



OPEN ACCESS

# Characterisation of T cell receptor repertoires in coeliac disease

Lik Wee Lee,<sup>1</sup> Shahin Shafiani ,<sup>1</sup> Beryl Crossley,<sup>1</sup> Ryan O Emerson,<sup>1</sup> David Williamson,<sup>1</sup> Anna Bunin,<sup>2</sup> Justin Vargas,<sup>2</sup> Arnold S Han,<sup>2</sup> Ian M Kaplan,<sup>1</sup> Peter H R Green,<sup>2</sup> Ilan Kirsch,<sup>1</sup> Govind Bhagat

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/jcp-2022-208541>).

<sup>1</sup>Computational Biology and Translational Medicine, Adaptive Biotechnologies Corp, Seattle, Washington, USA

<sup>2</sup>Department of Medicine, Celiac Disease Center, Columbia University Irving Medical Center, New York, New York, USA

<sup>3</sup>Department of Pathology and Cell Biology and Department of Medicine, Celiac Disease Center, Columbia University Irving Medical Center, New York, New York, USA

## Correspondence to

Dr Govind Bhagat, Department of Pathology and Cell Biology and Department of Medicine, Celiac Disease Center, Columbia University Irving Medical Center, New York, New York, USA; [gb96@cumc.columbia.edu](mailto:gb96@cumc.columbia.edu)

Received 8 August 2022

Accepted 23 November 2022

Published Online First

15 December 2022

## ABSTRACT

**Aims** Characterise T-cell receptor gene (TR) repertoires of small intestinal T cells of patients with newly diagnosed (active) coeliac disease (ACD), refractory CD type I (RCD I) and patients with CD on a gluten-free diet (GFD).

**Methods** Next-generation sequencing of complementarity-determining region 3 (CDR3) of rearranged T cell receptor  $\beta$  (TRB) and  $\gamma$  (TRG) genes was performed using DNA extracted from intraepithelial cell (IEC) and lamina propria cell (LPC) fractions and a small subset of peripheral blood mononuclear cell (PBMC) samples obtained from CD and non-CD (control) patients. Several parameters were assessed, including relative abundance and enrichment.

**Results** TRB and TRG repertoires of CD IEC and LPC samples demonstrated lower clonality but higher frequency of rearranged TRs compared with controls. No CD-related differences were detected in the limited number of PBMC samples. Previously published LP gliadin-specific TRB sequences were more frequently detected in LPC samples from patients with CD compared with non-CD controls. TRG repertoires of IECs from both ACD and GFD patients demonstrated increased abundance of certain CDR3 amino acid (AA) motifs compared with controls, which were encoded by multiple nucleotide variants, including one motif that was enriched in duodenal IECs versus the PBMCS of CD patients.

**Conclusions** Small intestinal TRB and TRG repertoires of patients with CD are more diverse than individuals without CD, likely due to mucosal recruitment and accumulation of T cells because of protracted inflammation. Enrichment of the unique TRG CDR3 AA sequence in the mucosa of patients with CD may suggest disease-associated changes in the TCR $\gamma\delta$  IEL lymphocyte (IEL) landscape.

## INTRODUCTION

Coeliac disease (CD) is an immune-mediated malabsorption syndrome occurring in genetically susceptible individuals intolerant to dietary gluten.<sup>1</sup> The pathogenesis of CD involves the generation of effector innate and adaptive immune responses.<sup>1</sup> The latter involves activation of gliadin-specific CD4<sup>+</sup> TCR $\alpha\beta$  T cells within regional lymphoid tissue, with trafficking of effector memory gliadin responsive CD4<sup>+</sup> T cells to the lamina propria (LP), and concomitant stimulation of humoral immune responses directed towards tissue transglutaminase

## WHAT IS ALREADY KNOWN ON THIS TOPIC

⇒ Studies aimed at delineating coeliac disease (CD) pathogenesis have mostly focused on characterising small intestinal and peripheral blood TR repertoires of enriched, gliadin-specific CD4<sup>+</sup> T cells.

## WHAT THIS STUDY ADDS

⇒ Analysis of the overall TR repertoires in the small intestinal IE and LP compartments of CD patients, in an unbiased manner, revealed a higher T cell density and lower clonality in both compartments in untreated (active) CD and refractory CD type I, and confirmed the presence of public gliadin-specific TRB sequences in the LP in a high proportion of CD cases. Certain public TCR $\gamma\delta$  T cell-derived TRG CDR3 AA motifs, encoded by multiple nucleotide variants, were found to be more abundant in the IE compartment of CD patients compared with controls.

## HOW THIS STUDY MIGHT AFFECT RESEARCH, PRACTICE OR POLICY

⇒ The identification of CD-specific public TRG sequence motifs may help uncover yet unknown facets of CD biology and assist in the development of novel diagnostic biomarkers.

type 2 and deamidated gliadin. Activation of innate intraepithelial (IE) CD8<sup>+</sup> TCR $\alpha\beta$  T cells has been shown to play a role in epithelial damage.<sup>2</sup> Recruitment of TCR $\gamma\delta$  T cells to the IE compartment is also a characteristic feature of CD.<sup>3</sup> Indeed, IE TCR $\gamma\delta$  T cells are increased in all stages of CD, with some reports suggesting immunoregulatory roles for a subset.<sup>4–9</sup>

Investigations of specific innate and adaptive immune cells have contributed to a better understanding of cellular networks and signalling pathway that are dysregulated in CD. However, the breadth of and temporal changes in immune responses, both locally and systemically, remain to be clarified. Analysis of the immune receptor repertoires to determine changes in B and T cell diversity in different tissue compartments has enabled dissection of antigen-specific and non-specific immune responses in an unbiased manner,<sup>10 11</sup> and for T cell-mediated diseases, presence of clonal T cell expansions have suggested the presence of disease



© Author(s) (or their employer(s)) 2024. Re-use permitted under CC BY. Published by BMJ.

**To cite:** Lee LW, Shafiani S, Crossley B, et al. *J Clin Pathol* 2024;**77**:116–124.

**Table 1** Details of samples used in the study

Donor status	TRB locus			TRG locus			CD8 <sup>+</sup> TCRαβ <sup>+</sup>			CD4 <sup>+</sup> TCRαβ <sup>+</sup>		
	IEC	LPC	PBMC	IEC	LPC	PBMC	IEC	LPC	PBMC	IEC	LPC	IEC
Active coeliac disease (CD)	10	8	2	4	2		10	8	2	4	2	2
Refractory CD type 1	3	3	0	0	0		2	2	0	0	0	0
CD—gluten-free diet	8	7	2	2	2		9	8	2	2	2	2
Controls	6	6	2	2	2		8	8	2	2	2	2
Total	27	24	6	8	6		29	26	6	8	6	6

Table provides a breakdown by donor type, including control and CD donors categorised by CD status, the number of actual samples by tissue type subjected to bulk population analysis, and IEC and LPC samples that were subjected to cell sorting including the sorted cell type.  
IEC, intraepithelial cell; LPC, lamina propria cell; PBMC, peripheral blood mononuclear cell.

triggering antigens.<sup>12</sup> Studies of LP CD4<sup>+</sup> TCRαβ T cell repertoires in CD have helped shed light on the adaptive immune response to gluten in the small intestinal mucosa,<sup>13</sup> but our knowledge of alterations in IE innate CD8<sup>+</sup> TCRαβ and TCRγδ T cell repertoires is more limited.<sup>14 15</sup>

A variety of methods have been used to study T cell repertoires in the past including V-region-specific monoclonal antibody-based flow cytometry and TR spectratyping (immunoscope analysis), but these modalities had limited resolution.<sup>10</sup> The introduction of next-generation sequencing (NGS) has enabled comprehensive characterisation of the TRB and TRG repertoires. The current study used an NGS-based method to sequence rearranged TRB and TRG CDR3 regions of lymphocytes from different small intestinal mucosal compartments and a limited number of peripheral blood mononuclear cell (PBMC) samples from patients with different clinical presentations or phases of CD and controls, enabling quantitative evaluation of disease-associated changes in the TR repertoires.<sup>16 17</sup>

## MATERIALS AND METHODS

### Sample acquisition and isolation of cellular fractions

Altogether 83 discrete samples from 30 patients were analysed, 22 with serologically confirmed CD and 8 with non-small intestinal disorders, serving as controls. The different types of samples analysed were as follows: total intraepithelial cells (IECs) from all 30 patients, total LP cells (LPCs) from 27 patients, PBMCs from 6 patients and 20 flow-sorted samples of different IEC and LPC T-cell subsets. The details of all the samples evaluated are provided in table 1 and online table S1. Information about patient classification and clinical, pathology and laboratory data are provided in online supplemental methods and table S1. Lymphocyte fractions were isolated from small intestinal biopsies as previously described,<sup>18 19</sup> pertinent details are provided in online supplemental methods. Aliquots of 500 000 IECs and LPCs were suspended in dimethyl sulfoxide (DMSO)/fetal bovine serum (FBS) and stored in liquid nitrogen. One aliquot of each cellular fraction was used for sequencing. The PBMC isolation method is described in online supplemental methods. One PBMC aliquot, comprising 1,000,000 cells, was used for sequencing.

### Fluorescence-activated cell sorting

Cell sorting was performed using a BD FACSAria II instrument, and FACSDiva V.6.1.3 software was used for data acquisition and analysis (BD Biosciences, New Jersey, USA). Three T cell populations were sorted from a subset of cases: TCRγδ (two samples each of active CD [ACD], gluten-free diet [GFD] and

controls), CD8<sup>+</sup> TCRαβ T cells from IEC fractions (four ACD samples and two samples each of GFD and controls), and CD4<sup>+</sup> TCRαβ T cells from LPC fractions (two samples each of ACD, GFD and controls), as listed in table 1. Further details on sorting strategy are provided in online supplemental methods and the number of cells sorted for the different T cell lineages from the IEC and LPC samples are included in online table S2.

### TR sequencing

Genomic DNA was extracted from aliquots of frozen total IEC, LPC and PBMC fractions as well as sorted T cell populations using the Qiagen DNeasy blood extraction Kit (Qiagen, Gaithersburg, Maryland, USA) and sequencing of rearranged TRB and TRG loci was performed (immunoSEQ, Adaptive Biotechnologies, Washington, USA) as reported previously<sup>16 17</sup> and described in online supplemental methods. The TR sequencing data of all the samples are available from the Adaptive Biotechnologies immuneACCESS repository: <https://doi.org/10.21417/LWL2022JCI>.<sup>20</sup>

### TR repertoire analysis

In order to obtain a representation of the total TRB and TRG repertoires for each unsorted (and sorted) sample, the nucleotide sequences representing the CDR3 regions were translated into their corresponding amino acid (AA) sequences. A number of parameters of the individual repertoires from each sample were analysed, including the total number of sequences identified, the overall clonality,<sup>17</sup> the frequency distribution of the individual sequences and the degree of convergent recombination (the number of nucleotide sequences generating the same AA sequence). The individual samples were compared for co-occurrence of sequences and a differential abundance algorithm was applied to identify AA sequences that were present in multiple samples, in order to assess for the presence of potential disease-associated sequences. These methods have been published previously.<sup>21 22</sup> We also investigated the presence of known gliadin-specific TRB clonotypes in LPC and available PBMC samples by querying the sequences against a set of published CD LP TRB sequences derived from HLA-DQ2.5 gliadin tetramer-sorted CD4<sup>+</sup> T cells.<sup>23</sup>

## RESULTS

### Characterisation of the TR repertoires

The overall TR repertoire in each sample was characterised using two summary statistics: TR repertoire clonality, which is a normalised measure of the evenness of the TR frequency

**Table 2** Mean productive nucleotide sequences by disease and sample type

Patient status	Patients	Avg TRG productive sequences			Avg TRB productive sequences		
		IEC	LPC	PBMC	IEC	LPC	PBMC
Active coeliac disease	10	5469	7383	69 133	6249	8598	152 197
Refractory coeliac disease type 1	3	7954	8808	N/A	8570	11 608	N/A
Coeliac disease—gluten-free diet	9	5459	6677	46 601	3553	5570	106 076
Controls	8	3826	6006	55 286	1538	5259	146 125
<b>Total</b>	<b>30</b>						

Mean of the total number of nucleotide sequences with an AA product for each locus, disease state and sample origin are provided. AA, amino acid; IEC, intraepithelial cell; LPC, lamina propria cell; N/A, not available; PBMC, peripheral blood mononuclear cell.

distribution,<sup>24</sup> and T cell density (obtained by counting the number of observed TR molecules and dividing by the amount of genomic DNA used as input).<sup>25</sup> Samples from 27 subjects were successfully sequenced for the TRB locus, including 24 paired, bulk (unsorted) IEC and LPC samples. Similarly, samples from 29 subjects were successfully sequenced for the TRG locus, including 25 with paired IEC and LPC bulk samples. Furthermore, six PBMC samples underwent successful sequencing of the TRB and TRG loci. **Table 2** shows the mean productive nucleotide sequences by disease and sample type and **figure 1** shows the distribution characteristics of the repertoire, separated by locus, tissue type and disease status. As shown in **figure 1**, the TR repertoire was fairly similar across loci and tissue types; among the 25 IEC/LPC pairs with informative TRG sequences and 24 IEC/LPC pairs with informative TRB sequences, both clonality and TR density were highly correlated between patients when comparing TRB vs TRG repertoires (Spearman's  $r = X@0.75$  and  $@0.81$ , respectively) and when comparing IEC vs LPC repertoires (Spearman's  $r = X@0.70$  and  $@0.71$ , respectively).

However, repertoire metrics differed across disease states; within each repertoire/locus/tissue combination. **Figure 1** shows the results of a Kruskal-Wallis test with a post hoc Dunn test to determine significant differences between pairs of disease states. IEC and LPC samples (the two categories with the highest sample sizes) from control and ACD patients were significantly different in 6 of 8 comparisons, with the remaining two comparisons trending in the same direction: TRB and TRG repertoires of ACD IEC and LPC samples exhibited lower clonality than the repertoires of control samples, while there was a trend towards higher T cell density in ACD and RCD I samples compared with control and GFD samples. Interestingly, samples from subjects on a GFD had mean clonalities closer to control subjects than subjects with ACD (and to some extent RCD samples). The lower clonality seen in CD samples indicated that diseased tissue has a fairly broad and diverse TRB and TRG repertoire. The small number of PBMC samples did not demonstrate significant differences in clonality or T cell density between subjects (data not shown).

#### Identification of CD-associated TR sequences in IEC samples

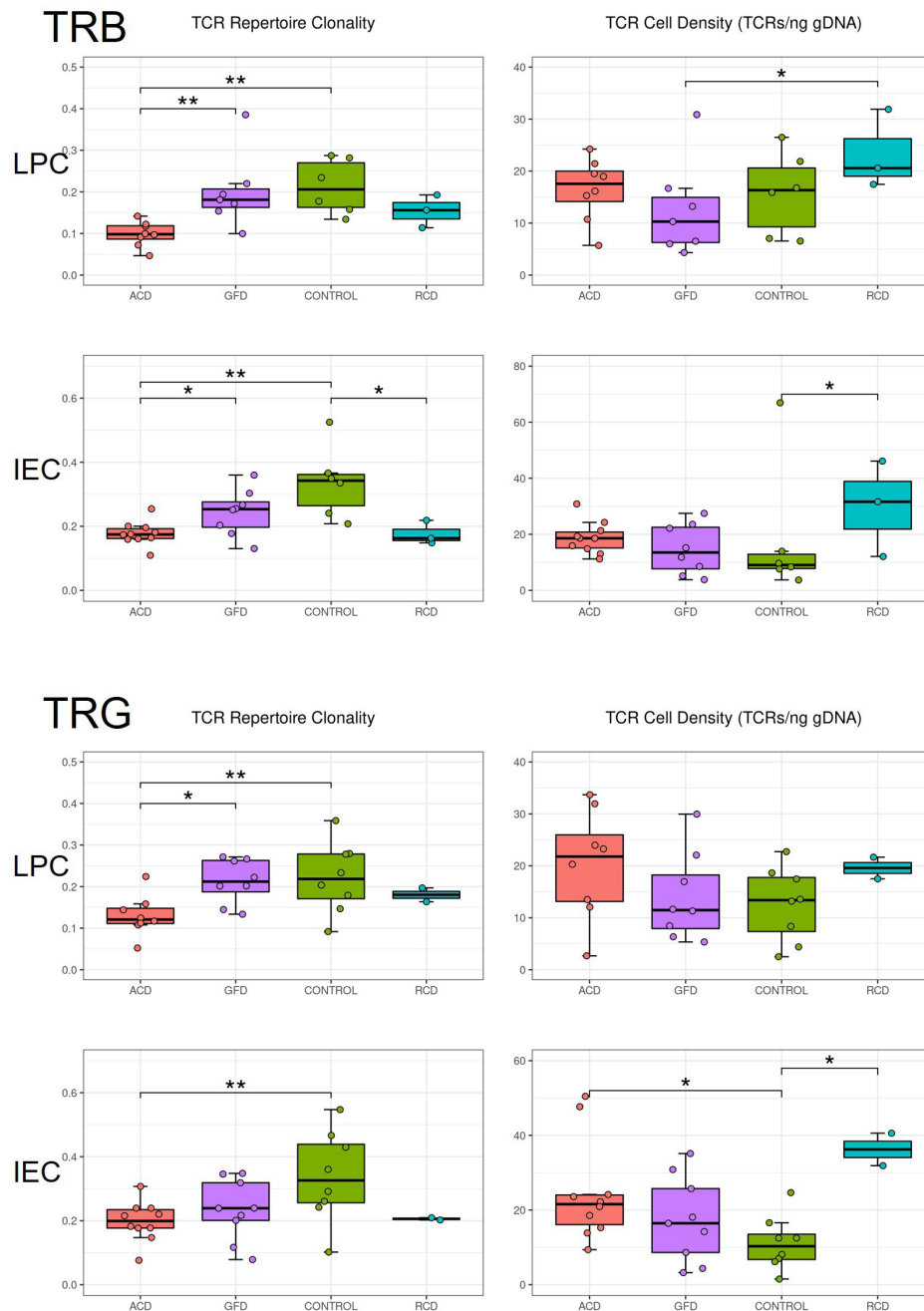
Public responses refer to identical, clonally dominant TRs shared by multiple individuals in a population<sup>26</sup> and in the context of broad disease-associated TR repertoires, some public TRs may be more commonly encountered. To evaluate whether there were specific TRs associated with CD in our study cohort, we searched for public TRs (defined as CDR3 AA sequences present in at least 80% of unsorted samples) and tested for increased abundance in the diseased samples. There were no TRB AA sequences shared by more than 10 samples (37% of samples). Of

114 527 TRB AA sequences across all samples, 112 072 (97.9%) were unique to an individual sample. This distribution was similar between CD and control samples. In addition, there was no statistically significant difference in the relative V-gene utilisation in IEC samples from ACD, RCD I and GFD patients, or controls (Kruskal-Wallis rank sum test  $p = 0.3122$ ). These results indicate that the TRB repertoires of both patients with CD and individuals without CD are diverse and largely private, with the frequency of potential public TRs too low to be identified by our analytical approach.

On the contrary, 48 TRG CDR3 AA sequences were detected in 80% or more of the samples sequenced (online table S4). As a first screen, we performed Fisher's exact test with a significance level of  $p < 0.05$  to identify sequences that were present more frequently in CD than in control repertoires, which revealed six different AA sequences. Among these, two sequences; CATWDGLNYYKKLF and CATWDRPEKLF, were significantly more abundant in CD samples versus controls as measured by a Mann-Whitney rank sum test with Bonferroni correction ( $p = 0.0114$  and  $0.0136$ , respectively), which incorporated both presence and clonal abundance. We also compared the abundance of these sequences in ACD and GFD samples, with no significant differences noted (CATWDGLNYYKKLF  $p = 0.6965$ , CATWDRPEKLF  $p = 0.6334$ ), indicating that the receptors with these AA motifs persist at comparable levels despite incorporation of GFD. When we compared the frequency distribution of these two TRG AA sequence motifs within the IEC samples from patients with CD, the shorter motif (CATWDRPEKLF), which was associated with the TRGV3 gene, had a significantly higher relative frequency distribution compared with CATWDGLNYYKKLF (**figure 2**).

#### Identification of CD-associated TR sequences in LPC samples

There were no TRB AA sequences shared by more than 12 samples. Of 158 877 TRB AA sequences across all samples, 154 447 (97.2%) were unique to an individual sample. There was no statistically significant difference in the relative V-gene utilisation across LPC samples from ACD, RCD I and GFD patients, or controls (Kruskal-Wallis rank sum test  $p = 0.9716$ ). For TRG, 153 sequences were detected in 80% or more of the samples. No TRs were significantly more abundant in the LPC repertoires of patients with CD compared with controls. The longer sequence (CATWDGLNYYKKLF) was not observed in the LPC fractions of patients with CD, however, CATWDRPEKLF, the other high abundance AA motif detected in CD IEC samples, was also observed in LPC fractions of patients with CD but not in controls (Bonferroni adjusted  $p = 0.063$ ). Since the method used to separate IECs from LPCs does not yield cellular fractions

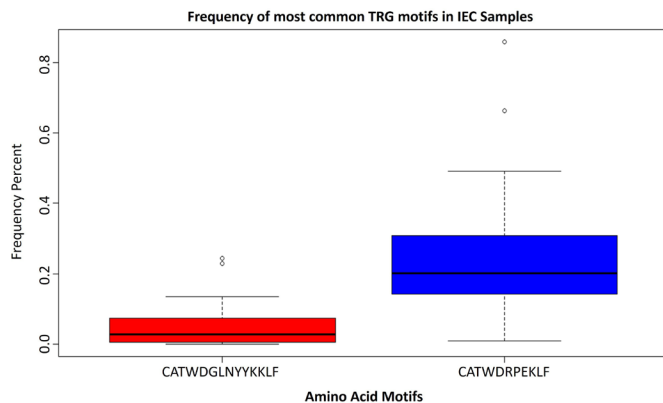


**Figure 1** TR repertoire metrics by tissue type, TR locus and disease status. TR repertoire clonality (measured from 0 to 1) and T cell density (estimated TCRs per ng gDNA input) are charted by disease type. According to TR locus (TRB or TRG) and tissue cells (LPC or IEC) for each unique sample. Data are plotted with median and quartiles; significance is calculated using a Kruskal-Wallis test with a Dunn post-test, and significance is shown with bars and asterisks: (\* $p \leq 0.05$ ; \*\* $p \leq 0.01$ ). Jitter has been applied to the images to resolve closely spaced data elements. ACD, active coeliac disease; GFD, gluten-free coeliac disease; IEC, intraepithelial cell; LPS, lamina propria cell; RCD I, refractory coeliac disease type I.

of high purity, the detection of this motif in LPC samples could be due to the presence of residual IECs.

Although the full spectrum of gluten responsive T cell clonotypes is not known, prior TR sequencing studies of HLA class II tetramer-gliadin peptide sorted CD4<sup>+</sup> T cells have identified public TR CDR3 sequences of HLA-DQ2.5 and DQ8 restricted gliadin-specific CD4<sup>+</sup> T cells in a significant proportion of small intestinal LP and peripheral blood samples from patients with CD.<sup>13 23</sup> To discern presence of gliadin-specific public TRB CDR3 sequences at a frequency below the stringent cut-off used in our study, we queried our bulk LPC (and PBMC) TRB CDR3 sequences against those generated in a recent large analysis of LP gliadin-specific CD4<sup>+</sup> T

cells isolated with HLA-DQ2.5-gliadin tetramers constituting four immunodominant epitopes.<sup>23</sup> Nine of 18 (50%) of our CD LPC samples (3/8 ACD, 4/7 GFD and 2/3 RCD I) with successful TRB sequencing harboured CDR3 sequences similar to the published sequences, with an average of 4.9 shared (public) sequences per sample (range 1–10 per sample). In contrast, only 1 of 6 control LPC samples possessed a shared sequence with the published gliadin-specific CDR3 sequences, with only a single copy of the sequence being detected in that sample (figure 3). Public gliadin responsive TRB sequences were also detected in 2 of 4 PBMC samples from patients with CD, although at a relatively lower clonal breadth and depth. Among the 218 published gliadin-specific sequences, 23



**Figure 2** Relative frequency distribution of the two most abundant IE TRG CDR3 AA motifs identified in CD. A boxplot of the relative frequency distribution of the AA motifs CATWDRPEKLF and CATWDGLNYYKKLF in IEC samples from patients with CD shows that the CATWDRPEKLF AA sequence is considerably more abundant than CATWDGLNYYKKLF (Wilcoxon  $p=2.2e-05$ ). AA, amino acid; CD, coeliac disease; IEC, intraepithelial cell.

were detected in our LPC and PBMC samples combined (details are provided in online table S3). Biased usage of specific TRBV genes by HLA-DQ2.5 and DQ8 restricted gliadin-specific  $CD4^+$  T cells has also been documented in CD.<sup>13</sup> Since high-resolution HLA typing information was not available for our samples, as a proof of principle, we decided to query biased use of the TRBV7-2 gene, which has been observed in HLA-DQ2.5 restricted T cells. The TRBV7-2 gene was identified in 10/19 (53%) of the public TRB sequences in the 9 CD LPC samples, while the TRBV05-01 gene was noted in

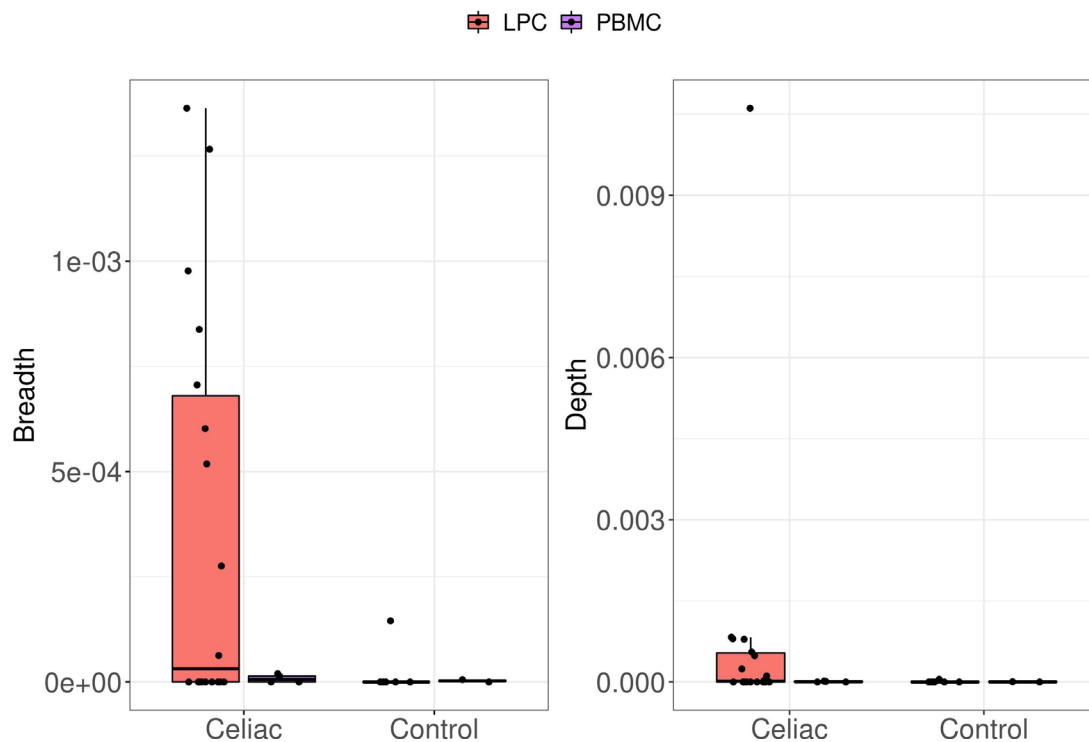
the only control sample with the single gliadin-specific public TRB sequence.

### Evaluation of sorted T cells from CD and control samples to determine the cellular source of TRG CDR3 motifs

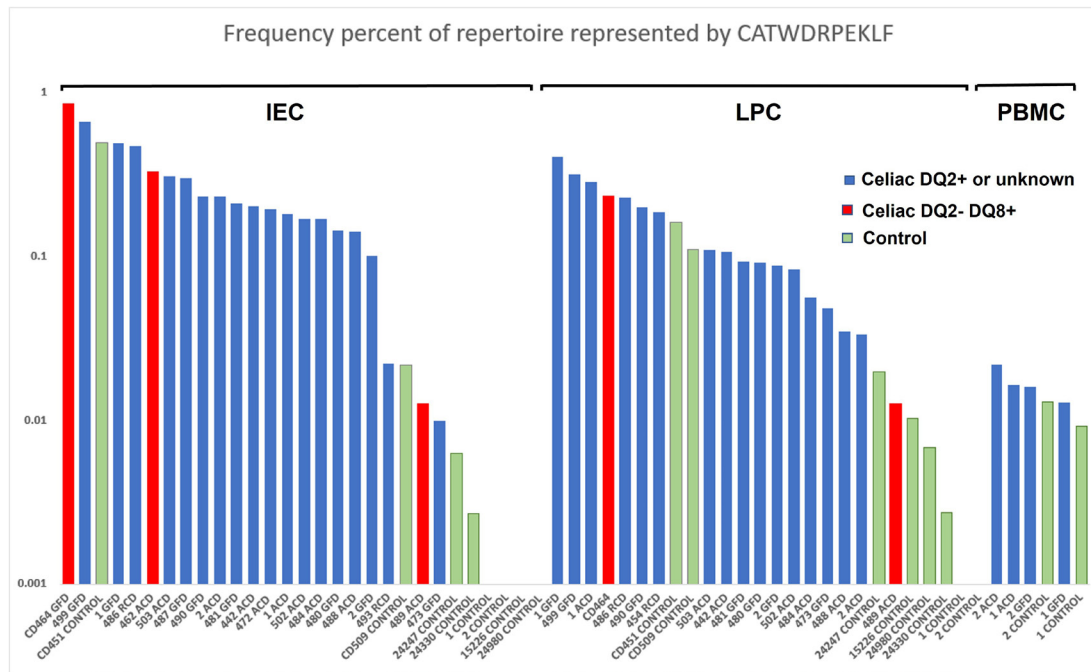
As highlighted in table 1 and online table S1, in addition to sequencing of bulk samples, a few sorted IE  $CD8^+$  TCR $\alpha\beta$  ( $n=8$ ), IE TCR $\gamma\delta$  ( $n=6$ ) and LP  $CD4^+$  TCR $\alpha\beta$  ( $n=6$ ) T cell samples were subjected to sequencing of the TRG locus, in order to confirm IE TCR $\gamma\delta$  T cells as the source of the TRG CDR3 sequences encoding the two AA motifs. The frequencies of the two motifs in unsorted samples were also compared with their frequencies in sorted samples. In these limited number of samples, the shorter motif (CATWDRPEKLF), in particular, was detected at a higher frequency in sorted IE TCR $\gamma\delta$  T cell samples compared with matched unsorted samples; 1.8–2.5 fold increase for ACD samples and 3.1–4.5 fold increase for the GFD samples (online table S5). Online figure S1 depicts the relative frequency count for each motif across the sorted samples analysed. Importantly, these sequences were not detected in sorted IE  $CD8^+$  TCR $\alpha\beta$  and LP  $CD4^+$  TCR $\alpha\beta$  T cells.

### The TRG CDR3 motif CATWDRPEKLF is enriched in CD mucosa

Since the disease-associated increase in abundance of the two TRG CDR3 AA motifs in CD IECs compared with controls could be due to an increase in IE TCR $\gamma\delta$  T cells that occurs in CD, we compared the frequency of the AA motifs in the blood and duodenal IECs/LPCs of patients with CD, using two ACD and two GFD donors for which matched samples were available. Intriguingly, among the four CD samples, the CATWDRPEKLF motif was on average enriched 12.1-fold (range 9.1–9.9 fold for ACD and 8.6–20.9 fold for GFD) in the IEC fractions ( $p<0.01$ ), and it was also enriched in



**Figure 3** Detection of common gliadin-specific TRB sequences in CD samples. Plots depict comparison of the breadth (left) and depth of the gliadin-specific TCR sequences found within the aggregated (ACD, GFD, RCD) CD and control LPC and PBMC samples. Each LPC and PBMC sample was queried against a list of 218 published sequences. Breadth was calculated as the number of unique, matching, productive rearrangements out of the total number of unique productive rearrangements, while depth accounts for the frequency of those rearrangements in the repertoire. ACD, active CD; CD, coeliac disease; GFD, gluten-free diet; LPC, lamina propria cell; PBMC, peripheral blood mononuclear cell.



**Figure 4** Abundance of CATWDRPEKLF in the study cohort. The frequency of the AA motif CATWDRPEKLF across all samples (logarithmic scale). The samples were separated according to origin; duodenal biopsy intraepithelial cells (IEC) or lamina propria cells (LPC) fractions or peripheral blood mononuclear cell (PBMC) samples. Patients with CD who were HLA DQ2<sup>+</sup> or of unknown HLA type were distinguished from the patients who were HLA DQ2<sup>neg</sup>DQ8<sup>+</sup>. Patients without CD were designated as controls. AA, amino acid; CD, coeliac disease.

the LPC samples. However, such differences were not observed for the longer AA motif (CATWDGLNYYKLF).

#### Distribution and HLA association of the putative CD-associated TRG CDR3 motif CATWDRPEKLF

At least 90% of patients with CD have the HLA-DQ2 risk allele, with the majority of the remainder carrying the HLA-DQ8 allele.<sup>27</sup> To assess any relationship between the patient HLA type and the putative CD-associated TRG CDR3 AA motif, we calculated the repertoire frequency of the motif in each sample and plotted the results in descending order of frequency based on sample types (IECs, LPCs and PBMCs). CD samples were grouped as HLA-DQ2<sup>neg</sup>DQ8<sup>+</sup>, HLA-DQ2<sup>+</sup>/HLA unknown and controls (figure 4). The receptor motif CATWDRPEKLF was relatively evenly distributed between the DQ2<sup>neg</sup>DQ8<sup>+</sup> and DQ2<sup>+</sup>/HLA unknown samples, consistent with absence of HLA-DQ2 or DQ8 restriction of the motif bearing TCR $\gamma\delta$  T cells.<sup>6</sup> Every sample with a CD diagnosis harboured this TR sequence motif, as did 12 of 18 control samples (inclusive of sorted TCR $\gamma\delta$  IEC samples). Abundance was high in most IEC and LPC samples from both HLA-DQ2<sup>+</sup>/HLA unknown and DQ2<sup>neg</sup>DQ8<sup>+</sup> CD patients and low in most control samples.

#### Evaluation of TR rearrangements constituting the TRG CDR3 motif CATWDRPEKLF

We next sought to determine if the abundant and tissue enriched TRG CDR3 AA motif CATWDRPEKLF in patients with CD was derived from unique clonal sequences in each sample or was the consequence of several TRG rearrangements with different sequences all leading to the same TRG CDR3 AA motif by convergent recombination. The constituent nucleotide sequences encoding the CDR3 AA motif CATWDRPEKLF in each sample were assessed. Table 3 shows the average number

of unique CDR3 nucleotide variants in each sample, by disease status and sample type. In contrast to control tissue samples, CD IEC and LPC samples had a mean average of ~23 separate nucleotide sequences (different TRG rearrangement events) encoding the TRG.CDR3 AA motif CATWDRPEKLF, demonstrating that the motif is recurrently produced during V-J recombination, and suggesting that several clonal lineages bearing the same TRG CDR3 sequence at the AA level are present in CD mucosa. As for PBMCs samples, an average of 17 distinct rearrangements (range 9–25) generated this motif in ACD, with no significant differences between CD subjects and controls overall. A small number of rearrangements common to both IEC and PBMC samples from the same subject were present at comparable levels to the same rearrangements between samples from unrelated subjects, indicative of a high generation probability for these rearrangements.

**Table 3** Unique nucleotide variants encoding CATWDRPEKLF in CD and control samples

CATWDRPEKLF	IEC		LPC		PBMC	
	Average	Range	Average	Range	Average	Range
ACD	21.7	5–42	14.1	1–33	17.0	9–25
GFD	20.9	2–62	19.6	8–46	11.5	11–12
RCD I	27	10–44	34	34		
Control	<1*	0–1	2	1–3	14.0	11–17

The mean and range of unique CDR3 nucleotide variants encoding the CDR3 region CATWDRPEKLF AA motif is shown for CD patient and control IEC, LPC and PBMC samples.

\*Absolute value in decimals not shown due to lack of biological meaning for values below 1.

ACD, active CD; CD, coeliac disease; GFD, gluten-free diet; IEC, intraepithelial cell; LPC, lamina propria cell; PBMC, peripheral blood mononuclear cell; RCD, refractory CD.

## DISCUSSION

In the current study, high throughput TR sequencing analysis of small intestinal IEC and LPC samples from CD patients, including those with ACD and RCD I as well as individuals on GFD, and non-CD controls, revealed significantly lower TRB and TRG clonality in CD compared to controls, indicative of increased diversity of the T cell repertoires in CD. Our findings differ from those of an earlier study that reported an oligoclonal IEL TRB repertoire in CD,<sup>14</sup> but corroborate the results of recent studies analysing small intestinal TR profiles of CD patients with different clinical presentations and utilising different assays, including PCR using the BIOMED II primers,<sup>28</sup> multiplex PCR followed by high throughput sequencing,<sup>29</sup> and single cell sequencing of TCR $\alpha\beta$  CD8<sup>+</sup> and TCR $\gamma\delta$  IELs.<sup>30–31</sup> Our sequencing and analytical approach demonstrated the small intestinal IE and LP TRB repertoires in CD to be unique overall, showing low overlap between samples and minimal publicity of the sequences, without differential abundance of any CDR3 AA motif in CD compared with controls. However, comparison of LP (and PBMC) TRB CDR3 sequences with published sequences derived from HLA-DQ2.5 gliadin tetramer-sorted LP CD4<sup>+</sup> T cells,<sup>23</sup> disclosed the presence of gliadin-specific public TRB sequence in 50% of our CD samples, at levels below the detection threshold employed in our study, and also biased TRBV7-2 gene usage. On the other hand, despite the increased diversity of the TRG repertoire in CD, we observed enrichment of certain TCR $\gamma\delta$  T cell derived TRG CDR3 sequence motifs in the small intestinal mucosa of CD patients, which has not been reported before.

As it is well established that the adaptive immune response to gluten-derived deamidated gliadin peptides is the prime driver of mucosal inflammation in CD, most of the studies to date have focused on characterising the repertoire, phenotype and functional attributes of gliadin-specific CD4<sup>+</sup> T cells within the small intestinal LP (and peripheral blood).<sup>13</sup> However, due to the rarity of gliadin responsive T cells in the LP of CD patients (<2% of T cells),<sup>32</sup> soluble HLA-DQ multimers bound with gliadin peptides and other enrichment strategies have been utilised to isolate CD4<sup>+</sup> T cells that recognise immunodominant epitopes of gliadin peptides presented by HLA-DQ2 and DQ8 molecules.<sup>33–37</sup> The use of pooled HLA-DQ2 gliadin tetramers has helped delineate the breadth of gluten-specific T cell responses in CD.<sup>23 38 39</sup> These studies have revealed unique characteristics of the gluten responsive CD4<sup>+</sup> TR repertoires in CD, including occurrence of gliadin-specific public TRs in a high proportion of cases (94%), although with limited overlap of the TRs (2–23 shared sequences) among patients, biased or preferential use of certain TRAV, TRBV and TRAV/TRBV pairs, and conserved CDR3 motifs. Our observation of a diverse, polyclonal TRB repertoire in the small intestinal LP in ACD and RCD I suggests similarities between these two disease states, and indicates a preponderance of gluten non-specific T cells in the inflamed LP that traffic to this site because of the altered cytokine milieu and impaired intestinal barrier function engendered by ongoing adaptive and innate mucosal immune responses. Dampening of inflammation on commencement of GFD leads to a shift in the T cell repertoire towards the normal state with increased TRB clonality.<sup>40</sup> The prevalence of gliadin responsive CD4<sup>+</sup> T cells has also been shown to be higher in the inflamed CD mucosa, which decrease on removal of gluten from the diet, but persist at low levels for long periods.<sup>39</sup> Since the goal of our study was to detect high abundance disease-specific public sequences that could be exploited for diagnostic purposes,

our analysis pipeline was unable to identify low frequency gliadin-specific public TRB sequences in an unbiased manner.

Dysregulated IL-15 expression by epithelial cells and LP macrophages incited by toxic gliadin peptides and cytokines elaborated by gliadin responsive LP CD4<sup>+</sup> T cells (eg, IL-2, IL-21, TNF $\alpha$ ) have been proposed to activate innate CD8<sup>+</sup> TCR $\alpha\beta$  IELs that kill epithelial cells, which have upregulated 'stress-induced' ligands, in a TCR-independent fashion.<sup>41–43</sup> Although clonal expansions of circulating gut homing CD8<sup>+</sup> TCR $\alpha\beta$  T cells with preferential usage of specific TRBV genes was reported after gluten challenge in CD patients,<sup>44</sup> the observation of a more diverse IE TRB repertoire by us, and CD8<sup>+</sup> TRA/B repertoire by others,<sup>31</sup> in ACD compared with GFD and control samples, and the absence of TRBV gene bias, supports the polyclonal nature of the IE CD8<sup>+</sup>  $\alpha\beta$  T cell immune response in CD. This was also true for RCD I, a subtype of RCD that is characterised by a polyclonal expansion of IE CD8<sup>+</sup> TCR $\alpha\beta$  T cells, as opposed to RCD II, which represents a clonal lymphoproliferative disorder of IELs manifesting an aberrant immunophenotype.<sup>45</sup> The detection of lower clonality and higher T cell density in RCD I, ascertained by TRB (and TRG) sequencing, indicates persistence of an inflammatory process that is similar in nature to ACD, in line with previous findings.<sup>29</sup>

TCR $\gamma\delta$  IELs are elevated in CD, but the roles of these lymphocytes in disease pathogenesis remains unclear. Besides an immunoregulatory role proposed for a subset,<sup>4</sup> they likely also subserve other functions. Mayassi *et al* reported loss of naturally occurring V $\gamma$ 4<sup>+</sup>/V $\delta$ 1<sup>+</sup> TCR $\gamma\delta$  IELs in ACD that have cytotoxic properties and recognise epithelial butyrophilin-like (BTNL) molecules (BTNL3/BTNL8), reduced epithelial BTNL8 expression and a concomitant expansion of gluten sensitive IFN- $\gamma$  producing TCR $\gamma\delta$  V $\delta$ 1<sup>+</sup> IELs whose ligands are unknown.<sup>46</sup> An expansion of circulating TCR $\gamma\delta$  T cells, with preferential usage of the TRDV1 gene has also been described in CD patients after gluten challenge.<sup>44</sup> In contrast, recent single cell sequencing analysis of TCR $\gamma\delta$  IELs and gut-homing TCR $\gamma\delta$  T cells in the peripheral blood by Eggesbø *et al* demonstrated biased usage of TRGV4 and TRDV1 genes in healthy individuals, but diverse TCR $\gamma\delta$  T cell repertoires in samples from ACD patients and those on GFD.<sup>30</sup> We also observed increased diversity of the IE TRG repertoires in CD, however higher abundance of a couple of public TRG CDR3 AA sequences was noted in samples from CD patients, including those on GFD, compared with controls. These sequences were confirmed to be derived from TCR $\gamma\delta$  T cells and not from companion TRG rearrangements in TCR $\alpha\beta$  T cells. One of the CDR3 AA motifs, CATWDRPEKLF, associated with the TRGV3 gene, was enriched in IECs compared with matched PBMC samples. Evaluation of the CDR3 nucleotide sequences revealed that the AA sequence was generated from multiple nucleotide variants, suggesting convergent evolution. Interestingly, many of the gliadin-specific public TRB CDR3 AA sequences reported previously also resulted from convergent recombination.<sup>39</sup>

Although Eggesbø *et al* detected several public TRG CDR3 AA sequences in small intestinal TCR $\gamma\delta$  IELs of CD patients, they did not identify significant enrichment of any sequence in CD samples.<sup>30</sup> It is unclear if this was due to a lower number of productive TRG sequences assessed by the authors. Of interest, using bulk sequencing of DNA extracted from duodenal biopsies and machine learning analysis, Foers *et al*, recently reported CD-associated changes in TRG and TRD repertoires, which persisted even after cessation of gluten consumption.<sup>15</sup> It is possible that restricted diversity of the TRG repertoire enabled

detection of the TRG AA motifs in CD samples. The significance of enrichment of CATWDRPEKLF motif-bearing TCR $\gamma\delta$  T cells in the small intestinal epithelium of CD patients, however, remains to be determined. We speculate that TCR $\gamma\delta$  T cells bearing this TR motif traffic to and accumulate (or proliferate) in the small intestinal epithelium in response to epithelial damage or stress induced by toxic gliadin peptides, since complete exclusion of dietary gluten can be difficult for CD patients. Paired TRG and TRD sequencing might enable better characterisation of such public TR bearing TCR $\gamma\delta$  T cells in CD and help ascertain similarities or differences between our findings and those published previously. If confirmed by others, the identified TRG CDR3 motif could serve as a novel biomarker of CD.

A limitation of the methodology used in this study is that it is unsuitable for detecting gliadin-specific T cells in an unbiased manner due to: scarcity of these T cells in the small intestinal mucosa (and peripheral blood), limited sharing of public TR sequences among patients, and incomplete knowledge of the breadth of gliadin epitopes in CD.

In summary, NGS-based analysis of the TR repertoires of small intestinal IEC and LPC samples revealed reduced clonality of TRB and TRG repertoires in CD, suggesting recruitment of diverse populations of TCR $\alpha\beta$  and TCR $\gamma\delta$  T cells to the inflamed small intestinal mucosa. An increased abundance of IE TCR $\gamma\delta$  T cells bearing specific public TRG CDR3 sequences was also noted in CD. Future studies, utilising single cell TR sequencing coupled with immunophenotypic and transcriptional analyses,<sup>47</sup> are required to elucidate the biologic attributes of these and other mucosal T cell populations in CD.

**Correction notice** This article has been corrected since it was first published. The open access licence has been updated to CC BY.

**Handling editor** Runjan Chetty.

**Acknowledgements** We thank Krishna Patel and Kristin MacIntosh for project management and editorial assistance.

**Contributors** GB conceived the study, acquired and analysed data and wrote the manuscript; PHRG provided clinically annotated patient samples and critically reviewed the manuscript; ASH contributed to study design refinement and critically reviewed the manuscript; AB and JV performed experiments including cell sorting and isolation and sample preparation and critically reviewed the manuscript; LWL, SS, BC, ROE, DW and IMK planned the experiments, analyzed data and assisted with writing and revising the manuscript; IK assisted with experimental design, data analysis and critically reviewed the manuscript. SS accepts full responsibility for the work and/or the conduct of the study, had access to the data, and controlled the decision to publish.

**Funding** This project was funded by Adaptive Biotechnologies.

**Competing interests** GB, PHRG, ASH, AB and JV declare no competing interests. LWL, SS, BC, ROE, DW, IMK and IK declare employment and equity ownership with Adaptive Biotechnologies.

**Patient consent for publication** Not applicable.

**Ethics approval** The study was performed in adherence to the principles of the Declaration of Helsinki, using a protocol (AAAB2472) approved by the Institutional Review Board of Columbia University and the written informed consent of patients for use of tissue samples for research.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data are available in a public, open access repository.

**Supplemental material** This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution 4.0 Unported (CC BY 4.0) license, which permits others to copy, redistribute, remix, transform and build upon this work for any purpose, provided the original work is properly cited, a link to the licence is given, and indication of whether changes were made. See: <https://creativecommons.org/licenses/by/4.0/>.

#### ORCID iDs

Shahin Shafiani <http://orcid.org/0000-0003-4641-5458>

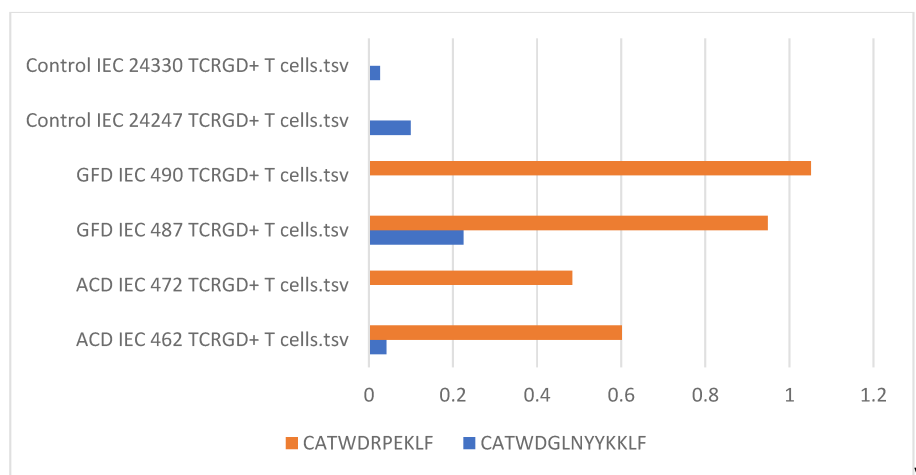
Govind Bhagat <http://orcid.org/0000-0001-6250-048X>

#### REFERENCES

- Green PHR, Jabri B. Celiac disease. *Annu Rev Med* 2006;57:207–21.
- Kim SM, Mayassi T, Jabri B. Innate immunity: actuating the gears of celiac disease pathogenesis. *Best Pract Res Clin Gastroenterol* 2015;29:425–35.
- Steenholt JV, Nielsen C, Baudewijn L, et al. The composition of T cell subtypes in duodenal biopsies are altered in coeliac disease patients. *PLoS One* 2017;12:e0170270.
- Bhagat G, Naiyer AJ, Shah JG, et al. Small intestinal CD8+TCR $\gamma\delta$ +NKG2A+ intraepithelial lymphocytes have attributes of regulatory cells in patients with celiac disease. *J Clin Invest* 2008;118:281–93.
- Halstensen TS, Scott H, Brandtzaeg P. Intraepithelial T cells of the TcR gamma/delta+ CD8- and V delta 1/J delta 1+ phenotypes are increased in coeliac disease. *Scand J Immunol* 1989;30:665–72.
- Spencer J, Isaacson PG, MacDonald TT, et al. Gamma/Delta T cells and the diagnosis of coeliac disease. *Clin Exp Immunol* 1991;85:109–13.
- Rust C, Kooy Y, Peña S, et al. Phenotypical and functional characterization of small intestinal TCR gamma delta + T cells in coeliac disease. *Scand J Immunol* 1992;35:459–68.
- Savilahti E, Arato A, Verkasalo M. Intestinal  $\gamma\delta$  receptor-bearing T lymphocytes in celiac disease and inflammatory bowel diseases in children. Constant increase in celiac disease. *Pediatr Res* 1990;28:579–81.
- Mäki M, Holm K, Collin P, et al. Increase in gamma/delta T cell receptor bearing lymphocytes in normal small bowel mucosa in latent coeliac disease. *Gut* 1991;32:1412–4.
- Rosati E, Dowds CM, Liaskou E, et al. Overview of methodologies for T-cell receptor repertoire analysis. *BMC Biotechnol* 2017;17:61.
- Bradley P, Thomas PG. Using T cell receptor repertoires to understand the principles of adaptive immune recognition. *Annu Rev Immunol* 2019;37:547–70.
- Maini MK, Casorati G, Dellabona P, et al. T-Cell clonality in immune responses. *Immunol Today* 1999;20:262–6.
- Lindeman I, Sollid LM. Single-Cell approaches to dissect adaptive immune responses involved in autoimmunity: the case of celiac disease. *Mucosal Immunol* 2022;15:51–63.
- Balas A, Garcia-Novo MD, Martinez J, et al. Intestinal alphabeta T cells of symptomatic celiac disease patients show oligoclonal TCRBV repertoire but polyclonal rearrangement patterns. *Hum Immunol* 2000;61:247–54.
- Foers AD, Shoukat MS, Welsh OE, et al. Classification of intestinal T-cell receptor repertoires using machine learning methods can identify patients with coeliac disease regardless of dietary gluten status. *J Pathol* 2021;253:279–91.
- Carlson CS, Emerson RO, Sherwood AM, et al. Using synthetic templates to design an unbiased multiplex PCR assay. *Nat Commun* 2013;4:2680.
- Robins HS, Campregher PV, Srivastava SK, et al. Comprehensive assessment of T-cell receptor beta-chain diversity in alphabeta T cells. *Blood* 2009;114:4099–107.
- Meresse B, Curran SA, Ciszewski C, et al. Reprogramming of CTLs into natural killer-like cells in celiac disease. *J Exp Med* 2006;203:1343–55.
- Meresse B, Chen Z, Ciszewski C, et al. Coordinated induction by IL15 of a TCR-independent NKG2D signaling pathway converts CTL into lymphokine-activated killer cells in celiac disease. *Immunity* 2004;21:357–66.
- Lee LW, Shafiani S, Crossley B, et al. Data from: Characterization of T cell receptor repertoires in celiac disease. immuneACCESS repository, November 18, 2021. Available: <https://doi.org/10.21417/LWL2022JCP>
- Sherwood AM, Emerson RO, Scherer D, et al. Tumor-Infiltrating lymphocytes in colorectal tumors display a diversity of T cell receptor sequences that differ from the T cells in adjacent mucosal tissue. *Cancer Immunol Immunother* 2013;62:1453–61.
- Emerson RO, DeWitt WS, Vignali M, et al. Immunosequencing identifies signatures of cytomegalovirus exposure history and HLA-mediated effects on the T cell repertoire. *Nat Genet* 2017;49:659–65.
- Dahal-Koirala S, Risnes LF, Neumann RS, et al. Comprehensive analysis of CDR3 sequences in Gluten-Specific T-cell receptors reveals a dominant R-Motif and several new minor motifs. *Front Immunol* 2021;12:639672.
- Shannon CE. A mathematical theory of communication. *Bell System Technical J* 1948;27:379–423.
- Wu D, Emerson RO, Sherwood A, et al. Detection of minimal residual disease in B lymphoblastic leukemia by high-throughput sequencing of IGH. *Clin Cancer Res* 2014;20:4540–8.



- 26 Venturi V, Kedzierska K, Price DA, *et al.* Sharing of T cell receptors in antigen-specific responses is driven by convergent recombination. *Proc Natl Acad Sci U S A* 2006;103:18691–6.
- 27 Lionetti E, Catassi C. Co-Localization of gluten consumption and HLA-DQ2 and -DQ8 genotypes, a clue to the history of celiac disease. *Dig Liver Dis* 2014;46:1057–63.
- 28 Hussein S, Gindin T, Lagana SM, *et al.* Clonal T cell receptor gene rearrangements in celiac disease: implications for diagnosing refractory coeliac disease. *J Clin Pathol* 2018;71:825–31.
- 29 Ritter J, Zimmermann K, Jöhrens K, *et al.* T-Cell repertoires in refractory coeliac disease. *Gut* 2018;67:644–53.
- 30 Eggesbø LM, Risnes LF, Neumann RS, *et al.* Single-Cell TCR sequencing of gut intraepithelial  $\gamma\delta$  T cells reveals a vast and diverse repertoire in celiac disease. *Mucosal Immunol* 2020;13:313–21.
- 31 Eggesbø LM, Risnes LF, Neumann RS, *et al.* Single-cell TCR repertoire analysis reveals highly polyclonal composition of human intraepithelial CD8<sup>+</sup>  $\alpha\beta$  T lymphocytes in untreated celiac disease. *Eur J Immunol* 2021;51:1542–5.
- 32 Qiao S-W, Dahal-Koirala S, Eggesbø LM, *et al.* Frequency of Gluten-Reactive T cells in active celiac lesions estimated by direct cell cloning. *Front Immunol* 2021;12:646163.
- 33 Dahal-Koirala S, Ciacchi L, Petersen J, *et al.* Discriminative T-cell receptor recognition of highly homologous HLA-DQ2-bound gluten epitopes. *J Biol Chem* 2019;294:941–52.
- 34 Ráki M, Fallang L-E, Brottveit M, *et al.* Tetramer visualization of gut-homing gluten-specific T cells in the peripheral blood of celiac disease patients. *Proc Natl Acad Sci U S A* 2007;104:2831–6.
- 35 Dahal-Koirala S, Risnes LF, Christophersen A, *et al.* TCR sequencing of single cells reactive to DQ2.5-glia- $\alpha$ 2 and DQ2.5-glia- $\omega$ 2 reveals clonal expansion and epitope-specific V-gene usage. *Mucosal Immunol* 2016;9:587–96.
- 36 Petersen J, van Bergen J, Loh KL, *et al.* Determinants of gliadin-specific T cell selection in celiac disease. *J Immunol* 2015;194:6112–22.
- 37 Broughton SE, Petersen J, Theodosis A, *et al.* Biased T cell receptor usage directed against human leukocyte antigen DQ8-restricted gliadin peptides is associated with celiac disease. *Immunity* 2012;37:611–21.
- 38 Sarna VK, Lundin KEA, Mørkrid L, *et al.* HLA-DQ-Gluten tetramer blood test accurately identifies patients with and without celiac disease in absence of gluten consumption. *Gastroenterology* 2018;154:886–96.
- 39 Risnes LF, Christophersen A, Dahal-Koirala S, *et al.* Disease-driving CD4<sup>+</sup> T cell clonotypes persist for decades in celiac disease. *J Clin Invest* 2018;128:2642–50.
- 40 Probert CSJ, Saubermann LJ, Balk S, *et al.* Repertoire of the alpha beta T-cell receptor in the intestine. *Immunol Rev* 2007;215:215–25.
- 41 Jabri B, Sollid LM. T cells in celiac disease. *J Immunol* 2017;198:3005–14.
- 42 Meresse B, Korneychuk N, Malamut G, *et al.* Interleukin-15, a master piece in the immunological jigsaw of celiac disease. *Dig Dis* 2015;33:122–30.
- 43 Kooy-Winkelaar YMC, Bouwer D, Janssen GMC, *et al.* CD4 T-cell cytokines synergize to induce proliferation of malignant and nonmalignant innate intraepithelial lymphocytes. *Proc Natl Acad Sci U S A* 2017;114:E980–9.
- 44 Han A, Newell EW, Glanville J, *et al.* Dietary gluten triggers concomitant activation of CD4<sup>+</sup> and CD8<sup>+</sup>  $\alpha\beta$  T cells and  $\gamma\delta$  T cells in celiac disease. *Proc Natl Acad Sci U S A* 2013;110:13073–8.
- 45 Soderquist CR, Bhagat G. Cellular and molecular bases of refractory celiac disease. *Int Rev Cell Mol Biol* 2021;358:207–40.
- 46 Mayassi T, Ladell K, Gudjonson H, *et al.* Chronic inflammation permanently reshapes tissue-resident immunity in celiac disease. *Cell* 2019;176:967–81.
- 47 Schattgen SA, Guion K, Crawford JC, *et al.* Integrating T cell receptor sequences and transcriptional profiles by clonotype neighbor graph analysis (CoNGA). *Nat Biotechnol* 2022;40:54–63.



**Figure S1. Relative frequency count of the two most significant TRG AA motifs within the sorted IEC TCR $\gamma\delta$  T cells of CD patients and controls.** Bar plot of the relative frequency count (sum across sequences belonging to a sample) of the AA motifs CATWDRPEKLF and CATWDGLNYYKKLF in IEC samples from CD patients shows the higher abundance of CATWDRPEKLF AA sequence and its absence in control samples.

## SUPPLEMENTAL METHODS

### Cell isolation

Fresh biopsy specimens were put in RPMI solution containing 1mM eEGTA (VWR), 1.5 mM MgCl<sub>2</sub> (Sigma Aldrich, St. Louis, USA) and 1% FBS (Mediatech, Virginia, USA) to isolate IECs. After stirring for one hour at 37°C, cell suspensions were pelleted from the supernatant and washed two times in culture media containing RPMI (Mediatech) supplemented with 10% FBS, 10,000 IU penicillin and 10,000 µg/mL streptomycin. The remaining tissue sample was incubated in collagenase to release the LPCs.

For PBMC isolation, equal amounts of PBS and blood (7-8 mL/sample) were added to a 50 mL tube and mixed, following which the mixture was layered over an equal volume of Ficoll-Hypaque (MilliporeSigma) using an automatic pipette and centrifuged at 1200 rpm for 20 minutes at 4°C. The interface between the plasma and PBS/Ficoll-Hypaque was transferred to a new 15 mL tube and centrifuged at 1300 rpm for 10 minutes. The pellet was reconstituted in 0.5 mL of FBS. After counting the cells, aliquots of 1 million cells were suspended in DMSO/FBS and stored in liquid nitrogen.

### CD patient categorization

CD patients were separated into categories based on disease status: newly diagnosed or active celiac disease (ACD), refractory celiac disease type 1 (RCD I, no response to the gluten-free diet for >12 months, with the IECs manifesting a normal immunophenotype) and CD controlled by a gluten-free diet (GFD). All 30 patients within the study were subjected to endoscopy and small intestinal biopsy. The clinical, histopathologic and laboratory data of all patients are provided in Table S1.

**Cell-sorting strategy**

The IECs and LPCs were stained with CD45 A700, CD3 PE, CD4 APC-Cy7, CD8 FITC, TCR $\beta$  PE-Cy7 and TCR $\gamma$  BV421. After gating on CD3/CD45 double positive events, a bivariate plot was created looking at TCR $\beta$  and TCR $\gamma$  expression in IEC samples. TCR $\gamma\delta$  T cells were sorted by gating on the TCR $\gamma^+$  TCR $\beta^-$  population. To identify TCR $\alpha\beta$  IEC and LPC populations of interest, after gating on TCR $\beta^+$  events, a bivariate plot was created for CD8 and CD4 expression, and TCR $\beta$ -gated CD8 $^+$  and CD4 $^+$  T cells were sorted from IEC and LPC samples, respectively.

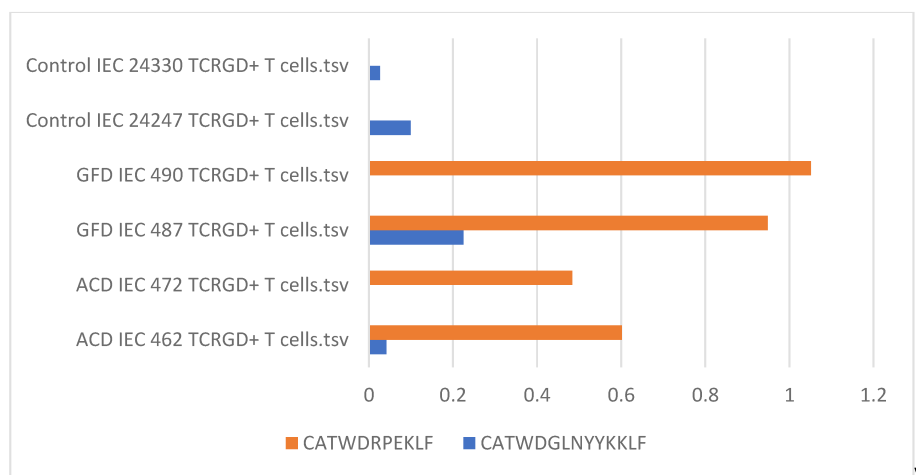
**TR sequencing approach**

Briefly, a multiplexed PCR method that uses a mixture of forward primers specific to V gene segments and reverse primers specific to J gene segments was employed. Sequencing of amplified CDR3 regions was performed using the Illumina HiSeq System (Illumina Inc., San Diego, USA). Raw HiSeq sequence data were preprocessed to remove errors in the primary sequence of each read. To remove both PCR and sequencing errors, a nearest neighbor algorithm was used to collapse the data into unique sequences by merging closely related sequences<sup>16,17</sup>.

Subject ID	Disease status	Sample type	T cell lineage sorted	Number of events/cells sorted
24330	Control	IEC	CD8 TCR $\alpha\beta$	100,000
24247	Control	IEC	CD8 TCR $\alpha\beta$	3683
488	ACD	IEC	CD8 TCR $\alpha\beta$	26523
489	ACD	IEC	CD8 TCR $\alpha\beta$	12646
472	ACD	IEC	CD8 TCR $\alpha\beta$	143000
462	ACD	IEC	CD8 TCR $\alpha\beta$	167000
490	GFD	IEC	CD8 TCR $\alpha\beta$	9575
487	GFD	IEC	CD8 TCR $\alpha\beta$	168000
Subject ID	Disease status	Sample type	T cell lineage sorted	Number of events/cells sorted
24330	Control	LPC	CD4 TCR $\alpha\beta$	204036
24247	Control	LPC	CD4 TCR $\alpha\beta$	9836
488	ACD	LPC	CD4 TCR $\alpha\beta$	5693
489	ACD	LPC	CD4 TCR $\alpha\beta$	8058
490	GFD	LPC	CD4 TCR $\alpha\beta$	2452
473	GFD	LPC	CD4 TCR $\alpha\beta$	11043
Subject ID	Disease status	Sample type	T cell lineage sorted	Number of events/cells sorted
24330	Control	IEC	TCR $\gamma\delta$	78294
24247	Control	IEC	TCR $\gamma\delta$	133000
472	ACD	IEC	TCR $\gamma\delta$	3400
462	ACD	IEC	TCR $\gamma\delta$	176000
487	GFD	IEC	TCR $\gamma\delta$	177000
490	GFD	IEC	TCR $\gamma\delta$	6600

Sample ID	CATWDRPEKLF Frequency
ACD_IE_462_TCRGD__T_CELLS	0.601
ACD_IE_462_UNSORTED_IEs	0.329
ACD_IE_472_TCRGD__T_CELLS	0.484
ACD_IE_472_UNSORTED_IEs	0.194
GFD_IE_487_TCRGD__T_CELLS	0.948
GFD_IE_487_UNSORTED_IEs	0.301
GFD_IE_490_TCRGD__T_CELLS	1.051
GFD_IE_490_UNSORTED_IEs	0.232
Control_IE_24247_TCRGD__T_CELLS	0.000
Control_IE_24247_UNSORTED_IEs	0.003
Control_IE_24330_TCRGD__T_CELLS	0.000
Control_IE_24330_UNSORTED_IEs	0.012

Sample ID	CATWDGLNYYKKLF Frequency
ACD_IE_462_TCRGD__T_CELLS	0.042
ACD_IE_462_UNSORTED_IEC	0.003
ACD_IE_472_TCRGD__T_CELLS	0.000
ACD_IE_472_UNSORTED_IEC	0.012
GFD_IE_487_TCRGD__T_CELLS	0.225
GFD_IE_487_UNSORTED_IEC	0.228
GFD_IE_490_TCRGD__T_CELLS	0.000
GFD_IE_490_UNSORTED_IEC	0.021
Control_IE_24247_TCRGD__T_CELLS	0.099
Control_IE_24247_UNSORTED_IEC	0.024
Control_IE_24330_TCRGD__T_CELLS	0.026
Control_IE_24330_UNSORTED_IEC	0.012



**Figure S1. Relative frequency count of the two most significant TRG AA motifs within the sorted IEC TCR $\gamma\delta$  T cells of CD patients and controls.** Bar plot of the relative frequency count (sum across sequences belonging to a sample) of the AA motifs CATWDRPEKLF and CATWDGLNYYKKLF in IEC samples from CD patients shows the higher abundance of CATWDRPEKLF AA sequence and its absence in control samples.

## SUPPLEMENTAL METHODS

### Cell isolation

Fresh biopsy specimens were put in RPMI solution containing 1mM eEGTA (VWR), 1.5 mM MgCl<sub>2</sub> (Sigma Aldrich, St. Louis, USA) and 1% FBS (Mediatech, Virginia, USA) to isolate IECs. After stirring for one hour at 37°C, cell suspensions were pelleted from the supernatant and washed two times in culture media containing RPMI (Mediatech) supplemented with 10% FBS, 10,000 IU penicillin and 10,000 µg/mL streptomycin. The remaining tissue sample was incubated in collagenase to release the LPCs.

For PBMC isolation, equal amounts of PBS and blood (7-8 mL/sample) were added to a 50 mL tube and mixed, following which the mixture was layered over an equal volume of Ficoll-Hypaque (MilliporeSigma) using an automatic pipette and centrifuged at 1200 rpm for 20 minutes at 4°C. The interface between the plasma and PBS/Ficoll-Hypaque was transferred to a new 15 mL tube and centrifuged at 1300 rpm for 10 minutes. The pellet was reconstituted in 0.5 mL of FBS. After counting the cells, aliquots of 1 million cells were suspended in DMSO/FBS and stored in liquid nitrogen.

### CD patient categorization

CD patients were separated into categories based on disease status: newly diagnosed or active celiac disease (ACD), refractory celiac disease type 1 (RCD I, no response to the gluten-free diet for >12 months, with the IECs manifesting a normal immunophenotype) and CD controlled by a gluten-free diet (GFD). All 30 patients within the study were subjected to endoscopy and small intestinal biopsy. The clinical, histopathologic and laboratory data of all patients are provided in Table S1.

**Cell-sorting strategy**

The IECs and LPCs were stained with CD45 A700, CD3 PE, CD4 APC-Cy7, CD8 FITC, TCR $\beta$  PE-Cy7 and TCR $\gamma$  BV421. After gating on CD3/CD45 double positive events, a bivariate plot was created looking at TCR $\beta$  and TCR $\gamma$  expression in IEC samples. TCR $\gamma\delta$  T cells were sorted by gating on the TCR $\gamma^+$  TCR $\beta^-$  population. To identify TCR $\alpha\beta$  IEC and LPC populations of interest, after gating on TCR $\beta^+$  events, a bivariate plot was created for CD8 and CD4 expression, and TCR $\beta$ -gated CD8 $^+$  and CD4 $^+$  T cells were sorted from IEC and LPC samples, respectively.

**TR sequencing approach**

Briefly, a multiplexed PCR method that uses a mixture of forward primers specific to V gene segments and reverse primers specific to J gene segments was employed. Sequencing of amplified CDR3 regions was performed using the Illumina HiSeq System (Illumina Inc., San Diego, USA). Raw HiSeq sequence data were preprocessed to remove errors in the primary sequence of each read. To remove both PCR and sequencing errors, a nearest neighbor algorithm was used to collapse the data into unique sequences by merging closely related sequences<sup>16,17</sup>.

Subject ID	Disease status	Sample type	T cell lineage sorted	Number of events/cells sorted
24330	Control	IEC	CD8 TCR $\alpha\beta$	100,000
24247	Control	IEC	CD8 TCR $\alpha\beta$	3683
488	ACD	IEC	CD8 TCR $\alpha\beta$	26523
489	ACD	IEC	CD8 TCR $\alpha\beta$	12646
472	ACD	IEC	CD8 TCR $\alpha\beta$	143000
462	ACD	IEC	CD8 TCR $\alpha\beta$	167000
490	GFD	IEC	CD8 TCR $\alpha\beta$	9575
487	GFD	IEC	CD8 TCR $\alpha\beta$	168000
Subject ID	Disease status	Sample type	T cell lineage sorted	Number of events/cells sorted
24330	Control	LPC	CD4 TCR $\alpha\beta$	204036
24247	Control	LPC	CD4 TCR $\alpha\beta$	9836
488	ACD	LPC	CD4 TCR $\alpha\beta$	5693
489	ACD	LPC	CD4 TCR $\alpha\beta$	8058
490	GFD	LPC	CD4 TCR $\alpha\beta$	2452
473	GFD	LPC	CD4 TCR $\alpha\beta$	11043
Subject ID	Disease status	Sample type	T cell lineage sorted	Number of events/cells sorted
24330	Control	IEC	TCR $\gamma\delta$	78294
24247	Control	IEC	TCR $\gamma\delta$	133000
472	ACD	IEC	TCR $\gamma\delta$	3400
462	ACD	IEC	TCR $\gamma\delta$	176000
487	GFD	IEC	TCR $\gamma\delta$	177000
490	GFD	IEC	TCR $\gamma\delta$	6600



Sample ID	CATWDRPEKLF Frequency
ACD_IE_462_TCRGD__T_CELLS	0.601
ACD_IE_462_UNSORTED_IEs	0.329
ACD_IE_472_TCRGD__T_CELLS	0.484
ACD_IE_472_UNSORTED_IEs	0.194
GFD_IE_487_TCRGD__T_CELLS	0.948
GFD_IE_487_UNSORTED_IEs	0.301
GFD_IE_490_TCRGD__T_CELLS	1.051
GFD_IE_490_UNSORTED_IEs	0.232
Control_IE_24247_TCRGD__T_CELLS	0.000
Control_IE_24247_UNSORTED_IEs	0.003
Control_IE_24330_TCRGD__T_CELLS	0.000
Control_IE_24330_UNSORTED_IEs	0.012

Sample ID	CATWDGLNYYKKLF Frequency
ACD_IE_462_TCRGD__T_CELLS	0.042
ACD_IE_462_UNSORTED_IEC	0.003
ACD_IE_472_TCRGD__T_CELLS	0.000
ACD_IE_472_UNSORTED_IEC	0.012
GFD_IE_487_TCRGD__T_CELLS	0.225
GFD_IE_487_UNSORTED_IEC	0.228
GFD_IE_490_TCRGD__T_CELLS	0.000
GFD_IE_490_UNSORTED_IEC	0.021
Control_IE_24247_TCRGD__T_CELLS	0.099
Control_IE_24247_UNSORTED_IEC	0.024
Control_IE_24330_TCRGD__T_CELLS	0.026
Control_IE_24330_UNSORTED_IEC	0.012