Shared TCRs in peripheral blood offer robust celiac disease classification independent of gluten Intake

Rebecca Elyanow[1], Rok Seon Choung[2], Eric V. Marietta[2], Ramit Bharanikumar[1], Wenyu Zhou[1], Haiyin Chen-Harris[1], Govind Bhagat[4], Peter H. Green[5], Bryan Howie[1], Harlan S Robins[1], Susan L. Neuhausen[3], Joseph A. Murray[2]

1: Adaptive Biotechnologies - 1165 Eastlake Ave E, Seattle, WA 98109

2: Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, MN

3: Department of Population Sciences, Beckman Research Institute, City of Hope, Duarte, CA, United States of America

4. Department of Pathology, Saint Louis University School of Medicine, SSM Health, Saint Louis, MO

5. Department of Medicine, Division of Gastroenterology, Celiac Disease Center, Columbia University Irving Medical Center, New York, NY

Corresponding Author

Joseph A. Murray

Murray.Joseph@mayo.edu

Author Contributions

JAM, SLN, RSC, EVM, GB, and PG designed the initial study. RE and RB performed data analysis with support from WZ and HCH. BH and HSR oversaw generation of the T-cell receptor sequencing data. RE, RB, JAM, RSC, and EBM wrote the manuscript, with contributions from SLN, WZ, and HC. All authors commented on and approved the final version of the manuscript.

Author Declarations

RE, RB, WZ, HCH, BH, and HSR have employment and equity ownership with Adaptive Biotechnologies.

Funding

SLN is partially funded by the Morris and Horowitz Families Endowed Professorship.

Data Availability

All clonal breadth and model scores used to generate figures will be made available as a supplementary file. Full TCR repertoire sequencing data for 520 individuals sequenced in this study will be made available at https://clients.adaptivebiotech.com/pub/elyanow-2024-s.

Abstract

Background and Aims

Celiac disease (CeD) is a chronic digestive autoimmune disorder affecting approximately 1% of the worldwide population. It is driven by T cells activated by specific HLA-DQ2 or HLA-DQ8 molecules leading to the destruction of intestinal villi. We aimed to characterize shared CeD-specific T cells in patients on and off a gluten-free diet (GFD) from a large cohort of cases and controls.

Methods

We performed bulk TCRβ immune sequencing of the peripheral blood of 1,604 biopsy-confirmed CeD patients (1,339 on GFD, 265 on normal diet) and 1,100 controls. We identified over 300 TCRβs enriched in CeD cases versus controls in an HLA-aware manner, controlling for CeD risk alleles.

Results

CeD-associated TCRβs were found to be more predictive of disease than previously characterized gliadin and glutenin-binding TCRs in a validation cohort. Furthermore, the clonal breadth of these TCRβs was associated with increased intestinal damage. Immune sequencing of the peripheral blood also uncovered repertoire-level differences between CeD patients and controls. CeD patients displayed significantly higher productive clonality compared to age-matched controls as well as expansion of TCRβs specific to cytomegalovirus (CMV) and Epstein-Barr Virus (EBV).

Conclusions

These findings underscore the value of unbiased immune repertoire sequencing to identify novel biomarkers for autoimmune disease and to discover new disease mechanisms which can improve both diagnosis and treatment of disease.

Introduction

Celiac disease (CeD) is an inflammatory gastrointestinal disease caused by the ingestion of gluten [1,2]. In individuals with CeD, consumption of gluten results in enteropathy, and a gluten-free diet is the standard treatment [3]. It is one of the most common chronic digestive disorders, affecting approximately 1% of the population worldwide, and the prevalence of the disease continues to increase [4]. CeD occurs only in individuals with susceptible genes, the class II human leukocyte antigen (HLA) alleles, HLA-DQ2 and HLA-DQ8 [1,5]. HLA-DQ2.5 confers the greatest celiac disease risk, with DQ2.2 and DQ7.5 forming DQ2.5 in trans configuration, further contributing to susceptibility [6]. In CeD, both DQ2 and DQ8 present gluten-derived peptides to CD4+ T cells as antigens [7]. Numerous studies have demonstrated that several gluten-derived peptides can stimulate CD4+ T cells in patients with CeD, most notably the 33-mer alpha-gliadin peptide with 6 overlapping immunogenic epitopes [8, 9]. These anti-gluten CD4+ T cells and associated upregulation of IL-15 provide intraepithelial lymphocytes (CD8+) with a “license to kill” intestinal epithelial cells, resulting in tissue destruction in the small bowel [9,10,11,12].

Despite excellent evidence for the causal role of gluten-derived antigens, several puzzles remain in understanding the pathogenesis of CeD. For example, what triggers an increase in antibodies to the self-antigen tissue transglutaminase in the absence of a strong response by T cells to gluten-derived foreign antigens [13,14]? And why do many individuals with HLA risk alleles never develop CeD? Exogenous antigens other than gluten, such as viral infections, especially gastrointestinal viruses, have been hypothesized to play a role in triggering disease, but no causal antigens or T cell receptors (TCRs) have been discovered [15]. While the gluten-specific CD4+ T cell is central to the pathogenesis of CeD, it is highly likely that T-cell responses to other self or exogenous antigens also occur in CeD and are important for disease progression. To understand the causal mechanism behind CeD, it is crucial to identify additional TCRs associated with CeD beyond the known TCRs responding to gluten.

Next generation sequencing of adaptive immune repertoires and advanced analytics have provided researchers with an unbiased approach to determining underlying TCRs involved in disease-specific T cell responses [16,17]. Several studies have identified disease-specific TCRs by comparing the T cell repertoire of a given disease cohort with that of control cohorts [18,19]. Prior studies have performed immune sequencing of celiac patients, but they are either restricted to tissue-specific T cells or gliadin-stimulated T cells [20-25]. In this study, we analyzed the peripheral circulating TCR repertoire of CeD patients, from both treated (gluten-free diet) and untreated healthy controls. This study represents the largest immune sequencing effort of the peripheral blood of CeD patients as well as the largest of treated CeD patients.

The diversity of the T cell immune repertoire is critical in maintaining an effective immune response. Several studies have shown that a decrease in T cell diversity is common to several autoimmune diseases, such as rheumatoid arthritis, multiple sclerosis, and type 1 diabetes [26-29]. However, the diversity of the TCR repertoire in CeD had not been previously studied in an unbiased way. We postulated that CeD patients would display a less diverse TCR repertoire compared to controls. Moreover, we hypothesized that some specific, public, TCRs may react to antigens beyond known targets.

In this study, we identified 370 shared TCRs specific to CeD patients but not controls, including 318 novel CeD-associated TCRs. We demonstrated that these novel TCRs were more diagnostic of CeD than previously characterized gliadin-specific TCRs, particularly for subjects adhering to a gluten-free diet. We additionally investigated whether CeD induced large-scale changed to the peripheral T cell repertoire and found that CeD patients had significantly less diverse repertoires than controls and exhibited clonal expansion of T cells specific to common viral exposures.

Materials and Methods

Study Population

Our study population comprised cohorts from three distinct study populations across the US and Canada, including: 613 celiac disease (CeD) patients and 696 controls from the Beckman Research Institute of City of Hope (BRICOH), 728 CeD patients and 147 controls from the Mayo Clinic (Rochester, MN), and 263 CeD patients and 257 controls from the Seattle area collected by Adaptive Biotechnologies. All CeD patients were confirmed via duodenal biopsy; controls have not been biopsy confirmed negative.

Our validation cohort included 8032 previously sequenced healthy donors and donors with other infectious and autoimmune diseases, including Covid, multiple sclerosis, Crohn’s disease, and type I diabetes (Supplemental Table 1). Testing for CeD was not performed on these donors, but donors were assumed to be CeD negative for the purpose of comparing CeD-associated TCR prevalence in the context of other autoimmune and infectious diseases.

Of the 1604 total CeD patients, 1339 were adhering to a gluten-free diet at the time of sampling and 265 were consuming gluten normally. We observed no significant difference in sequencing depth (total productive templates) or age between CeD cases and controls. CeD cases skewed significantly more towards female (76%) compared to controls (59%), but this did not result in biased frequency of CeD-specific TCRs (Supplemental Figure 1).

Published CeD TCRs

We collected a set of 445 published CeD-associated TCRβs from IEDB [30], VDJdb [31] and McPAS [32]. We included only CDR3s with curated V and J genes (rather than imputed). This sequence set included 10 TCRβ sequences binding to epitopes deriving from barley (*Hordeum vulgare*) and the remaining 435 binding to epitopes deriving from common wheat (*Triticum aestivum*).

Sequencing

Genomic DNA was extracted from frozen, plasma-depleted blood samples using the Qiagen DNeasy Blood Extraction Kit (Qiagen). As much as 18 µg of input DNA was then used to perform immunosequencing of the third complementarity determining (CDR3) regions of TCRβ chains using the ImmunoSEQ Assay (Adaptive Biotechnologies). Briefly, input DNA was amplified in a bias-controlled multiplex PCR, followed by high-throughput sequencing. Sequences were collapsed and filtered to identify and quantitate the absolute abundance (i.e. templates) of each unique TCRβ CDR3 region for further analysis, as previously described [33-35]. To quantify the proportion of T cells out of total nucleated cells input for sequencing, or T cell fraction, a panel of reference genes present in all nucleated cells was amplified simultaneously [36].

HLA Imputation from TCRβs

HLA genotypes for each subject in the sequence discovery and validation sets (described in the section below) were inferred from the T cell repertoire using TCRβ-based HLA classifiers. Briefly, TCRβ repertoire sequencing from a cohort of ~40k individuals were assembled. From this large database, TCRβs specific to each of the 145 most common HLAs were identified. These TCRβs were used to build 145 HLA classifiers, one classifier for each HLA [51].

HLA-aware CeD Enhanced Sequence Discovery

Identification of TCRβs enriched in the CeD patient population was performed using a similar procedure as described previously [34]. Briefly, we first performed HLA inference from TCRβ repertoires as described above, then divided our samples based on presence of one or more of the three main CeD risk alleles (DQ2.5, DQ2.2, and DQ8). For each risk allele, we assessed occurrence in CeD+ cases and healthy controls for each TCRβ in the sequence discovery set and determined statistical enrichment using a one-sided Fisher's Exact Test (FET), testing only for enrichment and not depletion in cases. TCRs with an FET p-value below the threshold were defined as Enhanced Sequences (ES). The best FET p-value threshold was determined as 0.002 by

evaluating the area under the curve (AUC) of a classifier trained on the clonal breadth (defined below) of ESs in 5-fold cross-validation (Supplemental Figure 2).

We split the TCRβ repertoires into training and validation sets. To reduce genetic biases and ensure generalizability of the classifier across the disease population, the model was trained with genetically matched cases and controls and tested on a validation set coming from two different locations and studies (Table 1). The training set contained 613 CeD-positive cases and 696 CeD-negative controls from the City of Hope study. Of the controls, 588 were first degree and 43 were second degree relatives of CeD subjects in the cohort. The validation set contained 991 CeD-positive cases and 8,436 controls.

Cohort	Diagnosis	Samples	Train Set	DQ2.5	DQ2.2	DQ8	GFD	Female %	Median age	Median productive rearr.	Median templates	Median clonality
Adaptive	Control	257	N	167	82	65	0	84%	40.0	789410	666772	0.059
	CeD	263	N	173	97	54	251	85.6%	38.0	729829	267658	0.065
BRICOH	Control	696	Y	604	139	133	0	53.3%	44.0	714140	285717	0.071
	CeD	613	Y	524	234	79	569	77.2%	50.0	663005	277954	0.092
Mayo Clinic	Control	147	N	42	23	43	0	76.2%	47.0	729879	238877	0.060
	CeD	728	N	602	188	131	519	71.6%	45.0	710985	293344	0.079
Other	Control	8032	N	1658	1518	2094	0	58.7%	39.0	735876	256862	0.06

Table 1. Balance of CeD cases and controls in the training and validation sets.

Clonal Breadth Analysis

Clonal breadth refers to the number of unique productive TCRβ nucleotide rearrangements corresponding to a given set of sequences in a patient's repertoire. The unit of the clonal breadth is rearrangements per million (RPM). Since most sequenced repertoires contain fewer than 1 million TCRβs (mean = 210,753), a sampling strategy was employed. Repertoires were randomly down sampled without replacement to 100k TCRβs 10 times to achieve an estimate of RPM. Repertoire with fewer than 100k TCRβs were omitted from analysis.

We used clonal breadth to evaluate the CeD specificity of three sets of TCRβ sequences: our novel set of 318 TCRβs, 409 previously published gliadin-specific TCRβs, and 68 gliadin-specific R-motif TCRβs. Clonal breadth for each of these sequence sets was computed for 884 CeD cases and 3,624 controls from the validation set that harbored at least one of the DQ2 risk alleles (DQ2.5 or DQ2.2).

To assess the significance of ES sets in relation to CeD severity, we carried out correlation analysis between clonal breadth and Marsh score for 367 DQ2+ CeD patients with Marsh score measurements. Marsh score is a metric measuring severity of intestinal damage on the scale of 1, 2, and 3, where 3 is further subcategorized into 3A, 3B, and 3C. A Marsh score of 1 indicates no villous atrophy while 3C indicates total villous atrophy.

Clonality in CeD Repertoires

To observe the large-scale effects of CeD on the T cell repertoire we computed Productive Simpson Clonality (the square root of Simpson's diversity index on all productive TCRβ rearrangements), where 1 corresponds to a monoclonal repertoire and 0 a polyclonal repertoire. To account for the consistent increase of clonality with age due to clonal expansion of memory T cells, we partitioned the data into 10-year age bins and performed a Mann-Whitney U (MWU) test within each age bin to assess significance of the difference between cases and controls.

CeD Associated Viral Activation Analysis

We hypothesized that CeD may stress the immune system and cause a hyper-reactive state, resulting in clonal expansion of TCRs responding to previous viral or other exposures. To evaluate this hypothesis, we divided the subjects based on inferred CMV and EBV positivity, predicted from 26,139 CMV and 9,741 EBV-associated TCRβs from a prior study of over 30k T cell repertoires [36,51]. We then compared clonal breadth associated with CMV and EBV between groups.

HLA Disease Risk Analysis

HLA enrichment in patients compared to controls was assessed for single and combination alleles using the Fisher's Exact Test (FET), with Bonferroni correction for multiple comparison. Odd's ratio (OR) was used to identify additive, non-additive and interaction effects at the HLA loci in CeD patients relative to healthy controls. For a given HLA allele A, let A^+ denote presence of the allele and A^- denote absence. The Odd's ratio is defined as:

$$OR = \frac{\frac{A^+\ patients}{A^-\ patients}}{\frac{A^+\ controls}{A^-\ controls}}$$

Results

CeD HLA Risk Alleles

Before identifying CeD-associated TCRβs, we first inferred HLAs for each subject in the training cohort from their full TCRβ repertoire for the purpose of correcting for HLA bias during in the case and control repertoires (see Methods). We confirmed previously reported HLA effects for different CeD risk alleles [6,38] including relatively equal risk of one copy of the DQ2.5 allele (OR = 6.29, CI = 5.54-7.14) and DQ2.5 in the trans configuration, determined by the DQ7.5/DQ2.2 genotype (OR = 6.70, CI = 5.04-8.88). Under an additive model, we found strong effects for DQ2.5 (OR = 9.46, 95% CI = 8.22-10.89), DQ8 (OR = 2.40, 95% CI = 1.97-2.92) and DQ2.2 (OR = 2.95, 95% CI = 2.41-3.61) in CeD cases relative to controls. Additionally, we observed previously reported [6] zygosity and interactions effects at these HLA loci (see Table 2). These results demonstrated the feasibility of identifying complex HLA interactions in the HLA loci contributing to Celiac and other autoimmune disease risk [53] using TCRβ repertoires.

Genotype	Reference	OR [95% CI]	p-value
DQ2.5	DQX	9.46 [5.04-8.88]	<0.0001
DQ2.5/DQ2.5	DQX/DQX	18.99 [14.09-25.76]	<0.0001
DQ2.5/DQX	DQX/DQY	6.29 [5.54-7.14]	<0.0001
DQ2.5/DQ2.5	DQ2.5/DQX	1.70 [1.44-2.01]	<0.0001
DQ2.5/DQ2.2	DQ2.5/DQX	2.08 [1.77-2.42]	<0.0001
DQ2.2	DQX	2.95 [2.41-3.61]	<0.0001
DQ2.2/DQ2.2	DQX/DQX	1.98 [0.98-3.66]	0.04
DQ2.2/DQX	DQX/DQY	2.95 [2.37-3.67]	<0.0001
DQ2.2/DQ2.2	DQ2.2/DQX	0.49 [0.26-0.87]	0.01
DQ8	DQX	2.40 [1.97-2.92]	<0.0001
DQ8/DQ8	DQX/DQX	5.42 [3.14-9.41]	<0.0001
DQ8/DQX	DQX/DQY	1.98 [1.60-2.46]	<0.0001
DQ8/DQ8	DQ8/DQX	1.09 [0.73-1.62]	0.60
DQ7.5/DQ7.5	DQX/DQX	0.32 [0.10-0.82]	<0.001
DQ7.5/DQX	DQX/DQY	1.22 [0.97-1.53]	0.07
DQ7.5/DQ2.2	DQX/DQX	6.70 [5.04-8.88]	<0.0001

Table 2. CeD risk conferred by zygosity and interaction effects at the HLA-DQ loci. DQX and DQY indicate a haplotype different from DQ2.5, DQ8, DQ2.2 and DQ7.5.

Enhanced Sequences predict CeD status better than published CeD TCRβs

Using our HLA-aware sequence discovery method (see Methods), we identified a total of 370 CeD enhanced sequences (ESs) from the training cohort (613 CeD cases and 696 controls). The controls included samples with CeD HLA risk haplotypes (DQ2.5, DQ2.2, DQ8 and DQ7.5) to ensure that observed ESs are associated with CeD rather than HLA-linked exposures. The V-gene and J-gene distributions of our ES were similar

to published CeD TCRs, with V7-2 being the most common V-gene and J2-7 being the most common J-gene. However, some ES V-genes, such as V2-1, V28-1, V06-05, and V27-1, were not represented in published sequences (Supplemental Figure 7).

Our CeD ESs included 50 TCRβs corresponding to the canonical DQ2.5 CDR3 motif CASSxRxTDTQYF+TRBV7-2/3+TRBJ2-3, known as the R-motif [39] as well as 2 TCRβs previously identified via tetramer sorting, CASSQGSGGNEQFF+TCRBV04-03+TCRBJ02-01 and CASSQGQDTEAFF+TCRBV07-03+TCRBJ01-01, which bind antigens DQ2.5-glia-ω2 and DQ2.5-glia-α1 respectively [40,41]. After removing published sequences, we were left with 318 novel CeD ESs. Comparing the 68 published sequences matching the R-motif and 409 other published TCRβs (see Methods), we observed that the R-motif was significantly more enriched in CeD patients than other published epitope-specific TCRβs (OR 3.45, p=4e-71). However, our novel ESs were far more enriched in CeD than the R-motif TCRs (OR 6.85, p=7e-80) (Table 3), suggesting these novel ESs are better disease biomarkers than what TCR published to date.

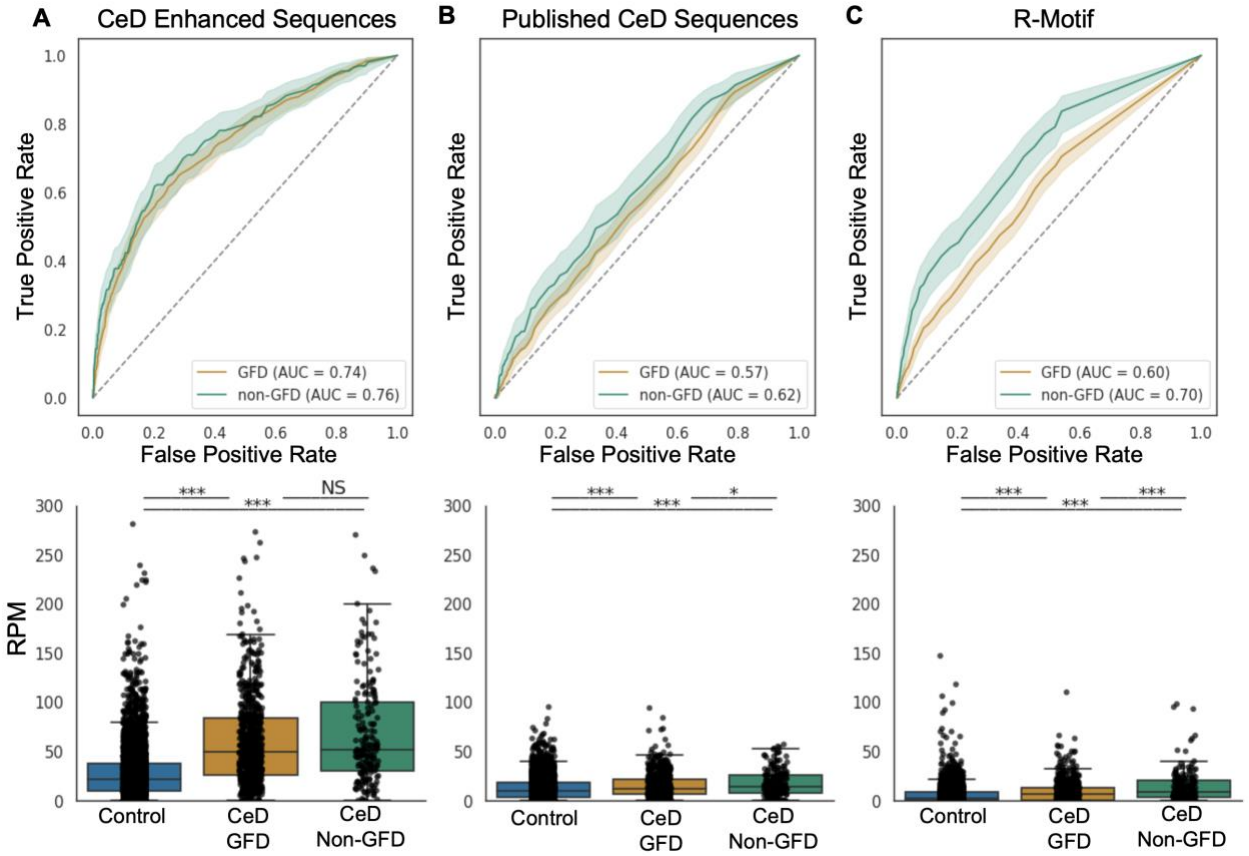

Figure 1 ROC curves (top panels) and clonal breadth (bottom panels) of Enhanced TCRs (N = 318, left panels), published TCRs (N = 409, middle panels), and the R-Motif (N = 68 TCRs, right

panels, present in both published and enhanced TCRs) in DQ2+ CeD patients and controls. *p< 0.05, **p< 0.01, ***p< 0.001.

We found that our novel ESs could distinguish CeD cases from controls harboring the DQ2 risk allele with significantly higher accuracy (AUC=0.74 GFD, 0.76 non-GFD) than published TCRs (AUC=0.57 GFD, 0.62 non-GFD), including the R-motif (AUC=0.60 GFD, 0.70 non-GFD, Figure 1 top panels). The presence of the risk alleles RPM of novel ESs was significantly elevated in CeD cases compared to controls with viral or bacterial infections (Covid and Lyme) as well as with other autoimmune disorders (type I diabetes, multiple sclerosis, ulcerative colitis and Crohn's disease) (MWU p<1e-10, Supplemental Figure 3a).

Interestingly, we found no significant difference in the enhanced sequence RPM for patients on or off a gluten-free diet (p > .05), while the RPM of the 409 published CeD-TCRβs and 68 R-motif TCRβs were significantly elevated in patients consuming gluten (R-motif, p<.001, other published, p < 0.05 Figure 1 lower). This suggests the novel CeD ES TCRβs are robust to gluten challenge or absence in the diet.

We note that neither our CeD ESs nor published DQ8-specific TCRs [42] distinguished cases from controls in the DQ8-specific context (Supplementary Figure 4a-b), indicating that the majority of public CeD TCRβs in the peripheral blood bind either DQ2.5 or DQ2.2.

Epitope Name	Epitope	# TCRs	CeD Fraction	Control Fraction	Odds Ratio	FET p-val
Novel CeD Enhanced Sequences	unknown	318	94%	69%	6.85	7e-80
R-Motif (DQ2.5-glia-a2)	PQPELPYPQPE	68	70%	41%	3.45	4e-71
DQ8-glia-a1	SGEGSFQPSQENP	16	3%	1%	3.09	1e-5
DQ2.5-glia-a1a	PFPQPELPY	87	53%	28%	2.87	1e-53
DQ2.2-glut-L1	QPPFSEQEQPVL	18	15%	7%	2.27	3e-14
DQ2.5-glia-a2	PQPELPYPQPE	198	54%	37%	2.00	1e-24
DQ2.5-glia-omega2	FPQPEQPFPWQP	57	16%	9%	1.96	2e-11
DQ2.5-glia-omega1	PQQPFPQPEQPFP	27	20%	12%	1.85	2e-11
DQ2.5-glia-a1a&a2	LQPFPQPELPYPQPQ	25	27%	20%	1.48	6e-7

Table 3. Prevalence of CeD TCRs in validation set grouped by epitope. Odds ratios and corresponding Fisher's exact p-values indicate enrichment of TCRs in CeD cases versus controls.

Enhanced TCR clonal breadth correlated with Marsh score in patients not on GFD

We observed that the clonal breadth of our novel CeD ESs correlated with Marsh score (Pearson r=0.105, p=0.037) (Figure 2a). In comparison, published CeD TCRs (r=0.014, p=0.79) including the R-Motif (r=0.064, p=0.21) were not correlated with Marsh score (Figure 2b-c). This finding suggests that profiling of CeD enhanced sequences in the peripheral blood may be a useful biomarker of severity, disease progression, or recovery.

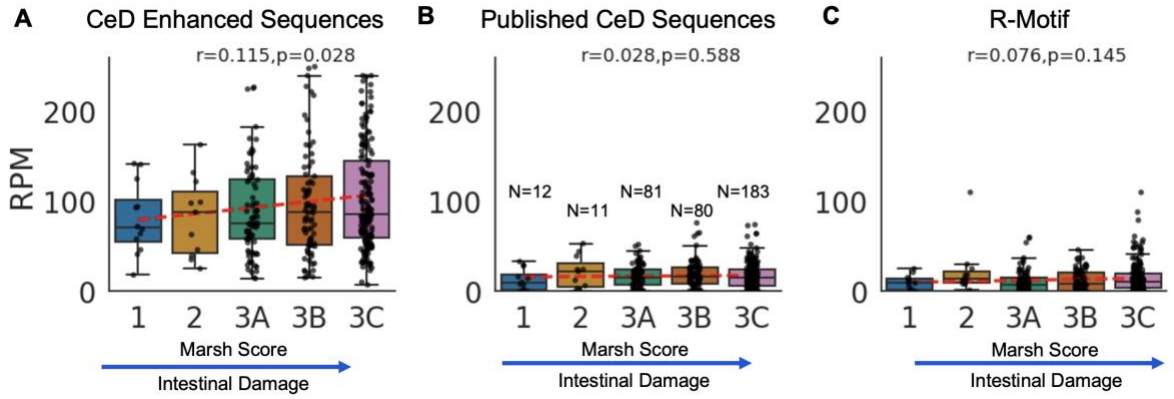

Figure 2. Correlation between Rearrangements Per Million (RPM) and Marsh score. (A) CeD enhanced sequence RPM was positively correlated with Marsh score (Pearson r = 1.15, p = 0.028). (B) CeD published sequence RPM were not significantly correlated with Marsh score. (C) CeD R-Motif sequence RPM was not significantly correlated with Marsh score.

Higher clonality in CeD repertoires

In addition to measuring the dynamics of CeD-specific TCRβs, we also observed large-scale effects of the disease on T-cell repertoire. We observed a significant increase in Simpson clonality in CeD patients compared to controls. Clonality was significantly increased in CeD cases in all age bins except for bin 70-80 ($p<0.001$, Figure 3a). This increase in clonality was significant for both CD8+ and CD4+ TCRs (Supplemental Figure 5) and was significant for patients adhering to a GFD (Supplemental Figure 6a) and consuming gluten normally (Supplemental Figure 6b).

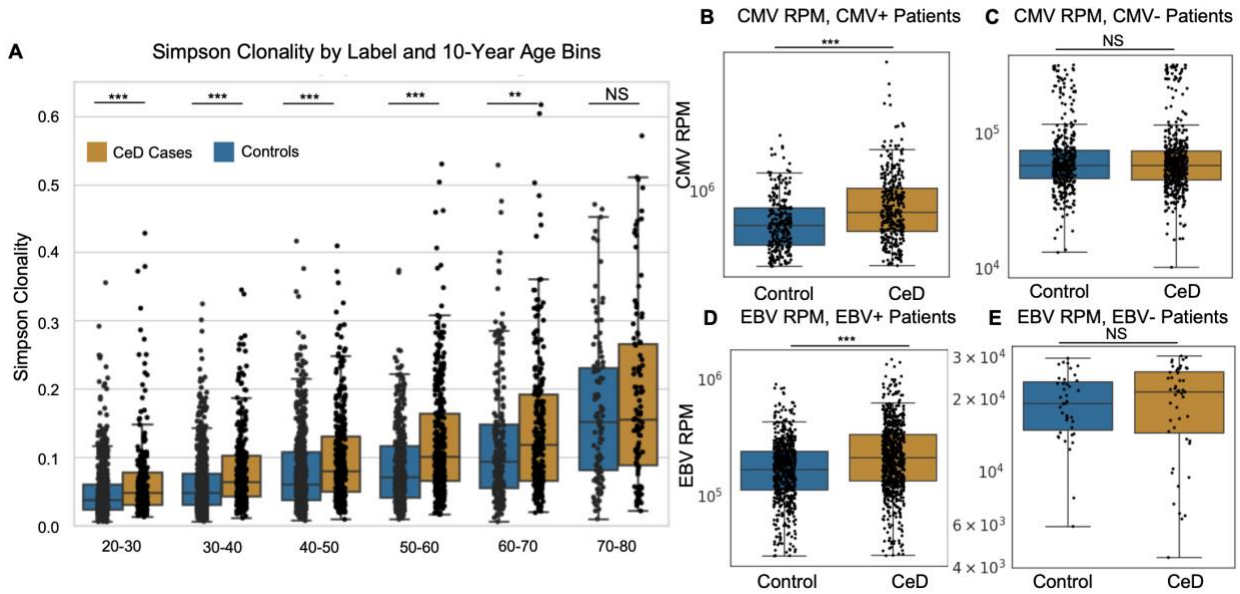

Figure 3. (A) Simpson clonality between CeD cases (orange) and controls (blue) grouped in 10-year age bins. (B) CMV RPM between CMV-positive CeD cases and controls. CeD cases had significantly elevated CMV RPM versus controls (MWU p<1e-9) (C) CMV RPM between CMV-negative CeD cases and controls. CeD cases did not significantly differ from controls (p>.6). (D) EBV RPM between EBV-positive CeD cases and controls. CeD cases had elevated EBV RPM versus controls (MWU p<1e-13). (E) EBV RPM between EBV-negative CeD cases and controls. CeD cases did not significantly differ from controls (MWU p>.29). *p< .05, **p< .01, ***p< .001.

Expansion of T cells from viral exposures

We confirmed that CeD patients positive for CMV and EBV had significantly higher CMV and EBV RPM respectively than CMV and EBV-positive controls (p < 1e-9, p<1e-13 respectively, Figure 3b,d). This indicated an expansion of viral memory T cells in subjects with CeD. As expected, we observed no difference in CMV and EBV RPM for CMV- and EBV- patients. Importantly, we note that the inferred rate of positivity for these viral infections was not increased in CeD cases compared to controls, with 43% of CeD cases and 44% of controls being CMV-positive and 94% of CeD cases and 94% of controls being EBV positive (p > .5). This result indicated that CMV and EBV infection are not likely to be necessary triggers for CeD, which differs from other autoimmune disorders such as multiple sclerosis, where EBV infection was a prerequisite for disease onset [45].

Discussion

Through immune sequencing of the TCRβ repertoires from the peripheral blood, we identified over 300 novel public TCRβ enhanced sequences (ESs) which were present

in significantly higher proportion in 696 CeD cases versus 613 controls. We demonstrated the utility of these TCRβs in a validation set consisting of 884 DQ2+ CeD cases and 3,624 DQ2+ controls. The novel sequence set significantly outperformed published TCRβs in distinguishing CeD cases from controls, especially for subjects adhering to a gluten-free diet. Most novel TCRβs had similar V-gene usage patterns as published ones (V7-2, V20-1), while some novel TCRβs had previously unobserved V-gene usage patterns (V27-1, V06-5, V28-1, V2-1). Additionally, the ESs ability to distinguish cases from controls despite sharing the CeD risk haplotypes further confirms that these sequences are primarily CeD driven and not associated with a specific HLA haplotype.

The discovery of novel public ESs that are robust to dietary gluten intake suggests that these TCRs may recognize antigens that are central to the pathogenesis of CeD but have remained elusive in previous studies. One possibility is that these ESs recognize non-gluten antigens. Another possibility is that they recognize gluten antigens but harbor a phenotype that promotes longer persistence in memory than other gluten-reactive T cells, such as those with the R-motif. Additionally, the identification of these TCRs in the peripheral blood could indicate a systemic immune response that transcends the localized gut inflammation typically associated with CeD. This finding opens new possibilities for exploring the role of systemic autoimmunity in CeD that may have implications for the identification of novel therapeutic targets that can modulate this immune response. However, this study was limited to only assessing the statistical association of TCRβs to CeD. Future studies will be needed to experimentally confirm the antigen-specificity of these public ESs to truly understand their relationship to disease pathology.

We further observed that the clonal breadth of these novel ESs correlated with increasing intestinal damage as measured by Marsh score. These public TCRβs could potentially serve as a non-invasive surrogate marker for assessing intestinal damage, potentially reducing the need for repeated biopsies. However, follow-up studies would be required to validate and calibrate this signal. The lack of correlation observed with other published sequences, including the gliadin-a2-binding “R-motif”, underscores the unique nature of these novel ESs and suggests that they may be capturing aspects of the disease process that are not reflected in previously identified TCR sequences. These findings demonstrate that public TCRβs in the peripheral blood are important biomarkers for CeD and highlight the advantage of unbiased sequence discovery through non-invasive bulk immune repertoire sequencing of case/control cohorts. This experimental design enables robust detection of autoimmune-associated TCRs without restricting to specific antigens.

We additionally presented evidence of increased repertoire clonality and clonal expansion of viral TCRβs in CeD. We believe this could have important implications for understanding the mechanisms underpinning autoimmunity in CeD. One mechanism that could explain the expansion of memory CMV and EBV TCRs and overall higher clonality is bystander activation. During bystander activation, self-reactive T cells induce activation of other antigen-specific memory T cells to produce inflammatory cytokines [43-45]. Another possible mechanism is viral reactivation. This occurs when a latent infection switches to a lytic stage of replication, resulting in an adaptive immune response [46]. CMV and EBV both have latent stages, where they remain dormant in infected cells. T cell reactivation associated with viruses has been studied in the context of autoimmunity [52], though prior work has not implicated CMV or EBV reactivation in CeD [47-49].

The hypothesis that bystander activation or viral reactivation could be contributing to the observed increase in clonality aligns with emerging evidence in other autoimmune diseases, where viral infections have been implicated as triggers for disease onset or flares. More research is needed to understand the role of latent viral infections and autoimmunity, especially in CeD. Future studies could focus on longitudinal monitoring of CeD patients with known viral exposures to assess changes in TCR repertoire and symptom severity over time.

While a gluten-free diet remains the standard of care for CeD management, many therapeutics are in active development. Our novel CeD public TCRβs possess the following desirable qualities 1) measurable with non-invasive technology, 2) robust to diet and likely associated to a broad set of celiac-associated antigens 3) strongly correlated with disease severity that can only be measured with invasive endoscopy. Furthermore, our observation of increased CMV and EBV clonal breadth in CeD patients highlights the usefulness of bulk TCR sequencing for monitoring viral reactivation in CeD, which may prove to be important for symptom management. Together, these properties suggest bulk TCR sequencing, and particularly public CeD TCRs, may be ideally suited in assessing treatment outcome in clinical trials.

References

1. Jabri, Bana, and Ludvig M Sollid. “Mechanisms of Disease: Immunopathogenesis of Celiac Disease.” *Nature Clinical Practice Gastroenterology & Hepatology* 3, no. 9 (September 2006): 516–25. https://doi.org/10.1038/ncpgasthep0582.
2. Dicke, W. K., H. A. Weijers, and J. H. V. D. Kamer. “Coeliac Disease *The Presence in Wheat of a Factor Having a Deleterious Effect in Cases of Coeliac Disease*.” *Acta Paediatrica* 42, no. 1 (January 1953): 34–42. https://doi.org/10.1111/j.1651-2227.1953.tb05563.x.
3. Husby, Steffen, and Julio C. Bai. “Follow-up of Celiac Disease.” *Gastroenterology Clinics of North America* 48, no. 1 (March 2019): 127–36. https://doi.org/10.1016/j.gtc.2018.09.009.
4. Kang, J. Y., A. H. Y. Kang, A. Green, K. A. Gwee, and K. Y. Ho. “Systematic Review: Worldwide Variation in the Frequency of Coeliac Disease and Changes over Time.” *Alimentary Pharmacology & Therapeutics* 38, no. 3 (August 2013): 226–45. https://doi.org/10.1111/apt.12373.
5. Seehusen, Dean A. “Comparative Accuracy of Diagnostic Tests for Celiac Disease.” *American Family Physician* 95, no. 11 (June 1, 2017): 726–28.
6. Lenz, Tobias L, Aaron J Deutsch, Buhm Han, Xinli Hu, Yukinori Okada, Stephen Eyre, Michael Knapp, et al. “Widespread Non-Additive and Interaction Effects within HLA Loci Modulate the Risk of Autoimmune Diseases.” *Nature Genetics* 47, no. 9 (September 2015): 1085–90. https://doi.org/10.1038/ng.3379.
7. Iversen, Rasmus, and Ludvig M. Sollid. “The Immunobiology and Pathogenesis of Celiac Disease.” *Annual Review of Pathology* 18 (January 24, 2023): 47–70. https://doi.org/10.1146/annurev-pathmechdis-031521-032634.
8. Rossjohn, J., and F. Koning. “A Biased View toward Celiac Disease.” *Mucosal Immunology* 9, no. 3 (May 2016): 583–86. https://doi.org/10.1038/mi.2016.17.
9. Qiao, Shuo-Wang, Elin Bergseng, Øyvind Molberg, Jiang Xia, Burkhard Fleckenstein, Chaitan Khosla, and Ludvig M. Sollid. “Antigen Presentation to Celiac Lesion-Derived T Cells of a 33-Mer Gliadin Peptide Naturally Formed by Gastrointestinal Digestion.” *The Journal of Immunology* 173, no. 3 (August 1, 2004): 1757–62. https://doi.org/10.4049/jimmunol.173.3.1757.
10. Sollid, Ludvig M. “Intraepithelial Lymphocytes in Celiac Disease: License to Kill Revealed.” *Immunity* 21, no. 3 (September 2004): 303–4. https://doi.org/10.1016/j.immuni.2004.08.002.
11. Jabri, Bana, and Ludvig M. Sollid. “T Cells in Celiac Disease.” *Journal of Immunology (Baltimore, Md.: 1950)* 198, no. 8 (April 15, 2017): 3005–14. https://doi.org/10.4049/jimmunol.1601693.
12. Abadie, Valérie, Sangman M. Kim, Thomas Lejeune, Brad A. Palanski, Jordan D. Ernest, Olivier Tastet, Jordan Voisine, et al. “IL-15, Gluten and HLA-DQ8 Drive Tissue Destruction in Coeliac Disease.” *Nature* 578, no. 7796 (February 2020): 600–604. https://doi.org/10.1038/s41586-020-2003-8.
13. Caproni, Marzia, Manuela Capone, Maria Caterina Rossi, Veronica Santarlasci, Laura Maggi, Alessio Mazzoni, Beatrice Rossettini, et al. “T Cell Response Toward Tissue-and Epidermal-Transglutaminases in Coeliac Disease Patients Developing Dermatitis Herpetiformis.” *Frontiers in Immunology* 12 (2021): 645143. https://doi.org/10.3389/fimmu.2021.645143.
14. Ciccocioppo, R., A. Finamore, E. Mengheri, D. Millimaggi, B. Esslinger, W. Dieterich, F. Papola, et al. “Isolation and Characterization of Circulating Tissue Transglutaminase-Specific T Cells in Coeliac Disease.” *International Journal of Immunopathology and Pharmacology* 23, no. 1 (2010): 179–91. https://doi.org/10.1177/039463201002300116.
15. Brown, Judy J., Bana Jabri, and Terence S. Dermody. “A Viral Trigger for Celiac Disease.” *PLoS Pathogens* 14, no. 9 (September 2018): e1007181. https://doi.org/10.1371/journal.ppat.1007181.
16. Delmonte, Ottavia M., Cihan Oguz, Kerry Dobbs, Katherine Myint-Hpu, Boaz Palterer, Michael S. Abers, Deborah Draper, et al. “Perturbations of the T-Cell Receptor Repertoire in Response to SARS-CoV-2 in Immunocompetent and Immunocompromised Individuals.” *The Journal of Allergy and Clinical Immunology* 153, no. 6 (June 2024): 1655–67. https://doi.org/10.1016/j.jaci.2023.12.011.

17. Mitchell, Angela M., Erin E. Baschal, Kristen A. McDaniel, Theodore Fleury, Hyelin Choi, Laura Pyle, Liping Yu, Marian J. Rewers, Maki Nakayama, and Aaron W. Michels. “Tracking DNA-Based Antigen-Specific T Cell Receptors during Progression to Type 1 Diabetes.” *Science Advances* 9, no. 49 (December 8, 2023): eadj6975. https://doi.org/10.1126/sciadv.adj6975.
18. Genolet, Raphael, Sara Bobisse, Johanna Chiffelle, Marion Arnaud, Rémy Petremand, Lise Queiroz, Alexandra Michel, et al. “TCR Sequencing and Cloning Methods for Repertoire Analysis and Isolation of Tumor-Reactive TCRs.” *Cell Reports Methods* 3, no. 4 (April 24, 2023): 100459. https://doi.org/10.1016/j.crmeth.2023.100459.
19. Musvosvi, Munyaradzi, Huang Huang, Chunlin Wang, Qiong Xia, Virginie Rozot, Akshaya Krishnan, Peter Acs, et al. “T Cell Receptor Repertoires Associated with Control and Disease Progression Following Mycobacterium Tuberculosis Infection.” *Nature Medicine* 29, no. 1 (January 2023): 258–69. https://doi.org/10.1038/s41591-022-02110-9.
20. Dahal-Koirala, S, L F Risnes, A Christophersen, V K Sarna, K Ea Lundin, L M Sollid, and S W Qiao. “TCR Sequencing of Single Cells Reactive to DQ2.5-Glia-A2 and DQ2.5-Glia-Ω2 Reveals Clonal Expansion and Epitope-Specific V-Gene Usage.” *Mucosal Immunology* 9, no. 3 (May 2016): 587–96. https://doi.org/10.1038/mi.2015.147.
21. Qiao, Shuo-Wang, Asbjørn Christophersen, Knut E. A. Lundin, and Ludvig M. Sollid. “Biased Usage and Preferred Pairing of α- and β-Chains of TCRs Specific for an Immunodominant Gluten Epitope in Coeliac Disease.” *International Immunology* 26, no. 1 (January 1, 2014): 13–19. https://doi.org/10.1093/intimm/dxt037.
22. Yohannes, Dawit A., Tobias L. Freitag, Andrea De Kauwe, Katri Kaukinen, Kalle Kurppa, Pirjo Wacklin, Markku Mäki, et al. “Deep Sequencing of Blood and Gut T-Cell Receptor β-Chains Reveals Gluten-Induced Immune Signatures in Celiac Disease.” *Scientific Reports* 7, no. 1 (December 21, 2017): 17977. https://doi.org/10.1038/s41598-017-18137-9.
23. Yohannes, Dawit A., Katri Kaukinen, Kalle Kurppa, Päivi Saavalainen, and Dario Greco. “Clustering Based Approach for Population Level Identification of Condition-Associated T-Cell Receptor β-Chain CDR3 Sequences.” *BMC Bioinformatics* 22, no. 1 (December 2021): 159. https://doi.org/10.1186/s12859-021-04087-7.
24. Foers, Andrew D, M Saad Shoukat, Oliver E Welsh, Killian Donovan, Russell Petry, Shelley C Evans, Michael Eb FitzPatrick, et al. “Classification of Intestinal T -cell Receptor Repertoires Using Machine Learning Methods Can Identify Patients with Coeliac Disease Regardless of Dietary Gluten Status.” *The Journal of Pathology* 253, no. 3 (March 2021): 279–91. https://doi.org/10.1002/path.5592.
25. Seitz, V., K. Gennermann, S. Elezkurtaj, D. Groth, S. Schaper, A. Dröge, N. Lachmann, et al. “Specific T-Cell Receptor Beta-Rearrangements of Gluten-Triggered CD8+ T-Cells Are Enriched in Celiac Disease Patients’ Duodenal Mucosa.” *Clinical Immunology* 256 (November 2023): 109795. https://doi.org/10.1016/j.clim.2023.109795.
26. Yang, Peiqing, Yijing He, Pingying Qing, Wangdong Xu, Dan Xie, Jean-Baptiste Cazier, Xiao Liu, et al. “Application of T-Cell Receptor Repertoire as a Novel Monitor in Dynamic Tracking and Assessment: A Cohort-Study Based on RA Patients.” *Journal of Cellular and Molecular Medicine* 26, no. 24 (December 2022): 6042–55. https://doi.org/10.1111/jcmm.17623.
27. Attaf, Meriem, Eric Huseby, and Andrew K. Sewell. “Aβ T Cell Receptors as Predictors of Health and Disease.” *Cellular & Molecular Immunology* 12, no. 4 (July 2015): 391–99. https://doi.org/10.1038/cmi.2014.134.
28. Mitchell, Angela M., Erin E. Baschal, Kristen A. McDaniel, Kimber M. Simmons, Laura Pyle, Kathleen Waugh, Andrea K. Steck, et al. “Temporal Development of T Cell Receptor Repertoires during Childhood in Health and Disease.” *JCI Insight* 7, no. 18 (September 22, 2022): e161885. https://doi.org/10.1172/jci.insight.161885.
29. Lossius, Andreas, Jorunn N. Johansen, Frode Vartdal, and Trygve Holmøy. “High-Throughput Sequencing of Immune Repertoires in Multiple Sclerosis.” *Annals of Clinical and Translational Neurology* 3, no. 4 (April 2016): 295–306. https://doi.org/10.1002/acn3.295.
30. Vita, Randi, Swapnil Mahajan, James A Overton, Sandeep Kumar Dhanda, Sheridan Martini, Jason R Cantrell, Daniel K Wheeler, Alessandro Sette, and Bjoern Peters. “The Immune Epitope Database (IEDB): 2018 Update.” *Nucleic Acids Research* 47, no. D1 (January 8, 2019): D339–43. https://doi.org/10.1093/nar/gky1006.

31. Shugay, Mikhail, Dmitriy V Bagaev, Ivan V Zvyagin, Renske M Vroomans, Jeremy Chase Crawford, Garry Dolton, Ekaterina A Komech, et al. “VDJdb: A Curated Database of T-Cell Receptor Sequences with Known Antigen Specificity.” *Nucleic Acids Research* 46, no. D1 (January 4, 2018): D419–27. https://doi.org/10.1093/nar/gkx760.
32. Tickotsky, Nili, Tal Sagiv, Jaime Prilusky, Eric Shifrut, and Nir Friedman. “McPAS-TCR: A Manually Curated Catalogue of Pathology-Associated T Cell Receptor Sequences.” Edited by Jonathan Wren. *Bioinformatics* 33, no. 18 (September 15, 2017): 2924–29. https://doi.org/10.1093/bioinformatics/btx286.
33. Carlson, Christopher S., Ryan O. Emerson, Anna M. Sherwood, Cindy Desmarais, Moon-Wook Chung, Joseph M. Parsons, Michelle S. Steen, et al. “Using Synthetic Templates to Design an Unbiased Multiplex PCR Assay.” *Nature Communications* 4, no. 1 (October 25, 2013): 2680. https://doi.org/10.1038/ncomms3680.
34. Gittelman, Rachel M., Enrico Lavezzo, Thomas M. Snyder, H. Jabran Zahid, Cara L. Carty, Rebecca Elyanow, Sudeb Dalai, et al. “Longitudinal Analysis of T Cell Receptor Repertoires Reveals Shared Patterns of Antigen-Specific Response to SARS-CoV-2 Infection.” *JCI Insight* 7, no. 10 (May 23, 2022): e151849. https://doi.org/10.1172/jci.insight.151849.
35. Robins, Harlan, Cindy Desmarais, Jessica Matthis, Robert Livingston, Jessica Andriesen, Helena Reijonen, Christopher Carlson, Gerold Nepom, Cassian Yee, and Karen Cerosaletti. “Ultra-Sensitive Detection of Rare T Cell Clones.” *Journal of Immunological Methods* 375, no. 1–2 (January 31, 2012): 14–19. https://doi.org/10.1016/j.jim.2011.09.001.
36. Pruessmann, Wiebke, Julie Rytlewski, James Wilmott, Martin C. Mihm, Grace H. Attrill, Beatrice Dyring-Andersen, Paul Fields, et al. “Molecular Analysis of Primary Melanoma T Cells Identifies Patients at Risk for Metastatic Recurrence.” *Nature Cancer* 1, no. 2 (January 20, 2020): 197–209. https://doi.org/10.1038/s43018-019-0019-5.
37. Emerson, Ryan O, William S DeWitt, Marissa Vignali, Jenna Gravley, Joyce K Hu, Edward J Osborne, Cindy Desmarais, et al. “Immunosequencing Identifies Signatures of Cytomegalovirus Exposure History and HLA-Mediated Effects on the T Cell Repertoire.” *Nature Genetics* 49, no. 5 (May 2017): 659–65. https://doi.org/10.1038/ng.3822.
38. Megiorni, Francesca, Barbara Mora, Margherita Bonamico, Maria Barbato, Raffaella Nenna, Giulia Maiella, Patrizia Lulli, and Maria Cristina Mazzilli. “HLA-DQ and Risk Gradient for Celiac Disease.” *Human Immunology* 70, no. 1 (January 2009): 55–59. https://doi.org/10.1016/j.humimm.2008.10.018.
39. Dahal-Koirala, Shiva, Louise Fremgaard Risnes, Ralf Stefan Neumann, Asbjørn Christophersen, Knut E. A. Lundin, Geir Kjetil Sandve, Shuo-Wang Qiao, and Ludvig M. Sollid. “Comprehensive Analysis of CDR3 Sequences in Gluten-Specific T-Cell Receptors Reveals a Dominant R-Motif and Several New Minor Motifs.” *Frontiers in Immunology* 12 (April 13, 2021): 639672. https://doi.org/10.3389/fimmu.2021.639672.
40. Ciacchi, Laura, Carine Farenc, Shiva Dahal-Koirala, Jan Petersen, Ludvig M. Sollid, Hugh H. Reid, and Jamie Rossjohn. “Structural Basis of T Cell Receptor Specificity and Cross-Reactivity of Two HLA-DQ2.5-Restricted Gluten Epitopes in Celiac Disease.” *Journal of Biological Chemistry* 298, no. 3 (March 2022): 101619. https://doi.org/10.1016/j.jbc.2022.101619.
41. Lee, Lik Wee, Shahin Shafiani, Beryl Crossley, Ryan O Emerson, David Williamson, Anna Bunin, Justin Vargas, et al. “Characterisation of T Cell Receptor Repertoires in Coeliac Disease.” *Journal of Clinical Pathology* 77, no. 2 (February 2024): 116–24. https://doi.org/10.1136/jcp-2022-208541.
42. Broughton, Sophie E., Jan Petersen, Alex Theodossis, Stephen W. Scally, Khai Lee Loh, Allan Thompson, Jeroen van Bergen, et al. “Biased T Cell Receptor Usage Directed against Human Leukocyte Antigen DQ8-Restricted Gliadin Peptides Is Associated with Celiac Disease.” *Immunity* 37, no. 4 (October 2012): 611–21. https://doi.org/10.1016/j.immuni.2012.07.013.
43. Shim, Chae-Hyeon, Sookyung Cho, Young-Mi Shin, and Je-Min Choi. “Emerging Role of Bystander T Cell Activation in Autoimmune Diseases.” *BMB Reports* 55, no. 2 (February 2022): 57–64. https://doi.org/10.5483/BMBRep.2022.55.2.183.
44. Tan, Linda C, Alastair G Mowat, Chrysoula Fazou, Tim Rostron, Helen Roskell, P Rod Dunbar, Claire Tournay, et al. “Specificity of T Cells in Synovial Fluid: High Frequencies of CD8+ T Cells That Are Specific for Certain Viral Epitopes.” *Arthritis Research & Therapy* 2, no. 2 (February 7, 2000): 154. https://doi.org/10.1186/ar80.

45. Pacheco, Yovana, Yeny Acosta-Ampudia, Diana M. Monsalve, Christopher Chang, M. Eric Gershwin, and Juan-Manuel Anaya. “Bystander Activation and Autoimmunity.” *Journal of Autoimmunity* 103 (September 2019): 102301. https://doi.org/10.1016/j.jaut.2019.06.012.
46. Robinson, William H., and Lawrence Steinman. “Epstein-Barr Virus and Multiple Sclerosis.” *Science* 375, no. 6578 (January 21, 2022): 264–65. https://doi.org/10.1126/science.abm7930.
47. Forte, Eleonora, Zheng Zhang, Edward B. Thorp, and Mary Hummel. “Cytomegalovirus Latency and Reactivation: An Intricate Interplay With the Host Immune Response.” *Frontiers in Cellular and Infection Microbiology* 10 (2020): 130. https://doi.org/10.3389/fcimb.2020.00130.
48. Gugliesi, Francesca, Selina Pasquero, Gloria Griffante, Sara Scutera, Camilla Albano, Sergio Fernando Castillo Pacheco, Giuseppe Riva, Valentina Dell’Oste, and Matteo Biolatti. “Human Cytomegalovirus and Autoimmune Diseases: Where Are We?” *Viruses* 13, no. 2 (February 8, 2021): 260. https://doi.org/10.3390/v13020260.
49. Nagata, Keiko, and Kazuhiko Hayashi. “Epstein–Barr Virus Reactivation-Induced Immunoglobulin Production: Significance on Autoimmunity.” *Microorganisms* 8, no. 12 (November 27, 2020): 1875. https://doi.org/10.3390/microorganisms8121875.
50. Zahid, H. Jabran, Ruth Taniguchi, Peter Ebert, I-Ting Chow, Chris Gooley, Jinpeng Lv, Lorenzo Pisani, et al. “Large-Scale Statistical Mapping of T-Cell Receptor *β* Sequences to Human Leukocyte Antigens.” Immunology, April 2, 2024. https://doi.org/10.1101/2024.04.01.587617.
51. May, Damon H., Steven Woodhouse, H. Jabran Zahid, Rebecca Elyanow, Kathryn Doroschak, Matthew T. Noakes, Ruth Taniguchi, et al. “Identifying Immune Signatures of Common Exposures through Co-Occurrence of T-Cell Receptors in Tens of Thousands of Donors.” Immunology, March 27, 2024. https://doi.org/10.1101/2024.03.26.583354.
52. ElAbd, Hesham, Mitchell Pesesky, Gabriel Innocenti, Brian K. Chung, Aya KH Mahdy, Valeriia Kriukova, Laila Kulsvehagen et al. "T and B cell responses against Epstein–Barr virus in primary sclerosing cholangitis." *Nature Medicine* (2025):1-11.https://doi.org/10.1038/s41591-025-03692-w
53. Pesesky, Mitchell, Ramit Bharanikumar, Lionel Le Bourhis, Hesham ElAbd, Elisa Rosati, Cara L. Carty, Namita Singh et al. "Antigen-driven expansion of public clonal T-cell populations in inflammatory bowel diseases." *Journal of Crohn's and Colitis* 19, no. 7 (2025): jjaf048. https://doi.org/10.1093/ecco-jcc/jjaf048