JCI insight

Temporal development of T cell receptor repertoires during childhood in health and disease

Angela M. Mitchell,[1] Erin E. Baschal,[1] Kristen A. McDaniel,[1] Kimber M. Simmons,[1,2] Laura Pyle,[1,3] Kathleen Waugh,[1] Andrea K. Steck,[1,2] Liping Yu,[1] Peter A. Gottlieb,[1,2,4] Marian J. Rewers,[1,2,4] Maki Nakayama,[1,2,5] and Aaron W. Michels[1,2,4,5]

[1]Barbara Davis Center for Diabetes and [2]Department of Pediatrics, University of Colorado School of Medicine, Aurora, Colorado, USA. [3]Department of Biostatistics and Informatics, University of Colorado School of Public Health, Aurora, Colorado, USA. [4]Department of Medicine and [5]Department of Immunology, University of Colorado School of Medicine, Aurora, Colorado, USA.

T cell receptor (TCR) sequences are exceptionally diverse and can now be comprehensively measured with next-generation sequencing technologies. However, a thorough investigation of longitudinal TCR repertoires throughout childhood in health and during development of a common childhood disease, type 1 diabetes (T1D), has not been undertaken. Here, we deep sequenced the TCR-β chain repertoires from longitudinal peripheral blood DNA samples at 4 time points beginning early in life (median age of 1.4 years) from children who progressed to T1D (*n* = 29) and age/sex-matched islet autoantibody-negative controls (*n* = 25). From 53 million TCR-β sequences, we show that the repertoire is extraordinarily diverse early in life and narrows with age independently of disease. We demonstrate the ability to identify specific TCR sequences, including those known to recognize influenza A and, separately, those specific for insulin and its precursor, preproinsulin. Insulin-reactive TCR-β sequences were more common and frequent in number as the disease progressed in those who developed T1D compared with genetically at risk nondiabetic children, and this was not the case for influenza-reactive sequences. As an independent validation, we sequenced and analyzed TCR-β repertoires from a cohort of new-onset T1D patients (*n* = 143), identifying the same preproinsulin-reactive TCRs. These results demonstrate an enrichment of preproinsulin-reactive TCR sequences during the progression to T1D, highlighting the importance of using disease-relevant TCR sequences as powerful biomarkers in autoimmune disorders.

Introduction

The T cell receptor (TCR) on the cell surface of T lymphocytes plays a central role in immune responses by the recognition of peptides in the context of major histocompatibility complexes (1, 2). Rearrangement of TCR α and β genes results in a vast and diverse repertoire of T cells within a given individual, allowing for recognition of millions of unique antigens. Advances in high-throughput sequencing technologies now enable deep sequencing of TCR-β chains from genomic DNA (3–5). Previous studies have analyzed TCR repertoires throughout the lifespan in a cross-sectional manner (6, 7), but none have longitudinally assessed the temporal changes in the repertoire beginning early in life and throughout childhood. The development of the TCR repertoire during childhood is beneficial for responding to foreign pathogens; however, recognition of self-peptides by T cells during this time is thought to contribute to development of autoimmune diseases such as type 1 diabetes (T1D), the immune-mediated form of diabetes.

T1D is a prototypical organ-specific autoimmune disease that develops in distinct stages marked by the presence of autoantibodies in peripheral blood targeting insulin and other pancreatic islet proteins such as glutamic acid decarboxylase (GADA), tyrosine phosphatase–related islet antigen-2 (IA-2A), and zinc transporter 8 (ZnT8A) (8, 9). Following islet autoantibody seroconversion, there is a period of time — generally years — prior to the development of hyperglycemia and clinical symptoms, and this requires lifelong treatment with exogenous insulin administration (10). Additionally, CD4 and CD8 T cells contribute to disease pathogenesis, as both cell types are present in the immune infiltrates of inflamed pancreatic islets

from T1D organ donors (11–14). The disease course provides defined periods to measure immune responses prior to the development of islet autoantibodies, after seroconversion, and then at clinical T1D onset.

In this study, we longitudinally deep sequenced the TCR-β chain repertoire from peripheral blood beginning early in life from children genetically at risk for T1D and throughout childhood into adolescence. Cases developed islet autoantibodies and then went on to clinical disease, while age- and sex-matched nondiabetic controls did not develop autoantibodies or T1D. We show that the TCR-β chain repertoire is extraordinarily diverse early in life and contracts over time. From our large TCR-β data set, we then evaluated the hypothesis that a subset of self-antigen–specific TCR-β sequences obtained from the target tissue, inflamed pancreatic islets of human T1D organ donors, would be present in blood and accumulate throughout the stages of diabetes development. We show that there is a specific enrichment of TCR-β sequences known to respond to insulin and its precursor, preproinsulin (PPI), during the progression to clinical diabetes. The ability to sequence TCRs from peripheral blood DNA has significant relevance for monitoring disease progression in T1D and other autoimmune disorders with distinct stages.

Results

Diversity of the TCR-β chain repertoire decreases throughout childhood. To gain a better understanding of the TCR repertoire early in life and during the stages of T1D development, we deep sequenced TCR-β chain repertoires from participants in the Diabetes Autoimmunity Study in the Young (DAISY), an established prospective birth cohort study following genetically at-risk children for the development of T1D-associated autoantibodies (islet autoantibodies) and clinical diabetes (15). Individuals in our nested case-control study were age and sex matched, and the vast majority of cases and controls carried 1 or both of the T1D human leukocyte antigen (HLA) class II risk alleles DQ8 and/or DQ2, with no statistical differences between the groups (Table 1 and Supplemental Data Set 1; supplemental material available online with this article; https://doi.org/10.1172/jci.insight.161885DS1). Cases ($n = 29$) progressed to clinical T1D and were sampled at 4 time points: early in life (mean age 2.1 years), prior to islet autoantibody seroconversion (5.5 years), after seroconversion (6.7 years), and immediately prior to T1D onset (12.2 years). Controls ($n = 25$) were sampled at comparable age-matched time points but did not develop islet autoantibodies or T1D (Figure 1A and Table 1). Importantly, controls and cases had similar numbers of TCR-β sequences at each time point (Supplemental Figure 1, A and B), with more than 50 million total sequences obtained. When analyzing the repertoires, the TCR-β sequences became more clonal over time (Figure 1B), and the Simpson productive clonality metric directly correlated with age in both cases and controls (Figure 1C). There was no difference in Simpson productive clonality between cases and controls at any single time point (Supplemental Figure 2).

Taken together, these results indicate that, early in life, TCR-β repertoires are diverse and become more narrowed with age, likely due to childhood infections and vaccinations resulting in clonal expansion of the responding T cells.

Distinct TCR V gene usage early in life identifies a subset of T1D cases. Next, we analyzed the repertoires for differences between controls and cases, beginning with overall V gene usage by the TCR-β chains. Remarkably, V gene usage in the total repertoire was nearly identical at the 4 time points within and between each group (Figure 2A). At the individual level, V gene usage was heterogeneous but remained consistent within any given individual over time (Supplemental Figure 3, A–D). Controls and cases were well matched for HLA class I and II alleles (Supplemental Figure 4, A and B), and this may contribute to the remarkable similarities in V gene usage between the groups. Although V gene usage did not differ between the groups over time, a subset of cases ($n = 11$) could be distinguished from controls ($n = 25$) and remaining cases ($n = 18$) by principal component analyses (PCA) based on percentages of V genes used in the total repertoire (Figure 2B). The variance accounting for the separation of these cases was driven by 5 TCR-β V genes: 28-01 (>88% of PC1), 04-03 (>28% of PC2), 07-02 (>22% of PC2), 27-01 (>17% of PC3), and 30-01 (>22% of PC3 and >23% of PC4; Figure 2B). As such, there were statistically significant increases in the use of Vβ chains 04-03 and 07-02, and there was a decrease in 28-01 that could be identified early in life and persisted over time in a subset of genetically at-risk children who developed T1D (Figure 2C).

Clinically, these cases who had the distinct V gene usage patterns were not different in sex, race, age at islet autoantibody seroconversion, or age at clinical T1D diagnosis compared with the remaining cases (Supplemental Figure 5, A and B). Furthermore, the subset of cases was also well matched for HLA class I and II alleles when compared with the remaining cases (Supplemental Figure 6, A and B), indicating

Table 1. Demographic and immunologic characteristics of DAISY study participants

	Control (n = 25)	Case (n = 29)	P value[A]
Age, mean (SD) in years			
Time point 1	1.9 (1.2)	2.1 (1.4)	0.63
Time point 2	5.2 (2.3)	5.5 (2.4)	0.69
Time point 3	6.6 (2.4)	6.7 (3.0)	0.93
Time point 4	11.7 (3.8)	12.2 (4.1)	0.65
Diabetes diagnosis	NA	12.8 (4.3)	
Sex			1.0
Female % (number)	36% (n = 9)	38% (n = 11)	
Race/ethnicity			0.69
Non-Hispanic White	84% (n = 21)	90% (n = 26)	
Hispanic White	16% (n = 4)	10% (n = 3)	
T1D autoantibodies present[B], % (number)			
Time point 1	0% (n = 0)	7% (n = 2)	0.49
Time point 2	0% (n = 0)	62% (n = 18)	<0.001
Time point 3	0% (n = 0)	93% (n = 27)	<0.001
Time point 4	0% (n = 0)	86% (n = 25)	<0.001
HLA-DQ fenotype, % (number)			
DQ8/X[C]	28% (n = 7)	41% (n = 12)	0.40[C]
DQ2/X[D]	20% (n = 5)	24% (n = 7)	0.75
DQ2/8	36% (n = 9)	28% (n = 8)	0.57
DQX/X	16% (n = 4)	7% (n = 2)	0.40

[A]P values for continuous variables use t test; categorical variables use Fisher's exact test. [B]One case developed clinical type 1 diabetes without autoantibodies. [C]DQ8 consists of the alleles DQA1*03:01-DQB1*03:02 or DQA1*03:03-DQB1*03:02. [D]DQ2 consists of the alleles DQA1*05:01-DQB1*02:01 or DQA1*02:01-DQB1*02:02.

that HLA allele differences were unlikely to segregate the subset of cases. However, there were notable differences in their islet autoantibody profiles. Individuals in the subset of cases were more likely to develop autoantibodies directed against GADA compared with the remaining cases (100% versus 61%, $P = 0.026$), trended toward being more likely to develop tyrosine phosphatase–related IA-2A (82% versus 50%, $P = 0.126$), and were more likely to develop both GADA and IA-2A (82% versus 28%, $P = 0.008$; Figure 2D and Supplemental Figure 7). These differences were not observed for development of insulin autoantibodies (IAA) (55% versus 67%, $P = 0.697$) or ZnT8 autoantibodies (64% versus 67%, $P = 1.0$, Figure 2D).

Overall, there was not a single dominant TCR-β V gene that differentiated cases and controls over time; however, PCA of TCR-β V genes identified a subset of cases early in life that eventually developed distinct islet autoantibody profiles and clinical diabetes.

Influenza-reactive TCR sequences are present during childhood. Influenza A virus is a common pathogen with initial exposure during childhood, and we aimed to use our database of TCR-β sequences to detect antigen-specific sequences for influenza. Using 6 influenza-reactive αβTCRs, 3 CD4 sequences, and 3 CD8 sequences, with known specificities to hemagglutinin or the matrix protein of the virus and restricted to diabetes-risk HLA molecules (16–18), we searched all 4 time points for the presence of the influenza TCR-β chains (identical V, J, and CDR3 amino acid sequences). Three of these sequences were identified comprising hundreds of templates in numerous individuals (Supplemental Table 1). A similar number of cases and controls had influenza sequences at each time point, and TCR-β template numbers in both groups were comparable (Figure 3A). As expected, influenza-reactive TCR-β chain presence and template number correlated with increasing age in both cases and controls with a trend toward statistical significance (Figure 3B). Next, we examined whether a given influenza TCR-β with multiple identical amino acid templates came from different DNA clonotypes. There were 18 samples that had more than 3 identical amino acid sequences for a single influenza TCR-β, and in all of these samples there were multiple DNA clonotypes that converged to a single amino acid TCR-β sequence. This indicates that there was in vivo antigen presentation and expansion of these T cell clones, as opposed to multiple amplifications of a single nucleotide sequence during the polymerase chain reaction used for sequencing.
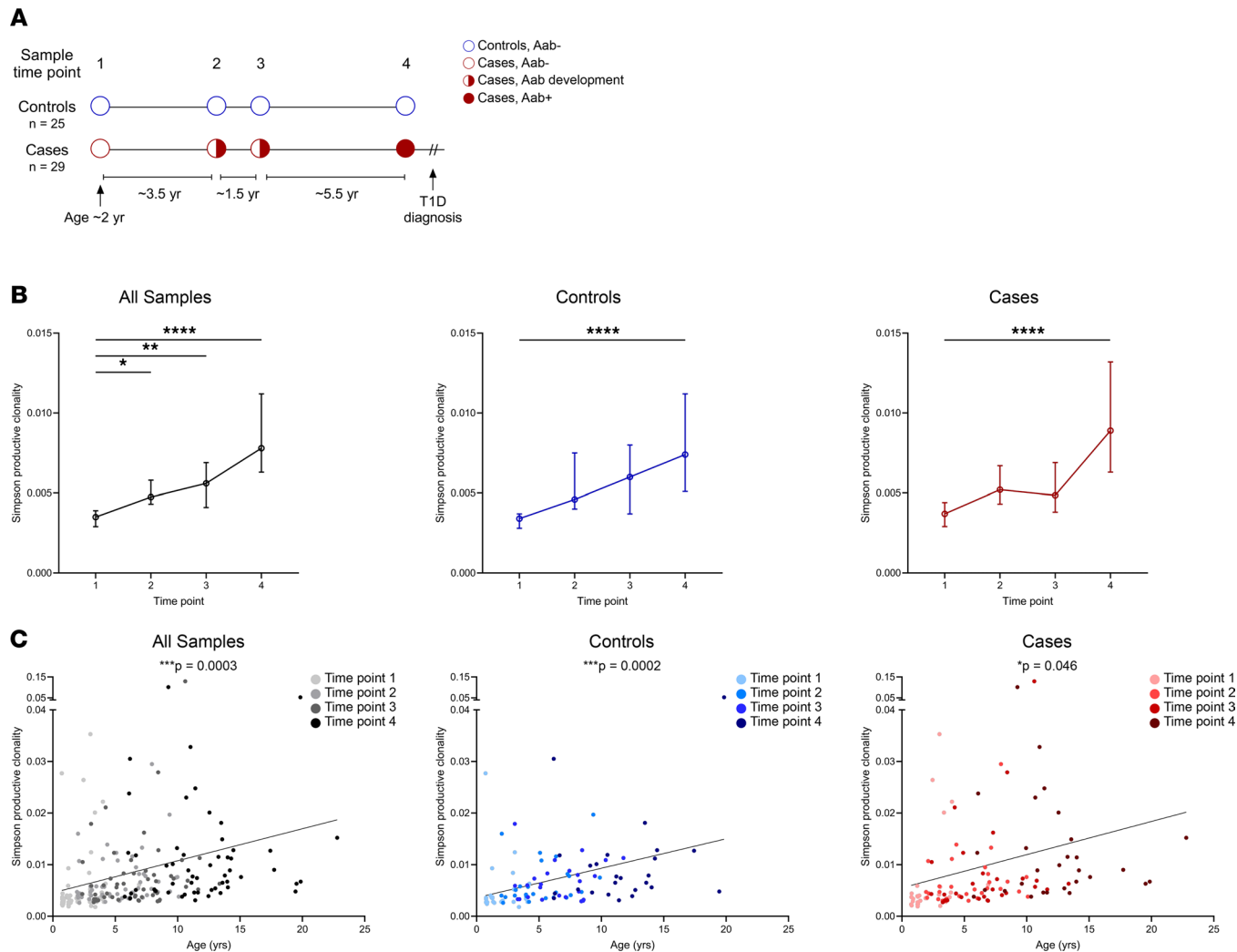
Figure 1. The TCR-β chain repertoire is diverse early in life and contracts with age. (**A**) Study design showing the time points and mean ages at which TCR-β chain sequencing was performed for nondiabetic controls (blue, *n* = 25) and cases (red, *n* = 29) who went on to develop type 1 diabetes, with islet autoantibody-negative visits as open circles, seroconversion as half-filled circles, and autoantibody-positive visits as filled-in circles. (**B**) Plots showing Simpson productive clonality at each time point for all samples (black, left), controls (blue, middle), and cases (red, right). Depicted are medians with 95% CI. (**C**) Scatterplots of Simpson productive clonality relative to age (years) for all samples (black, left), controls (blue, middle), and cases (red, right), with darker colors indicating later time points. *P* values were calculated using mixed-effects models to account for multiple measurements and comparisons for clonality at time point 1 versus all other time points; linear regression was used to compare clonality versus age. *$P < 0.05$, **$P < 0.01$, ****$P < 0.0001$.

In summary, we are able to search our TCR database and identify β chain sequences used by influenza-specific T cells, thus providing a proof of concept to detect antigen-specific TCR-β sequences.

Pancreatic islet–derived CD4 TCRs responding to insulin are more frequent in those who develop diabetes. With the ability to detect and quantitate viral-specific TCR-β chains, we next aimed to identify self-antigen–specific TCR sequences during the course of T1D development. Insulin is a major self-antigen in T1D, and its precursor, PPI, has been identified as a target for T cells within the inflamed pancreatic islets of T1D organ donors. To date, 44 PPI-reactive αβTCRs, 14 CD4, and 30 CD8, have been identified from the residual islets of multiple T1D donors (12, 14, 19–22). To determine whether these TCR-β chains were differentially expressed in cases and controls, we searched all 4 time points for the presence of the PPI TCR-β chains (identical V, J, and CDR3). A total of 15 PPI TCR-β chains were found throughout the time points in both groups (6 CD4 and 9 CD8 sequences; Figure 4A and Table 2). Notably, a larger proportion of cases had CD4 PPI TCR-β chain sequences compared with controls (72% versus 36%, $P = 0.0126$; Figure 4A), and the number of these templates was much higher in cases (180 versus 14 for cases and controls respectively, $P < 0.001$). This was not the case for CD8 PPI TCR-β sequences, which were present in similar proportions between cases and controls (52% versus 64%, $P = 0.417$; Figure 4A) and template numbers (29 versus 25 for cases and controls, respectively).
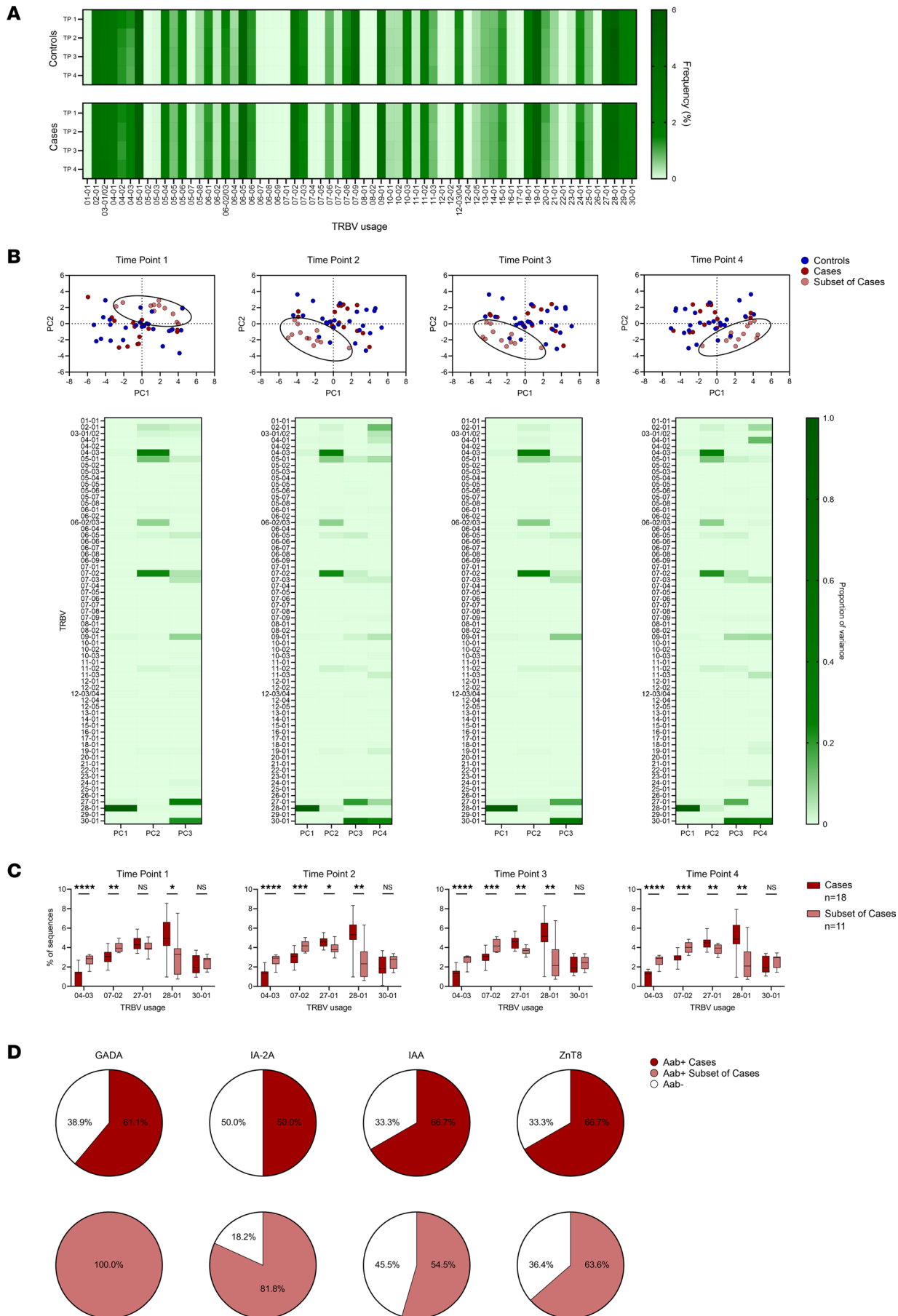
Figure 2. TCR-β V gene usage remains consistent throughout childhood. (**A**) Heatmaps of TCR-β V gene usage at the 4 time points in controls (top) and cases (bottom), with darker green indicating a higher frequency of a given gene. (**B**) Plots of principal component analyses depicting Vβ gene usage by individual at the 4 time points for controls (blue, $n = 29$), cases (dark red $n = 18$), and a subset of cases (light red, $n = 11$). Ellipses denote the same subset of cases throughout the 4 time points. Below each PCA plot is a heatmap quantifying the principal components (Vβ genes) contributing to the variance of each plot, with darker green indicating a higher proportion of the variance. (**C**) Box-and-whisker plots displaying Vβ gene usage accounting for the highest proportion of variance by cases (dark red, $n = 18$) and the subset of cases (light red, $n = 11$). The black center line denotes the median value (50th percentile), while the black box contains the 25th to 75th percentiles of the data set. The black whiskers mark the 10th and 90th percentiles. (**D**) Pie graphs showing the percentage of cases (dark red, top, $n = 18$) and the subset of cases (light red, bottom, $n = 11$) who developed each of the 4 islet autoantibodies tested: glutamic acid decarboxylase autoantibodies (GADA), tyrosine phosphatase–related islet antigen-2 autoantibodies (IA-2A), insulin autoantibodies (IAA), and zinc transporter 8 autoantibodies (ZnT8). Percentages of individuals in each group who were negative for each autoantibody are indicated in white. Time points in cases: 1, early in life; 2, before islet autoantibody positivity; 3, after islet autoantibody positivity; and 4, visit prior to clinical T1D diagnosis. Controls were age matched to cases at each time point. *P* values were calculated using Mann-Whitney *U* tests for Vβ gene usage in cases compared with the subset of cases at each time point. $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, $****P < 0.0001$.

Since islet autoantibodies precede the development of clinical T1D (23), we investigated the timing of the appearance of pancreatic PPI TCR-β sequences in the peripheral blood in relation to islet autoantibody seroconversion. CD4 and CD8 PPI TCR-β sequences were present before or at seroconversion in a few individuals (Figure 4B); however, CD4 PPI TCR-β templates increased considerably after the development of islet autoantibodies, especially when compared with controls at similar time points (Figure 4B). Since CD4 T cells undergo cognate interactions with B cells that can lead to antibody production, we investigated the presence and timing of CD4 PPI TCR-β templates in cases with and without IAA. Cases with IAA ($n = 18$) expressed more CD4 PPI TCR-β templates compared with IAA⁻ cases ($n = 11$), 167 versus 13 ($P < 0.001$, Figure 4C), respectively. Furthermore, the CD4 PPI TCR-β templates primarily accumulated after seroconversion to any islet autoantibody (Figure 4C). Four of the cases with IAA had 8 or more identical amino acid sequence templates for a single CD4 PPI TCR-β, and we examined whether different DNA clonotypes gave rise to these templates. Case 22 had 11 templates of a single CD4 PPI TCR-β that originated from 3 different DNA clonotypes (Supplemental Data Set 1), indicating in vivo presentation and expansion of this T cell clone. In contrast to the CD4 PPI TCR-β sequences, the frequency and accumulation of CD8 PPI TCR-β chains was not as robust as that for CD4 in cases with or without IAA (Figure 4C).

Taken together, these results demonstrate that cases who develop T1D more frequently expressed pancreatic islet–derived CD4 PPI TCR-β sequences compared with controls and had higher numbers of these TCR-β chains. Additionally, IAA⁺ cases expressed significantly more CD4 PPI TCR-β templates after autoantibody seroconversion compared with IAA negative cases, indicating a potential role for CD4 PPI TCRs in the affinity-matured antibody responses to insulin and B cell antigen presentation of PPI to these CD4 T cells.

Clusters of PPI-reactive TCR-β chains clonally expand prior to diabetes diagnosis. Although pancreatic PPI TCR-β chains were found in cases and controls, the overall numbers were much lower than that for influenza A–reactive TCR-β chains, despite having identified 15 shared PPI TCR-β chain sequences compared with only 3 for influenza. Therefore, we sought to identify other TCR-β chains that may recognize similar or nearly identical antigens as the identified PPI-reactive TCRs. Within αβTCRs, each chain contains a unique complementarity determining region 3 (CDR3), which encodes for the portion of the TCR that is most directly in contact with the antigen. The clustering program Grouping of Lymphocyte Interactions by Paratope Hotspots (GLIPH) analyzes TCR sequences and groups TCRs based upon a number of factors including HLA alleles, V gene usage, CDR3 sequence, and CDR3 length (24). Thus, the algorithm generates clusters of TCRs that may recognize shared or similar antigens to a parent TCR sequence. We clustered TCR-β sequences within an individual at each of the 4 time points using GLIPH2 (Figure 5A) (25). Tens of thousands of TCR-β clusters were identified in a single individual at 1 time point (Figure 5B), and we narrowed our focus to those clusters that included a known PPI TCR-β sequence. The most frequent PPI TCR-β found in this study, termed CD4 TCR #1, recognizes the well-characterized insulin B chain amino acids 9-23 peptide (B:9-23, Table 2) (26–30). After clustering all of the samples, 4 cluster motifs were identified which contained CD4 TCR #1 and additional, nearly identical TCR-β sequences (Figure 5C). From these TCR-β sequences, a single CDR3 motif was determined for each of the 4 clusters (i.e., sequence logo) (Figure 5C). Although the 4 clusters consisted of TCR-βs with nearly identical CDR3 sequences, each CDR3 motif contained 1 position, indicated by an asterisk, with varied amino acids at that particular position. As such, all TCR-β sequences within a given cluster shared 1 CDR3 motif, with variability among the TCR-β chain CDR3 sequences at the indicated positions. Importantly, these clusters containing the related TCR-β chains could be found at additional time points in those who expressed the parent PPI TCR-β, as well as in individuals who did not express the parent sequence.
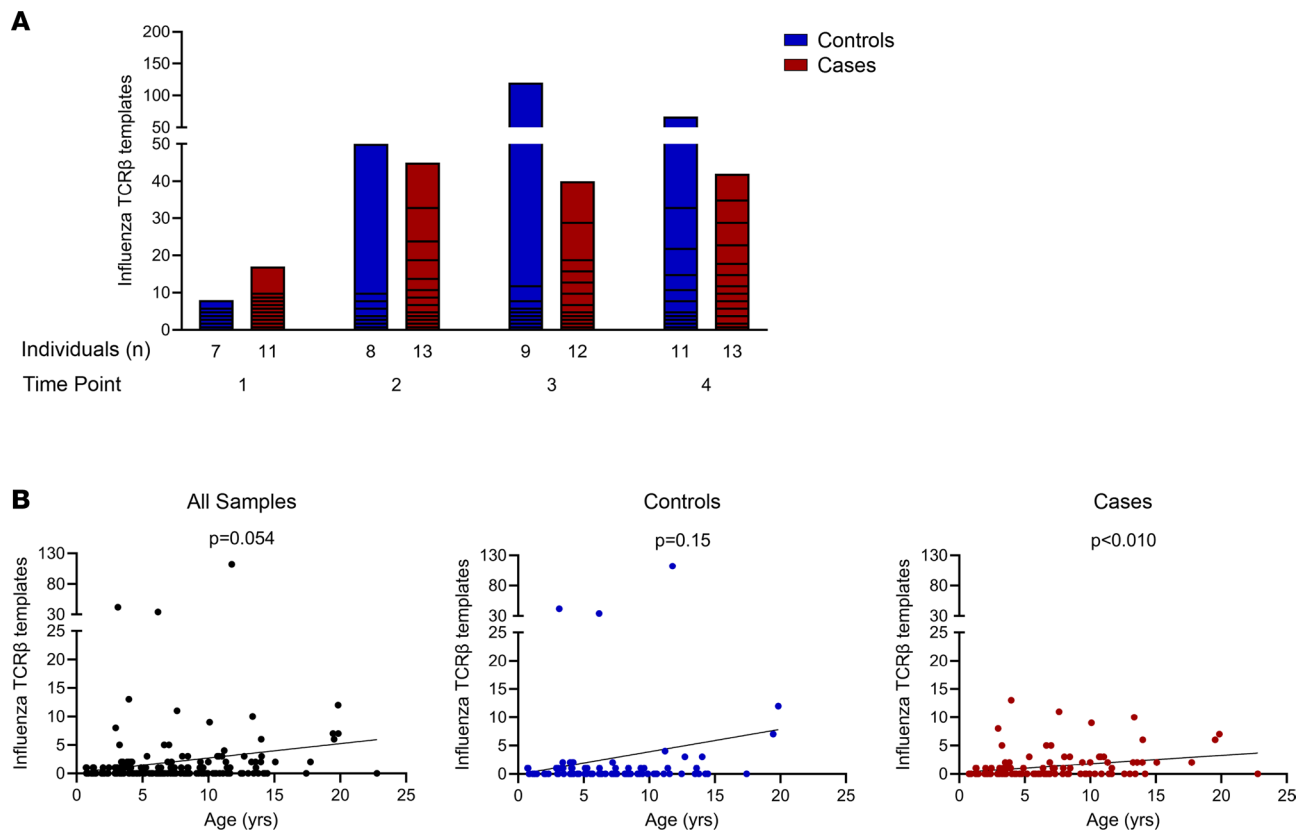
Figure 3. Presence of influenza-reactive TCR-β chain sequences correlates with increasing age. All TCR-β chain templates across individuals and time points were searched for the presence of 6 known influenza-responsive TCR-β chains (identical V, J, and CDR3 amino acid sequences), with 3 sequences identified in our data set. (**A**) Bar graphs showing influenza TCR-β chain template numbers in controls (blue) and cases (red) at the 4 time points. Time points in cases: 1, early in life; 2, before islet autoantibody positivity; 3, after islet autoantibody positivity; and 4, visit prior to clinical T1D diagnosis. Controls were age matched to cases at each time point. (**B**) Scatterplots of influenza TCR-β chain template numbers relative to age (years) for all samples (black, left), controls (blue, middle), and cases (red, right). *P* values were calculated using linear regression for template number versus age.

Among these 4 clusters, the temporal dynamics differed between cases and controls. The *AET cluster contained similar numbers of templates for both controls and cases at time point 1 (Figure 5C, first row), but templates decreased in controls and expanded in cases over time with a trend toward statistical significance ($P = 0.065$; Supplemental Table 2 for statistical comparisons using mixed effects models to account for multiple measurements and comparisons). The S*ET cluster dynamics were again similar in cases and controls until time point 4, in which there were more templates and CDR3 sequences in cases compared with controls (Figure 5C, second row). Interestingly a third cluster, SA*T, contained more unique CDR3s at time point 1 in controls that decreased over time, along with template numbers, indicating a contraction of this cluster repertoire (Figure 5C, third row). In contrast, this SA*T cluster in cases stayed relatively stable in terms of CDR3 diversity, but template numbers increased over time, signifying a clonal expansion. The fourth cluster, SAE*, was found infrequently in both groups (Figure 5C, fourth row). Next, we analyzed all samples for clusters containing the second and third most frequent PPI TCR-βs, CD4 #2 and CD8 #7 (Table 2). For CD4 #2, 4 motifs were identified, with 1 cluster showing differences in diversity at time point 1 between cases and controls (Supplemental Figure 8 and Supplemental Table 3). There were 6 distinct clusters for CD8 #7 with differences in template number and diversity at specific time points when comparing cases and controls (Supplemental Figure 9 and Supplemental Table 4).

Overall, TCR clustering allowed for the identification and visualization of hundreds of additional TCR-β chain sequences predicted to recognize similar antigens to those as PPI-specific TCR-βs. Several of these clusters preferentially expanded during diabetes development, providing a potentially powerful biomarker to monitor the progression during T1D development.

Validation of PPI-reactive TCR-β chain sequences and clusters in a new-onset T1D cohort. Lastly, we aimed to validate our findings from DAISY by performing the same sequencing and analyses in a separate
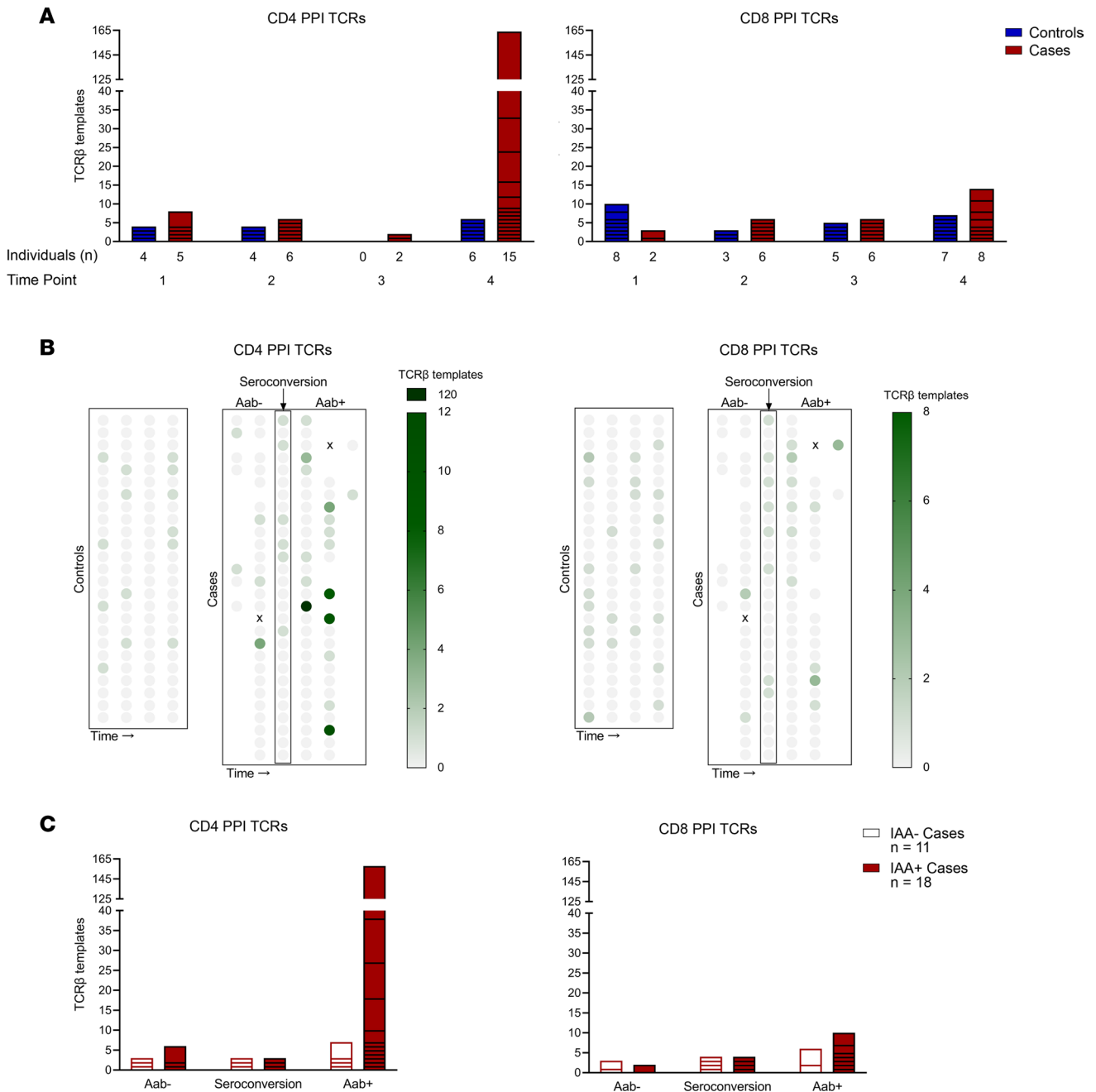
Figure 4. Preproinsulin-reactive TCR-β chain sequences are enriched during progression to type 1 diabetes. All TCR-β chain templates across individuals and time points were searched for the presence of 44 known preproinsulin-responsive (PPI-responsive) TCR-β chains (identical V, J, and CDR3) with 15 sequences identified in our data set. (**A**) Bar graphs showing template numbers of CD4 (left) and CD8 (right) PPI TCR-βs in controls (blue) and cases (red) by time point. Time points in cases: 1, early in life; 2, before islet autoantibody positivity; 3, after islet autoantibody positivity; and 4, visit prior to clinical T1D diagnosis. Controls were age matched to cases at each time point. (**B**) Multivariable plots displaying CD4 (left) and CD8 (right) PPI TCR-β template numbers in controls and cases, with visits aligned by islet autoantibody seroconversion in cases. Darker green indicates a higher template number. (**C**) Stacked bar graphs depicting CD4 (left) and CD8 (right) PPI TCR-β templates in cases who developed insulin autoantibodies (IAA⁺, red) and those who remained insulin autoantibody negative (IAA⁻, white), aligned by timing of seroconversion to any of the 4 islet autoantibodies.

cohort of new-onset T1D patients (*n* = 143). The new-onset cohort was well matched to DAISY cases and controls in terms of sex, ethnicity, HLA risk for T1D, and numbers of TCR-β chains sequenced (Supplemental Table 5 and Supplemental Data Set 1). Importantly, samples from the new-onset individuals were collected on average 5 days (median 2 days) after starting exogenous insulin treatment, such that insulin injections were unlikely to alter the TCR repertoire. We identified 8 of 15 PPI TCR-β sequences in new-onset individuals (2 CD4 and 6 CD8), and both of the most frequently identified PPI

Table 2. Pancreatic islet–derived preproinsulin-reactive TCR-β chain sequences.

TCR ID	T cell Subset	TCR-β chain			HLA restriction[A]	PPI epitope	TCR-β total templates (No.)[B]		
		V gene	J gene	CDR3 amino acid sequence			DAISY controls (*n* = 25)	DAISY cases (*n* = 29)	New-onset cohort (*n* = 143)
1	CD4	02-01	02-05	CASSAETQYF	DQ8	33-47 (B:9-23)	11	170	17
2	CD4	11-02	01-06	CASSLQSSYNSPLHF	DR4	87-101	2	6	0
3	CD4	27-01	01-01	CASSPGPGNTEAFF	DQ8-trans	86-100	0	2	1
4	CD4	04-01	02-07	CASSQDLAGVREQYF	DQ8-trans	72-87	1	0	0
5	CD4	05-01	02-05	CASSFRQGEQETQYF	DP4	29-43	0	1	0
6	CD4	07-02	02-01	CASSLTAGLASTYNEQFF	DQ8	33-47 (B:9-23)	0	1	0
7	CD8	19-01	02-01	CASSIAGGNEQFF	C*03:04	1-11	12	15	22
8	CD8	09-01	02-03	CASSVVGLGTDTQYF	A*02:01	3-11	4	3	2
9	CD8	09-01	02-01	CASSVGMDPGLGYNEQFF	C*03:04	69-79	0	6	0
10	CD8	06-05	02-01	CASRPTSGGYNEQFF	C*05:01	91-99	3	1	2
11	CD8	27-01	01-01	CASSLIGLNTEAFF	A*02:01	31-41	2	1	7
12	CD8	07-02	02-02	CASSLVVELFF	A*01:01	96-103	1	1	0
13	CD8	07-09	02-01	CASSLAQREQFF	B*41:02	91-100/ 92-102	2	0	0
14	CD8	29-01	02-01	CSVQVYNEQFF	B*18:01	91-100	1	1	1
15	CD8	07-09	02-02	CASSGGNTGELFF	A*29:02	45-53	0	1	2

[A]DQ8 consists of the alleles DQA1*03:01-DQB1*03:02. DR4 is the DRB1*04:04 allele. DQ8-trans contains the alleles DQA1*05:01-DQB1*03:02. DP4 consists of the alleles DPA1*01:03-DPB1*04:01. [B]DAISY cases and controls had TCR-β sequencing performed at 4 separate time points†New-onset T1D cohort had sequencing performed once at diagnosis.

TCR-β sequences in DAISY, CD4 #1, and CD8 #7 were also present within multiple individuals in the new-onset cohort (Table 2). Furthermore, we analyzed the new-onset samples expressing the PPI TCR-β chains for the presence of the clusters identified in the DAISY cohort. The 4 clusters from PPI CD4 #1 and the 6 from PPI CD8 #7 were identified in new-onset individuals who expressed the parent TCR-β chain (Supplemental Figure 10, A and B), confirming the DAISY results in another cohort consisting of new-onset T1D patients.

Taken together, these findings reveal that identical V, J, and CDR3 sequences from 8 pancreatic PPI TCR-β chains and clusters around these antigen-specific sequences can be identified in the peripheral blood of those at risk and with newly diagnosed T1D, indicating a continued role for these TCRs throughout disease development.

Discussion

Relatively little is known about the composition, dynamics, and development of human TCR repertoires early in life and into childhood; however, decades-spanning cross-sectional (6, 7) studies have shown that TCR-β diversity narrows with increasing age, and short-term longitudinal studies (up to 2 years) indicate repertoire stability (31, 32). Our studies now provide a robust longitudinal data set of TCR-β repertoires from children early in life throughout childhood, and we demonstrate that TCR-β repertoire diversity decreases with age while clonality increases, as demonstrated by expansion of influenza A–reactive TCR-β sequences. Importantly, the notion of clonal expansion over time is applicable to T cells specific to the self-antigen insulin, as these TCR-β sequences were enriched in individuals developing T1D. Understanding TCR repertoire changes early in life is exceedingly important for delineating the role of T cells in diseases that present during childhood, including T1D.

By sequencing the TCR-β repertoire from peripheral blood in children at risk for T1D, we demonstrate that, based upon Vβ gene usage early in life, a subset of children could be identified who later developed distinct islet autoantibody profiles with antibodies directed against GAD and IA-2A followed by clinical diabetes. Similarly, in a recent longitudinal study from children genetically at risk for T1D, gene expression analyses identified differences in peripheral blood NK cell and memory CD4 T cell transcriptome signatures

A TCRβ Sequences → TCRβ Clustering → Clusters → Specific Clusters → Cluster Characteristics
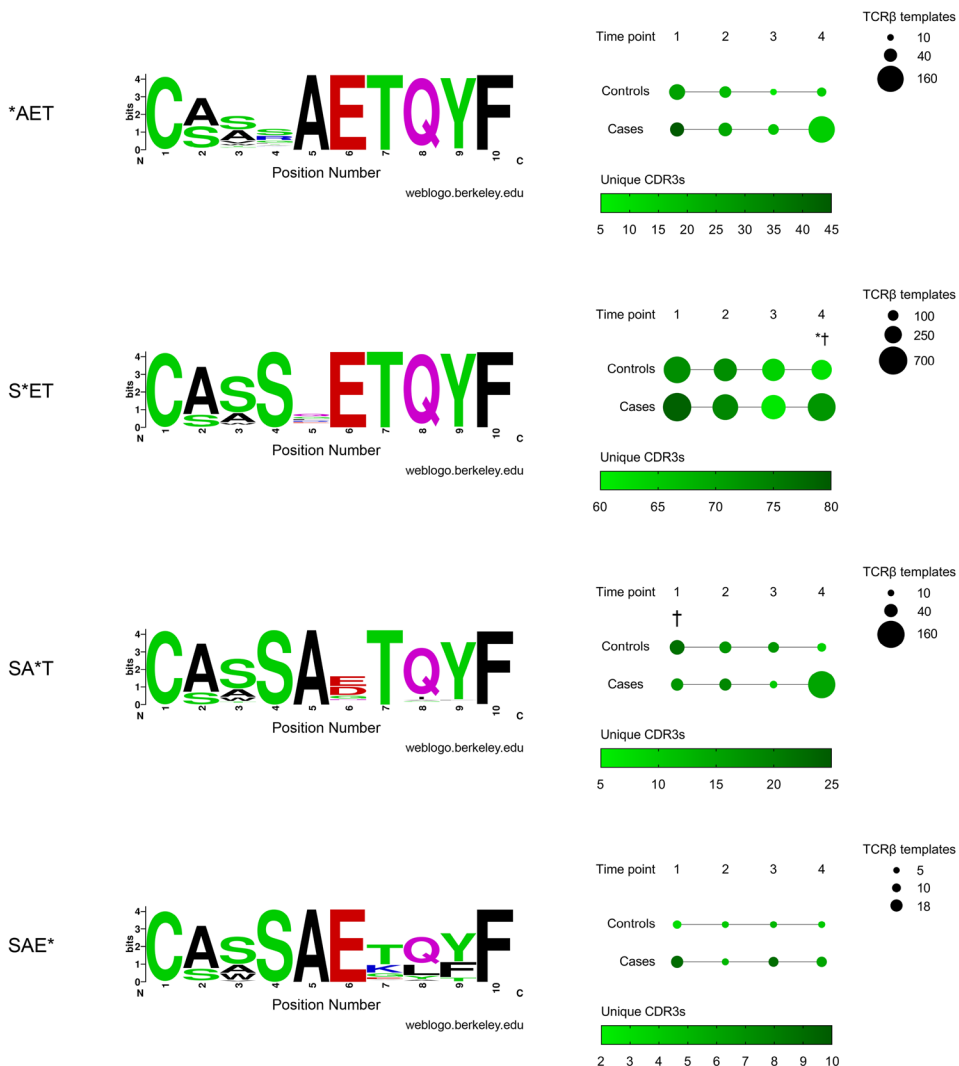
B
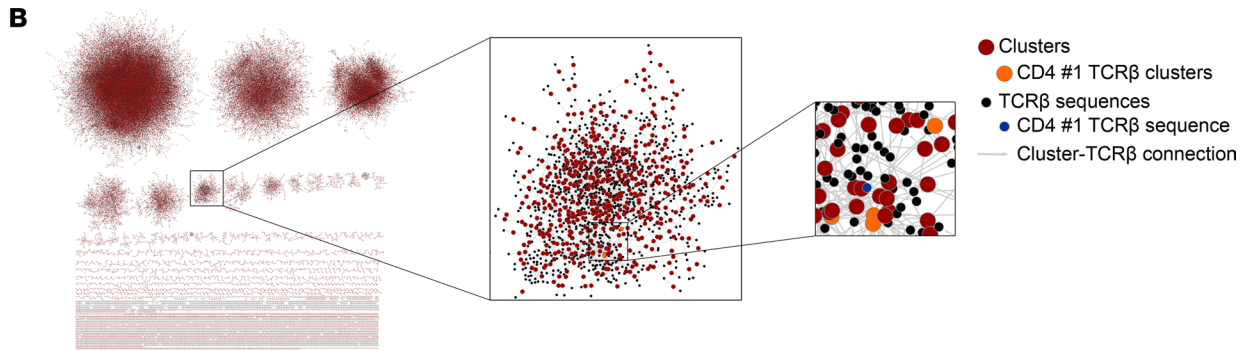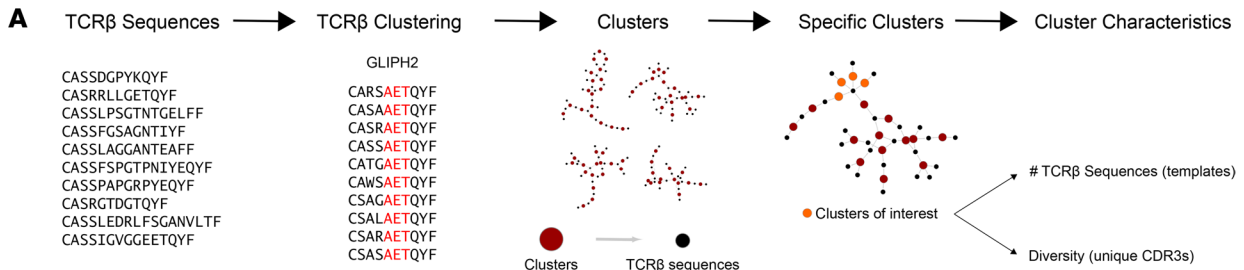
C CD4 TCRβ #1: CASSAETQYF

Figure 5. Clusters of TCR-β chain sequences, predicted to recognize similar antigens to preproinsulin, expand during diabetes development. (A) Schematic depicting the clustering of TCR-β sequences and subsequent characterization of identified clusters containing preproinsulin-reactive TCR-β chains. **(B)** A representation of TCR-β chain sequence clustering from a single individual at 1 time point generated using the GLIPH2 program; depicted are 68,639 total clusters from case #16 at time point 3. TCR-β chain sequences (black) are grouped into clusters (red) based upon predictions to bind similar peptides. CD4 TCR #1 (blue) that recognizes insulin B chain amino acids 9–23 were found in 4 TCR clusters (orange) at this time point. **(C)** Sequence logos for the 4 clusters that contain the PPI-reactive CD4 TCR #1 showing the frequency of amino acids at each position within the CDR3β sequences (left). Larger letters indicate a higher prevalence of an amino acid at a particular position. Multivariable plots in panels depict the 4 cluster motifs at each time point in controls and cases (right). Dot size indicates the number of TCR-β chain templates composing the cluster, while a darker green color depicts a higher number of unique CDR3β sequences in the cluster (a measure of TCR diversity). Time points in cases: 1, early in life; 2, before islet autoantibody positivity; 3, after islet autoantibody positivity; and 4, visit prior to clinical T1D diagnosis. Controls were age matched at each time point. P values were calculated using mixed-effects models to account for multiple measurements and comparisons between controls and cases at each time point for either template number (*) or CDR3β diversity (†). *$P < 0.05$, †$P < 0.05$. Supplemental Table 2 provides full statistics for temporal changes within a cohort and for comparisons between cases and controls.

that correlated with the development of initial islet autoantibody specificity and clinical diabetes onset (33). Furthermore, specific gene signatures alongside patterns of islet autoantibodies tracked with progression to T1D at an individual level. These results and our findings indicate that changes in immunological signatures, transcriptome profiling, and TCR-β chain sequencing during early childhood may help delineate individuals who are more likely to develop islet autoantibodies and progress to clinical T1D.

In general, the TCR repertoire within a given individual is primarily composed of private TCRs and minimally overlaps with those of others (5, 34–36). However, advances in next-generation sequencing have allowed for deep sequencing of the TCR repertoire from more individuals, resulting in the identification of specific sequences that are shared between individuals and associated with different disease states (37–39). We were able to identify both known influenza A–reactive and PPI-responding TCR-β chains in our data set. We focused on PPI-reactive T cells, as these have been well characterized from the pancreatic islet infiltrates of T1D organ compared with T cells directed against GAD and IA-2A (12, 14, 19–22), despite finding specific Vβ genes early in life associated with a subset of cases that developed GAD and IA-2A autoantibodies. Of the 44 PPI-reactive TCR-β chains identified from pancreatic islets of T1D organ donors, we show that the exact V, J, and CDR3 from 15 of 44 sequences (34%) are shared in the peripheral blood of individuals genetically at risk for T1D. Additionally, 8 of the 15 insulin-reactive TCR-β chains were found in a separate cohort of new-onset T1D patients, indicating a continued role for these TCRs throughout disease development. The lower number of TCR-β sequences in our new-onset cohort may be indicative of some TCRs being more involved earlier in the disease process, prior to clinical symptoms. Alternatively, there may be more β cell destruction in new-onset individuals compared with those in the preclinical stages of T1D, as evidenced by a recent study demonstrating that proliferation of proinsulin-specific CD4 T cells in children with new-onset T1D correlated to residual β cell function (40).

Interestingly, CD4 PPI–reactive TCR-β sequences were more prevalent than PPI CD8 sequences in the peripheral blood prior to clinical T1D and in our new-onset cohort. It is appreciated that HLA class I is overexpressed on the residual insulin containing pancreatic islets from T1D organ donors (41), and CD8 T cells are the more abundant T cell subtype infiltrating pancreatic islets (11, 13, 21). Thus, it is plausible that the CD8 PPI–reactive TCR-β sequences studied here may have been more localized to the pancreas and not as prevalent in the peripheral blood as compared with PPI CD4 TCR-β chain sequences.

It should be noted that our study focused on TCR-β sequencing, and as sequencing technology continues to advance, it will be important to deep sequence both TCR α and β chains, including paired sequences from a large number of single T cells. Having paired α and β chains from individual T cells would then allow for confirmation of antigen specificity of a given TCR, such as those identified in our clustering analyses. Undersampling of the vast TCR repertoire is also a limitation, and future studies should focus on deeper sequencing TCR-β chains. The potential exists to target specific Vβ genes and to deep sequence those TCR-β chains to identify more individuals with a given receptor of interest, such as those from PPI-reactive T cells in individuals progressing through the stages of T1D.

Importantly, we demonstrate that clusters containing TCR-β chains predicted to bind PPI peptides could be found at additional time points in individuals who expressed the parent TCR-β chain, as well as at multiple time points of those who never expressed it, thus allowing for a visualization of disease-related TCR numbers and CDR3 diversity over time. One advantage of sequencing and clustering TCRs is that it obviates the need for functional T cell assays to monitor disease-relevant cells. Although informative, functional T cell assays (e.g., fluorescent peptide-MHC multimer staining, proliferation assays with fluorescent

cell staining dyes, and cytokine enzyme linked immunospot [ELISpot] assays) are often challenging to perform due to variables such as the necessity for large blood volumes, the need for living cells, and biases that may occur due to operator use.

While sequencing TCRs directly from DNA has advantages, cell phenotype information cannot be obtained. However, Bonifacio and colleagues longitudinally examined the gene expression profiles of CD4 T cells that proliferated to proinsulin and GAD65 beginning early in life from children at risk for T1D (42). Islet-responsive naive CD4 T cells were present in all children at 6 months of age; however, these cells developed a memory phenotype with a predominant inflammatory gene expression profile after islet autoantibody seroconversion. Furthermore, the proinsulin- and GAD-reactive CD4 T cells proliferated more robustly following islet autoantibody development compared with children without autoantibodies. These results support the findings in our study in which PPI CD4 TCR-β sequences and clusters containing these sequences are present early in life and then expand following islet autoantibody seroconversion indicating that preproinsulin- and proinsulin-responsive CD4 T cells are present early and mature during disease progression.

In conclusion, our studies reveal that, in T1D, where disease-relevant TCRs have been obtained from the target organ, sequencing and clustering of related TCR-β chains in the peripheral blood provide insight into disease progression, as there is an enrichment of PPI-reactive sequences during the progression to clinical disease. TCR sequencing may provide a robust and reproducible biomarker during the course of T1D development and aid in the timing and monitoring of therapies for disease prevention. Our results also provide a framework for using TCR sequences in fields of medicine where disease-relevant sequences are known, including autoimmune diseases, infections, and cancers.

Methods

Study design. Subjects were recruited into the DAISY, which is an ongoing prospective birth cohort study following genetically at-risk children and adolescents for the development of islet autoantibodies and clinical T1D (15). Study subjects for the new-onset T1D cohort were recruited from the Barbara Davis Center for Diabetes Clinics. Peripheral blood was obtained for islet autoantibody measurements, HLA typing, and TCR-β sequencing.

Islet autoantibody measurements and HLA genotyping. Serum obtained from peripheral blood was used to measure islet autoantibodies to GADA, tyrosine phosphatase–related IA-2A, ZnT8A, and IAA using fluid-phase radio-binding assays as previously described (43, 44). ZnT8A was measured from participants beginning in 2010 and later. Harmonized assays for GADA and IA-2A were available starting in 2010, and index values obtained before this time were converted to DK units. The same positive cut-off values were applied to all participants with an index of 0.010 for IAA, 20 DK units for GADA, 5 DK units for IA-2A, and an index 0.020 for ZnT8A. Previously collected and stored DNA samples from DAISY participants were used to type HLA class I (A, B, and C) and class II alleles (DQA1, DQB1, DRB1, DPA1, and DPB1) using a targeted next-generation sequencing assay with hybrid capture technology following the manufacturer's instructions (AlloSeq Tx17 from CareDx Lab Solutions Inc.). For the new-onset T1D cohort, HLA typing of the HLA-DRB1, DQA1, and DQB1 alleles was performed using oligonucleotide probes as previously described (45).

TCR-β chain immunosequencing. Genomic DNA was extracted from cryopreserved buffy coat samples obtained from peripheral blood following the manufacturer's instructions using the QIAamp Blood Extraction kit (Qiagen). TCR-β chains were sequenced using the immunoSEQ Assay (Adaptive Biotechnologies) as previously described (4, 46). The identification of specific V and J genes was carried out according to the definitions established by the International ImMunoGeneTics (IMGT) collaboration (47). Raw data processing and sequencing analyses were performed using the immunoSEQ Analyzer 3.0 online platform (http://www.adaptivebiotech.com/immunoseq). The following analyses were performed directly through the immunoSEQ Analyzer 3.0: Simpson productive clonality measurements, number of productive TCR-β templates, and PPI TCR-β searches. The calculations for TCR-β V gene (TRBV) usage were performed in Microsoft Excel after downloading the raw data from the immunoSEQ Analyzer 3.0. Subsequently, the TRBV usage data were then used in the PCA conducted using GraphPad Prism 9.2 (GraphPad Software).

TCR-β chain clustering algorithm and analyses. After the sequencing data for each individual at each time point were downloaded from the immunoSEQ Analyzer 3.0 platform, clustering analyses were performed using the GLIPH2 online platform (http://50.255.35.37:8080/) (25). For all clustering, the reference database version 2.0 was selected (reference data set consisting of bulk sequencing data from T cells) with HLA typing included. All samples were first run against the CD4 reference database to identify the CD4 PPI TCR-β clusters, and then all samples were run against the CD8 reference database to identify the CD8 PPI

TCR-β clusters. Further analyses of clusters were then performed using Microsoft Excel. Both Cytoscape version 3.8.2 and GraphPad Prism 9.2 software were used to graph the clustering data.

Data and materials availability. All immunosequencing data generated in this study are freely available for analysis and download from the Adaptive Biotechnologies immuneACCESS site under the immune-ACCESS terms of use (https://clients.adaptivebiotech.com/pub/mitchell-2022-JCII).

Statistics. Statistical analyses were performed using R software version 4.1 (R Core Team, Vienna), GraphPad Prism 9.2 (GraphPad Software), and SAS 9.4 (SAS Institute). $P < 0.05$ was considered significant. Continuous variables (age at time points) were compared using 2-tailed t tests, and Fisher's exact test was used for categorical variables (sex, race, ethnicity, presence of islet autoantibodies, proportion of individuals with a given HLA type, proportions having specific TCR-β chain sequences, and TCR-β chain template numbers). Linear regression analyses were used to compare Simpson productive clonality to age. Mann-Whitney U tests were used to compare Simpson productive clonality in cases versus controls at each time point and to compare TRBV gene usage, age at seroconversion, and age at onset in cases versus the subset of cases. Mixed-effects models were used to compare groups and time points while accounting for the correlation of multiple measurements within a participant. Tukey's Honestly Significant Difference test was used to control for multiple testing when performing pairwise comparisons.

Study approval. The clinical investigation in this study was conducted in accordance with the Declaration of Helsinki principles, with study approvals provided by the Colorado Multiple IRB. All participants — or participant guardians, if the participant was less than 18 years of age — provided written informed consent.

Author contributions

AMM, AKS, KMS, MN, PAG, MJR, and AWM designed the studies. AMM, EEB, KAM, KW, and LY performed the research. LP conducted an independent statistical analysis of the study data. AMM, AKS, KMS, MN, PAG, MJR, and AWM wrote the manuscript. AWM is the guarantor of this work, had full access to all the data in the study, and takes responsibility for the data integrity and accuracy of analyses.

Acknowledgments

Address correspondence to: Aaron Michels, Barbara Davis Center for Diabetes, 1775 Aurora Court, Building M20, A140, Aurora, Colorado 80045, USA. Phone: 303.724.1923; Email: aaron.michels@cuanschutz.edu.

1. Huppa JB, Davis MM. The interdisciplinary science of T-cell recognition. *Adv Immunol*. 2013;119:1–50.
2. Rossjohn J, et al. T cell antigen receptor recognition of antigen-presenting molecules. *Annu Rev Immunol*. 2015;33:169–200.
3. Freeman JD, et al. Profiling the T-cell receptor beta-chain repertoire by massively parallel sequencing. *Genome Res*. 2009;19(10):1817–1824.
4. Robins HS, et al. Comprehensive assessment of T-cell receptor beta-chain diversity in alphabeta T cells. *Blood*. 2009;114(19):4099–4107.
5. Robins HS, et al. Overlap and effective size of the human CD8+ T cell receptor repertoire. *Sci Transl Med*. 2010;2(47):47ra64.
6. Britanova OV, et al. Dynamics of individual T cell repertoires: from cord blood to centenarians. *J Immunol*. 2016;196(12):5005–5013.
7. Britanova OV, et al. Age-related decrease in TCR repertoire diversity measured with deep and normalized sequence profiling. *J Immunol*. 2014;192(6):2689–2698.
8. Bluestone JA, et al. Genetics, pathogenesis and clinical interventions in type 1 diabetes. *Nature*. 2010;464(7293):1293–1300.
9. Atkinson MA, Eisenbarth GS, Michels AW. Type 1 diabetes. *Lancet*. 2014;383(9911):69–82.
10. Insel RA, et al. Staging presymptomatic type 1 diabetes: a scientific statement of JDRF, the Endocrine Society, and the American Diabetes Association. *Diabetes Care*. 2015;38(10):1964–1974.
11. Coppieters KT, et al. Demonstration of islet-autoreactive CD8 T cells in insulitic lesions from recent onset and long-term type 1 diabetes patients. *J Exp Med*. 2012;209(1):51–60.
12. Pathiraja V, et al. Proinsulin-specific, HLA-DQ8, and HLA-DQ8-transdimer-restricted CD4+ T cells infiltrate islets in type 1 diabetes. *Diabetes*. 2015;64(1):172–182.
13. Campbell-Thompson M, et al. Insulitis and β-cell mass in the natural history of type 1 diabetes. *Diabetes*. 2016;65(3):719–731.
14. Babon JAB, et al. Analysis of self-antigen specificity of islet-infiltrating T cells from human donors with type 1 diabetes. *Nat Med*. 2016;22(12):1482–1487.
15. Rewers M, et al. Newborn screening for HLA markers associated with IDDM: diabetes autoimmunity study in the young

(DAISY). *Diabetologia*. 1996;39(7):807–812.

16. Hennecke J, Wiley DC. Structure of a complex of the human alpha/beta T cell receptor (TCR) HA1.7, influenza hemagglutinin peptide, and major histocompatibility complex class II molecule, HLA-DR4 (DRA*0101 and DRB1*0401): insight into TCR cross-restriction and alloreactivity. *J Exp Med*. 2002;195(5):571–581.

17. Ostrov DA, et al. Methyldopa blocks MHC class II binding to disease-specific antigens in autoimmune diabetes. *J Clin Invest*. 2018;128(5):1888–1902.

18. Chow I-T, et al. Discriminative T cell recognition of cross-reactive islet-antigens is associated with HLA-DQ8 transdimer-mediated autoimmune diabetes. *Sci Adv*. 2019;5(8):eaaw9336.

19. Landry LG, et al. Proinsulin-reactive CD4 T cells in the islets of type 1 diabetes organ donors. *Front Endocrinol (Lausanne)*. 2021;12:622647.

20. Michels AW, et al. Islet-derived CD4 T cells targeting proinsulin in human autoimmune diabetes. *Diabetes*. 2017;66(3):722–734.

21. Anderson AM, et al. Human islet T cells are highly reactive to preproinsulin in type 1 diabetes. *Proc Natl Acad Sci U S A*. 2021;118(41):e2107208118.

22. Delong T, et al. Pathogenic CD4 T cells in type 1 diabetes recognize epitopes formed by peptide fusion. *Science*. 2016;351(6274):711–714.

23. Ziegler AG, et al. Seroconversion to multiple islet autoantibodies and risk of progression to diabetes in children. *JAMA*. 2013;309(23):2473–2479.

24. Glanville J, et al. Identifying specificity groups in the T cell receptor repertoire. *Nature*. 2017;547(7661):94–98.

25. Huang H, et al. Analyzing the Mycobacterium tuberculosis immune response by T-cell receptor clustering with GLIPH2 and genome-wide antigen screening. *Nat Biotechnol*. 2020;38(10):1194–1202.

26. Alleva DG, et al. A disease-associated cellular immune response in type 1 diabetics to an immunodominant epitope of insulin. *J Clin Invest*. 2001;107(2):173–180.

27. Lee KH, et al. Structure of a human insulin peptide-HLA-DQ8 complex and susceptibility to type 1 diabetes. *Nat Immunol*. 2001;2(6):501–507.

28. Yang J, et al. Autoreactive T cells specific for insulin B:11-23 recognize a low-affinity peptide register in human subjects with autoimmune diabetes. *Proc Natl Acad Sci U S A*. 2014;111(41):14840–14845.

29. Nakayama M, et al. Regulatory versus inflammatory cytokine T-cell responses to mutated insulin peptides in healthy and type 1 diabetic subjects. *Proc Natl Acad Sci U S A*. 2015;112(14):4429–4434.

30. Wang Y, et al. C-terminal modification of the insulin B:11-23 peptide creates superagonists in mouse and human type 1 diabetes. *Proc Natl Acad Sci U S A*. 2018;115(1):162–167.

31. Chu ND, et al. Longitudinal immunosequencing in healthy people reveals persistent T cell receptors rich in highly public receptors. *BMC Immunol*. 2019;20(1):19.

32. Servaas NH, et al. Longitudinal analysis of T-cell receptor repertoires reveals persistence of antigen-driven CD4+ and CD8+ T-cell clusters in systemic sclerosis. *J Autoimmun*. 2021;117:102574.

33. Xhonneux LP, et al. Transcriptional networks in at-risk individuals identify signatures of type 1 diabetes progression. *Sci Transl Med*. 2021;13(587):eabd5666.

34. Li H, et al. Determinants of public T cell responses. *Cell Res*. 2012;22(1):33–42.

35. Venturi V, et al. A mechanism for TCR sharing between T cell subsets and individuals revealed by pyrosequencing. *J Immunol*. 2011;186(7):4285–4294.

36. Venturi V, et al. The molecular basis for public T-cell responses? *Nat Rev Immunol*. 2008;8(3):231–238.

37. Mitchell AM, Michels AW. T cell receptor sequencing in autoimmunity. *J Life Sci (Westlake Village)*. 2020;2(4):38–58.

38. Kirsch I, et al. T-cell receptor profiling in cancer. *Mol Oncol*. 2015;9(10):2063–2070.

39. Schultheiß C, et al. Next-generation sequencing of T and B cell receptor repertoires from COVID-19 patients showed signatures associated with severity of disease. *Immunity*. 2020;53(2):442–455.

40. Musthaffa Y, et al. Proinsulin-specific T-cell responses correlate with estimated c-peptide and predict partial remission duration in type 1 diabetes. *Clin Transl Immunology*. 2021;10(7):e1315.

41. Richardson SJ, et al. Islet cell hyperexpression of HLA class I antigens: a defining feature in type 1 diabetes. *Diabetologia*. 2016;59(11):2448–2458.

42. Heninger A-K, et al. A divergent population of autoantigen-responsive CD4+ T cells in infants prior to β cell autoimmunity. *Sci Transl Med*. 2017;9(378):eaaf8848.

43. Yu L, et al. Antiislet autoantibodies usually develop sequentially rather than simultaneously. *J Clin Endocrinol Metab*. 1996;81(12):4264–4267.

44. Wenzlau JM, et al. The cation efflux transporter ZnT8 (Slc30A8) is a major autoantigen in human type 1 diabetes. *Proc Natl Acad Sci U S A*. 2007;104(43):17040–17045.

45. Rewers A, et al. Ethnic differences in the associations between the HLA-DRB1*04 subtypes and type 1 diabetes. *Ann N Y Acad Sci*. 2003;1005:301–309.

46. Carlson CS, et al. Using synthetic templates to design an unbiased multiplex PCR assay. *Nat Commun*. 2013;4:2680.

47. Yousfi Monod M, et al. IMGT/JunctionAnalysis: the first tool for the analysis of the immunoglobulin and T cell receptor complex V-J and V-D-J JUNCTIONs. *Bioinformatics*. 2004;20 Suppl 1:i379-85.