  clonotypes in T1D was suggested by studies identifying GAD65-specific CD4<sup>+</sup> TCR $\alpha$  clonotypes and IGRP-specific CD8<sup>+</sup> TCR $\alpha$  clonotypes present at a high frequency across different patients with T1D [52,53]. Another study identified circulating islet-reactive CD8<sup>+</sup> T cells with public TCR $\alpha$  chains in most individuals, but their homing to pancreas was mostly limited to T1D patients [54]. The development of more powerful techniques is required to be able to characterize the TCR $\beta$  repertoire paired with these public TCR $\alpha$  chains. However, it is interesting to speculate whether the public TCR $\alpha$  chains in T1D are recurrent enough to provide a target for intervention.

## Funding

University of Helsinki Research Foundation, Emil Aaltonen Foundation, The Finnish Medical Foundation, Biomedicum Helsinki Foundation, and MD PhD program of the University of Helsinki.

## Author contributions

NH: conceptualization, investigation, visualization and writing the original draft. SS formal analysis, software and visualization. JM investigation. TH investigation and resources. MK conceptualization and resources. ELI investigation. TK conceptualization. IPM resources. JS conceptualization, formal analysis, software. TPA conceptualization, writing the original draft. All authors contributed to the review and editing of the manuscript.

## Declaration of competing interest

The authors declare no potential competing financial interests.

## Acknowledgements

We thank Tamás Bazsinka for technical support. This study has been funded by the University of Helsinki Research Foundation, Emil Aaltonen Foundation, The Finnish Medical Foundation, Biomedicum Helsinki Foundation, and the MD PhD program of the University of Helsinki.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jaut.2021.102616>.

## References

- [1] M.M. Davis, P.J. Bjorkman, T-cell antigen receptor genes and T-cell recognition, *Nature* 334 (1988) 395–402, <https://doi.org/10.1038/334395a0>.
- [2] T.P. Arstila, A. Casrouge, V. Baron, J. Even, J. Kanellopoulos, P. Kourilsky, A direct estimate of the human alpha T cell receptor diversity, *Science* 286 (1999) 958–961. <https://www.ncbi.nlm.nih.gov/pubmed/10542151>.
- [3] H.S. Robins, S.K. Srivastava, P.V. Campregher, C.J. Turtle, J. Andriessen, S. R. Riddell, C.S. Carlson, E.H. Warren, Overlap and effective size of the human CD8 + T cell receptor repertoire, *Sci. Transl. Med.* 2 (2010) 47, <https://doi.org/10.1126/scitranslmed.3001442>, ra64.
- [4] Q. Qi, Y. Liu, Y. Cheng, J. Glanville, D. Zhang, J.-Y. Lee, R.A. Olshen, C.M. Weyand, S.D. Boyd, J.J. Goronzy, Diversity and clonal selection in the human T-cell repertoire, *Proc. Natl. Acad. Sci. U. S. A* 111 (2014) 13139–13144, <https://doi.org/10.1073/pnas.1409155111>.
- [5] R. Vanhanen, N. Heikkilä, K. Aggarwal, D. Hamm, H. Tarkkila, T. Pättilä, T. S. Jokiranta, J. Saramäki, T.P. Arstila, T cell receptor diversity in the human thymus, *Mol. Immunol.* 76 (2016) 116–122, <https://doi.org/10.1016/j.molimm.2016.07.002>.
- [6] M.G. Rudolph, R.L. Stanfield, I.A. Wilson, How TCRs bind MHCs, peptides, and coreceptors, *Annu. Rev. Immunol.* 24 (2006) 419–466, <https://doi.org/10.1146/annurev.immunol.23.021704.115658>.
- [7] V. Venturi, D.A. Price, D.C. Douek, M.P. Davenport, The molecular basis for public T-cell responses? *Nat. Rev. Immunol.* 8 (2008) 231–238, <https://doi.org/10.1038/nri2260>.
- [8] H. Li, C. Ye, G. Ji, J. Han, Determinants of public T cell responses, *Cell Res.* 22 (2012) 33–42, <https://doi.org/10.1038/cr.2012.1>.
- [9] V. Venturi, H.Y. Chin, T.E. Asher, K. Ladell, P. Scheinberg, E. Bornstein, D. van Bockel, A.D. Kelleher, D.C. Douek, D.A. Price, M.P. Davenport, TCR beta-chain sharing in human CD8<sup>+</sup> T cell responses to cytomegalovirus and EBV, *J. Immunol.* 181 (2008) 7853–7862. <http://www.jimmunol.org/content/181/11/7853.full.pdf>.
- [10] A. Madi, E. Shifrut, S. Reich-Zeliger, H. Gal, K. Best, W. Ndifon, B. Chain, I. R. Cohen, N. Friedman, T-cell receptor repertoires share a restricted set of public and abundant CDR3 sequences that are associated with self-related immunity, *Genome Res.* 24 (2014) 1603–1612, <https://doi.org/10.1101/gr.170753.113>.
- [11] Y. Zhao, P. Nguyen, P. Vogel, B. Li, L.L. Jones, T.L. Geiger, Autoimmune susceptibility imposed by public TCRbeta chains, *Sci. Rep.* 6 (2016) 37543, <https://doi.org/10.1038/srep37543>.
- [12] D. Benati, M. Galperin, O. Lambotte, S. Gras, A. Lim, M. Mukhopadhyay, A. Nouel, K.A. Campbell, B. Lemercier, M. Claireaux, S. Hendou, P. Lechat, P. de Truchis, F. Boufassa, J. Rossjohn, J.F. Delfraissy, F. Arenzana-Seisdedos, L.A. Chakrabarti, Public T cell receptors confer high-avidity CD4 responses to HIV controllers, *J. Clin. Invest.* 126 (2016) 2093–2108, <https://doi.org/10.1172/JCI83792>.
- [13] I.V. Zvyagin, M.V. Pogorelyy, M.E. Ivanova, E.A. Komech, M. Shugay, D.A. Bolotin, A.A. Shelenkov, A.A. Kurnosov, D.B. Staroverov, D.M. Chudakov, Y.B. Lebedev, I. Z. Mamedov, Distinctive properties of identical twins' TCR repertoires revealed by high-throughput sequencing, *Proc. Natl. Acad. Sci. U. S. A* 111 (2014) 5980–5985, <https://doi.org/10.1073/pnas.1319389111>.
- [14] N. Heikkilä, R. Vanhanen, D.A. Yohannes, I. Kleino, I.P. Mattila, J. Saramäki, T. P. Arstila, Human thymic T cell repertoire is imprinted with strong convergence to shared sequences, *Mol. Immunol.* 127 (2020) 112–123, <https://doi.org/10.1016/j.molimm.2020.09.003>.
- [15] N. Heikkilä, R. Vanhanen, D.A. Yohannes, P. Saavalainen, S. Meri, T.S. Jokiranta, H. Jarva, I.P. Mattila, D. Hamm, S. Sormunen, J. Saramäki, T.P. Arstila, Identifying the inheritable component of human thymic T cell repertoire generation in monozygous twins, *Eur. J. Immunol.* 50 (2020) 748–751, <https://doi.org/10.1002/eji.201948404>.
- [16] A. Parkkola, T. Härkönen, S.J. Ryhänen, J. Ilonen, M. Knip, Extended family history of autoimmune diseases and phenotype and genotype of children with newly diagnosed type 1 diabetes, *Eur. J. Endocrinol.* 169 (2013) 171–178, <https://doi.org/10.1530/EJE-13-0089>.
- [17] N. Tickotsky, T. Sagiv, J. Prilusky, E. Shifrut, N. Friedman, McPAS-TCR: a manually curated catalogue of pathology-associated T cell receptor sequences, *Bioinformatics* 33 (2017) 2924–2929, <https://doi.org/10.1093/bioinformatics/btx286>.
- [18] D.V. Bagaev, R.M.A. Vroomans, J. Samir, U. Stervbo, C. Rius, G. Dolton, A. Greenshields-Watson, M. Attaf, E.S. Egorov, I.V. Zvyagin, N. Babel, D.K. Cole, A. J. Godkin, A.K. Sewell, C. Kesmir, D.M. Chudakov, F. Luciani, M. Shugay, VDJdb in 2019: database extension, new analysis infrastructure and a T-cell receptor motif compendium, *Nucleic Acids Res.* 48 (2020) D1057–D1062, <https://doi.org/10.1093/nar/gkz874>.
- [19] M. So, C.M. Elso, E. Tresoldi, M. Pakusch, V. Pathiraja, J.M. Wentworth, L. C. Harrison, B. Krishnamurthy, H.E. Thomas, C. Rodda, F.J. Cameron, J. McMahon, T.W.H. Kay, S.I. Mannering, Proinsulin C-peptide is an autoantigen in people with type 1 diabetes, *Proc. Natl. Acad. Sci. U. S. A* 115 (2018) 10732–10737, <https://doi.org/10.1073/pnas.1809208115>.
- [20] E.-L. Ihantola, H. Ilmonen, A. Kaillaanmäki, M. Rytönen-Nissinen, A. Azam, B. Maillère, C.S. Lindestam Arlehamn, A. Sette, K. Motwani, H.R. Seay, T. M. Brusko, M. Knip, R. Veijola, J. Toppari, J. Ilonen, T. Kinnunen, Characterization of proinsulin T cell epitopes restricted by type 1 diabetes-associated HLA class II molecules, *J. Immunol.* (2020), <https://doi.org/10.4049/jimmunol.1901079>.
- [21] K. Cerosaletti, F. Barahmand-Pour-Whitman, J. Yang, H.A. DeBerg, M.J. Dufort, S. A. Murray, E. Israelsson, C. Speake, V.H. Gersuk, J.A. Eddy, H. Reijonen, C. J. Greenbaum, W.W. Kwok, E. Wambre, M. Prlic, R. Gottardo, G.T. Nepom, P. S. Linsley, Single-cell RNA sequencing reveals expanded clones of islet antigen-



- reactive CD4+ T cells in peripheral blood of subjects with type 1 diabetes, *J. Immunol.* 199 (2017) 323–335, <https://doi.org/10.4049/jimmunol.1700172>.
- [22] D.X. Beringer, F.S. Kleijwegt, F. Wiede, A.R. van der Slik, K.L. Loh, J. Petersen, N. L. Dudek, G. Duinkerken, S. Laban, A. Joosten, J.P. Vivian, Z. Chen, A.P. Uldrich, D.I. Godfrey, J. McCluskey, D.A. Price, K.J. Radford, A.W. Purcell, T. Nikolic, H. H. Reid, T. Tiganis, B.O. Roep, J. Rossjohn, T cell receptor reversed polarity recognition of a self-antigen major histocompatibility complex, *Nat. Immunol.* 16 (2015) 1153–1161, <https://doi.org/10.1038/ni.3271>.
- [23] O.V. Bolkhovskaya, D.Y. Zorin, M.V. Ivanchenko, Assessing T cell clonal size distribution: a non-parametric approach, *PLoS One* 9 (2014), e108658, <https://doi.org/10.1371/journal.pone.0108658>.
- [24] J.A. Noble, A.M. Valdes, Genetics of the HLA region in the prediction of type 1 diabetes, *Curr. Diabetes Rep.* 11 (2011) 533–542, <https://doi.org/10.1007/s11892-011-0223-x>.
- [25] L. Klein, B. Kyewski, P.M. Allen, K.A. Hogquist, Positive and negative selection of the T cell repertoire: what thymocytes see (and don't see), *Nat. Rev. Immunol.* 14 (2014) 377–391, <https://doi.org/10.1038/nri3667>.
- [26] A. Pugliese, M. Zeller, A. Fernandez, L.J. Zalcberg, R.J. Bartlett, C. Ricordi, M. Pietropaolo, G.S. Eisenbarth, S.T. Bennett, D.D. Patel, The insulin gene is transcribed in the human thymus and transcription levels correlate with allelic variation at the INS VNTR-IDDM2 susceptibility locus for type 1 diabetes, *Nat. Genet.* 15 (1997) 293–297, <https://doi.org/10.1038/ng0397-293>.
- [27] M. Sospedra, X. Ferrer-Francesch, O. Dominguez, M. Juan, M. Foz-Sala, R. Pujol-Borrell, Transcription of a broad range of self-antigens in human thymus suggests a role for central mechanisms in tolerance toward peripheral antigens, *J. Immunol.* 161 (1998) 5918–5929, <https://www.ncbi.nlm.nih.gov/pubmed/9834072>.
- [28] J. Gotter, B. Brors, M. Hergenhanh, B. Kyewski, Medullary epithelial cells of the human thymus express a highly diverse selection of tissue-specific genes colocalized in chromosomal clusters, *J. Exp. Med.* 199 (2004) 155–166, <https://doi.org/10.1084/jem.20031677>.
- [29] L. Amaya-Urbe, M. Rojas, G. Azizi, J.-M. Anaya, M.E. Gershwin, Primary immunodeficiency and autoimmunity: a comprehensive review, *J. Autoimmun.* 99 (2019) 52–72, <https://doi.org/10.1016/j.jaut.2019.01.011>.
- [30] L. Guilherme, K. Faé, S.E. Oshiro, J. Kalil, Molecular pathogenesis of rheumatic fever and rheumatic heart disease, *Expert Rev. Mol. Med.* 7 (2005) 1–15, <https://doi.org/10.1017/S146239940501015X>.
- [31] D.L. Petersel, L.H. Sigal, Reactive arthritis, *Infect. Dis. Clin.* 19 (2005) 863–883, <https://doi.org/10.1016/j.idc.2005.07.001>.
- [32] M.A. Brown, T. Kenna, B.P. Wordsworth, Genetics of ankylosing spondylitis—insights into pathogenesis, *Nat. Rev. Rheumatol.* 12 (2016) 81–91, <https://doi.org/10.1038/nrrheum.2015.133>.
- [33] A. Bar-Or, M.P. Pender, R. Khanna, L. Steinman, H.-P. Hartung, T. Maniari, E. Croze, B.T. Aftab, G. Giovannoni, M.A. Joshi, Epstein-barr virus in multiple sclerosis: theory and emerging immunotherapies, *Trends Mol. Med.* 26 (2020) 296–310, <https://doi.org/10.1016/j.molmed.2019.11.003>.
- [34] J.-F. Bach, Infections and autoimmune diseases, *J. Autoimmun.* 25 (Suppl) (2005) 74–80, <https://doi.org/10.1016/j.jaut.2005.09.024>.
- [35] D. Lebrun, M. Hentzien, L. Cuzin, D. Rey, V. Joly, L. Cotte, C. Allavena, P. Dellamonica, A. Servettaz, F. Bani-Sadr, The Dat AIDS study group, Epidemiology of autoimmune and inflammatory diseases in a French nationwide HIV cohort, *AIDS* 31 (2017) 2159–2166, <https://doi.org/10.1097/QAD.0000000000001603>.
- [36] J.J. Damba, M. Laskine, Y. Jin, L. Sinyavskaya, M. Durand, Incidence of autoimmune diseases in people living with HIV compared to a matched population: a cohort study, *Clin. Rheumatol.* (2020), <https://doi.org/10.1007/s10067-020-05500-x>.
- [37] Y. Rodríguez, L. Novelli, M. Rojas, M. De Santis, Y. Acosta-Ampudia, D. M. Monsalve, C. Ramírez-Santana, A. Costanzo, W.M. Ridgway, A.A. Ansari, M. E. Gershwin, C. Selmi, J.-M. Anaya, Autoinflammatory and autoimmune conditions at the crossroad of COVID-19, *J. Autoimmun.* 114 (2020) 102506, <https://doi.org/10.1016/j.jaut.2020.102506>.
- [38] L. Novelli, F. Motta, M. De Santis, A.A. Ansari, M.E. Gershwin, C. Selmi, The JANUS of chronic inflammatory and autoimmune diseases onset during COVID-19 - a systematic review of the literature, *J. Autoimmun.* 117 (2020) 102592, <https://doi.org/10.1016/j.jaut.2020.102592>.
- [39] A.B. Sioofy-Khojine, J. Lehtonen, N. Nurminen, O.H. Laitinen, S. Oikarinen, H. Huhtala, O. Pakkanen, T. Ruokoranta, M.M. Hankaniemi, J. Toppari, M. Vaha-Makila, J. Ilonen, R. Veijola, M. Knip, H. Hyoty, Cocksackievirus B1 infections are associated with the initiation of insulin-driven autoimmunity that progresses to type 1 diabetes, *Diabetologia* 61 (2018) 1193–1202, <https://doi.org/10.1007/s00125-018-4561-y>.
- [40] O.H. Laitinen, H. Honkanen, O. Pakkanen, S. Oikarinen, M.M. Hankaniemi, H. Huhtala, T. Ruokoranta, V. Lecouturier, P. André, R. Harju, S.M. Virtanen, J. Lehtonen, J.W. Almond, T. Simell, O. Simell, J. Ilonen, R. Veijola, M. Knip, H. Hyoty, Cocksackievirus B1 is associated with induction of  $\beta$ -cell autoimmunity that portends type 1 diabetes, *Diabetes* 63 (2014) 446–455, <https://doi.org/10.2337/db13-0619>.
- [41] J. Ilonen, J. Lempainen, A. Hammis, A.P. Laine, T. Harkonen, J. Toppari, R. Veijola, M. Knip, Primary islet autoantibody at initial seroconversion and autoantibodies at diagnosis of type 1 diabetes as markers of disease heterogeneity, *Pediatr. Diabetes* 19 (2018) 284–292, <https://doi.org/10.1111/pedi.12545>.
- [42] M.K. Smatti, F.S. Cyprian, G.K. Nasrallah, A.A. Al Thani, R.O. Almishal, H. M. Yassine, Viruses and autoimmunity: a review on the potential interaction and molecular mechanisms, *Viruses* 11 (2019), <https://doi.org/10.3390/v11080762>.
- [43] L. Cleenewerk, J. Garssen, A. Hogenkamp, Clinical use of schistosoma mansoni antigens as novel immunotherapies for autoimmune disorders, *Front. Immunol.* 11 (2020) 1821, <https://doi.org/10.3389/fimmu.2020.01821>.
- [44] P. Bradley, P.G. Thomas, Using T cell receptor repertoires to understand the principles of adaptive immune recognition, *Annu. Rev. Immunol.* 37 (2019) 547–570, <https://doi.org/10.1146/annurev-immunol-042718-041757>.
- [45] F. Rubelt, C.R. Bolen, H.M. McGuire, J.A. Vander Heiden, D. Gadala-Maria, M. Levin, G.M. Euskirchen, M.R. Mamedov, G.E. Swan, C.L. Dekker, L.G. Cowell, S. H. Kleinstein, M.M. Davis, Individual heritable differences result in unique cell lymphocyte receptor repertoires of naive and antigen-experienced cells, *Nat. Commun.* 7 (2016) 11112, <https://doi.org/10.1038/ncomms11112>.
- [46] M. Nakayama, N. Abiru, H. Moriyama, N. Babaya, E. Liu, D. Miao, L. Yu, D. R. Wegmann, J.C. Hutton, J.F. Elliott, G.S. Eisenbarth, Prime role for an insulin epitope in the development of type 1 diabetes in NOD mice, *Nature* 435 (2005) 220–223, <https://doi.org/10.1038/nature03523>.
- [47] H. Tanno, T.M. Gould, J.R. McDaniel, W. Cao, Y. Tanno, R.E. Durrett, D. Park, S. J. Cate, W.H. Hildebrand, C.L. Dekker, L. Tian, C.M. Weyand, G. Georgiou, J. J. Goronzy, Determinants governing T cell receptor  $\alpha/\beta$ -chain pairing in repertoire formation of identical twins, *Proc. Natl. Acad. Sci. U. S. A.* 117 (2020) 532–540, <https://doi.org/10.1073/pnas.1915008117>.
- [48] M. Nakatsugawa, Y. Yamashita, T. Ochi, S. Tanaka, K. Chamoto, T. Guo, M. O. Butler, N. Hirano, Specific roles of each TCR hemichain in generating functional chain-centric TCR, *J. Immunol.* 194 (2015) 3487–3500, <https://doi.org/10.4049/jimmunol.1401717>.
- [49] N.A. Bowerman, M.T. Falta, D.G. Mack, F. Wehrmann, F. Crawford, M.M. Mroz, L. A. Maier, J.W. Kappler, A.P. Fontenot, Identification of multiple public TCR repertoires in chronic beryllium disease, *J. Immunol.* 192 (2014) 4571–4580, <https://doi.org/10.4049/jimmunol.1400007>.
- [50] K.M. Armstrong, K.H. Piepenbrink, B.M. Baker, Conformational changes and flexibility in T-cell receptor recognition of peptide-MHC complexes, *Biochem. J.* 415 (2008) 183–196, <https://doi.org/10.1042/BJ20080850>.
- [51] A.K. Sewell, Why must T cells be cross-reactive? *Nat. Rev. Immunol.* 12 (2012) 669–677, <https://doi.org/10.1038/nri3279>.
- [52] A. Eugster, A. Lindner, M. Catani, A.K. Heninger, A. Dahl, S. Klemroth, D. Kuhn, S. Dietz, M. Bickle, A.G. Ziegler, E. Bonifacio, High diversity in the TCR repertoire of GAD65 autoantigen-specific human CD4+ T cells, *J. Immunol.* 194 (2015) 2531–2538, <https://doi.org/10.4049/jimmunol.1403031>.
- [53] Y.F. Fuchs, A. Eugster, S. Dietz, C. Sebelesky, D. Kuhn, C. Wilhelm, A. Lindner, A. Gavrigan, J. Knoop, A. Dahl, A.G. Ziegler, E. Bonifacio, CD8(+) T cells specific for the islet autoantigen IGRP are restricted in their T cell receptor chain usage, *Sci. Rep.* 7 (2017) 44661, <https://doi.org/10.1038/srep44661>.
- [54] S. Culina, A.I. Lalanne, G. Afonso, K. Cerosaletti, S. Pinto, G. Sebastiani, K. Kuranda, L. Nigi, A. Eugster, T. Østerbye, A. Maugein, J.E. McLaren, K. Ladell, E. Larger, J.-P. Beressi, A. Lissina, V. Appay, H.W. Davidson, S. Buus, D.A. Price, M. Kuhn, E. Bonifacio, M. Battaglia, S. Caillat-Zucman, F. Dotta, R. Scharfmann, B. Kyewski, R. Mallone, ImMaDiab Study Group, Islet-reactive CD8+ T cell frequencies in the pancreas, but not in blood, distinguish type 1 diabetic patients from healthy donors, *Sci Immunol* 3 (2018), <https://doi.org/10.1126/sciimmunol.aao4013>.