# Supplementary material I

## I. Supplementary methods

### Immunohistochemistry and histology

FFPE tissue sections (4 µm) were pretreated to evoke antigen retrieval by cooking for 2 min in citrate buffer (pH 6.0).[1] For immunostaining the streptavidin-biotin-alkaline phosphatase method using FastRed as the chromogen was applied with the following antibodies: CD3, CD4, CD8 (all from DAKO, Glostrup, Denmark). The duodenal biopsies were reviewed by pathologists and villous atrophy was classified according to Marsh and colleagues with modifications introduced by Oberhuber.[2, 3] For quantification of CD3-positive T-cells in frozen tissue samples that were subjected to HTS, immunofluorescence was done as previously reported using an anti-CD3 antibody. The average T-cell content was 12.1% ± 1.2% (range 5.6% - 19.6%**; Figure S1**).

### DNA isolation

DNA was extracted from snap-frozen duodenal biopsies that were collected from the same patient in the identical upper GI endoscopy as for the FFPE-based immunohistology. To ensure highly comparable findings as in the FFPE material, neighboring sections of the frozen material were generated, H&E-stained and microscoped before DNA was isolated. For DNA isolation, the QIAamp Mini Kit (Qiagen, Hilden, Germany) was used according to the manufacturer's instructions. Quantification of the isolated DNA was done employing the NanoDrop ND-1000 spectrophotometer (NanoDrop, Wilmington, USA).

### Capillary electrophoresis

TCRβ gene rearrangements were amplified using the BIOMED-2 primer sets and protocols developed for PCR-based clonality analysis.[4] The fluorescence-labeled amplicons were size separated and detected via laser scanning by a 3130 Genetic Analyzer (Applied Biosystems; Darmstadt, Germany).

### Amplification and sequencing of TCRβ gene rearrangements

Amplification of TCRβ gene rearrangements was performed using 100 ng DNA and a multiplex PCR approach which includes 58 primers in the first PCR reaction. In a second PCR step universal Illumina adapter sequences including additional barcodes were added to the generated TCRβ amplicons for parallel sequencing on the Illumina MiSeq platform. Generated amplicons covered the entire CDR3 length as well as parts of the adjacent Vβ- and Jβ-segments.[5] Identical sequence reads and those differing only by one base pair in their CDR3 proportion were clustered and designated as unique clonotypes. Single reads and clonotypes consisting of <0.03% of all reads were eliminated since they were assumed to represent background signals.

The Vβ- and Jβ-segment usage of the clonotypes was assigned according to the international ImMunoGeneTics information system *(http://www.imgt.org)* based on the short sequence proportions created by the TCRβ multiplex PCR. Vβ- and Jβ-segment usage in RCD cases was determined by grouping according to the IMGT-segment families. To this end, the percentage of reads in the different patient groups were summed up for the respective Vβ- and Jβ-segments and normalized.

The *in silico* TCRβ-CDR3 sequence was defined as all amino acids (AA) starting from the conserved 5'cysteine in the Vβ-segment and ending at the conserved 3'phenylalanine in the Jβ segment. To determine the robustness and reliability of our HTS analysis from the duodenal mucosal tissue, we previously evaluated the TCRβ repertoire from peripheral blood samples of six healthy volunteers.[5]

**Peripheral blood T cells of healthy donors**

Peripheral blood mononuclear cells (PBMCs) were obtained by density gradient centrifugation (Ficoll, GE Healthcare UK) of whole blood samples of six healthy volunteers (refer to [5] for more details). PBMCs were stained with anti-CD4 FITC (SK3), anti-CD8 PE (SK1), anti-CD14 PerCP (MφP9, all clones BD Biosciences), and anti-TCRαβ (BW242/412, Miltenyi Biotec) for flow cytometric sorting. $CD4^+$ and $CD8^+$ T-cells were sorted into separate tubes achieving a purity of >98.0%. DNA was extracted from sorted T-cells using the Qiagen QlIPrep DNA/RNA Mini Kit (Qiagen, Hilden, Germany) and TCRβ HTS analysis were processed as described above.

**HLA typing by sequence-specific oligonucleotides (SSO)**

DNA isolated from intestinal mucosa was analyzed using the LABType® SSO kit (One Lambda, Inc), which hybridizes reverse sequence-specific oligonucleotide probes tagged with a unique color-coded microsphere to identify HLA class II alleles.[8] Target DNA was previously PCR-amplified using group-specific primers, biotinylated and tagged by R-Phycoerythrin-conjugated Streptavidin. After denaturation of the PCR product it is hybridized to complementary DNA probes conjugated to fluorescently code microspheres and detected by the Luminex® Flow Analyser. The assignment of HLA alleles is based on the reaction pattern of the various beads compared to patterns with known HLA alleles.

## II. Supplementary figures & tables

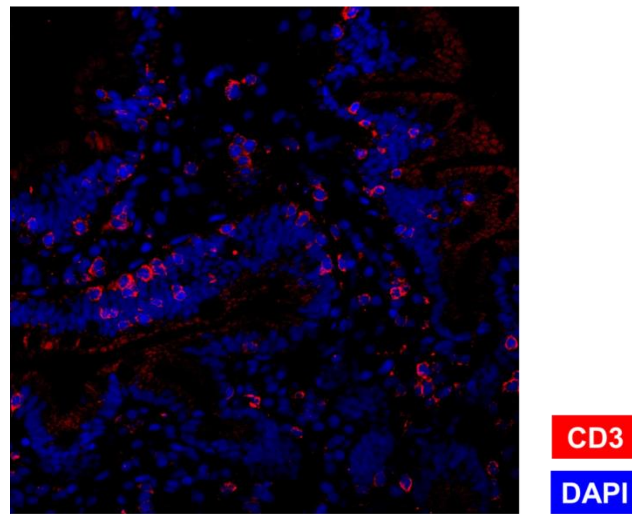



**Figure S1.** CD3 staining of a representative control small intestinal mucosa to determine percentage of T-cells. Red: CD3 (indirect immunofluorescence); blue: nuclear staining (DAPI).

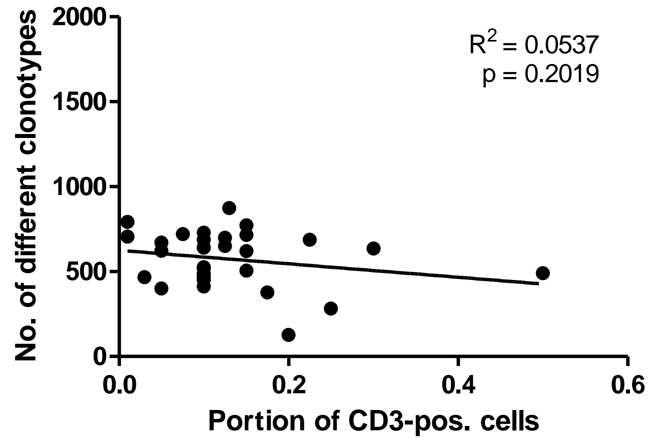
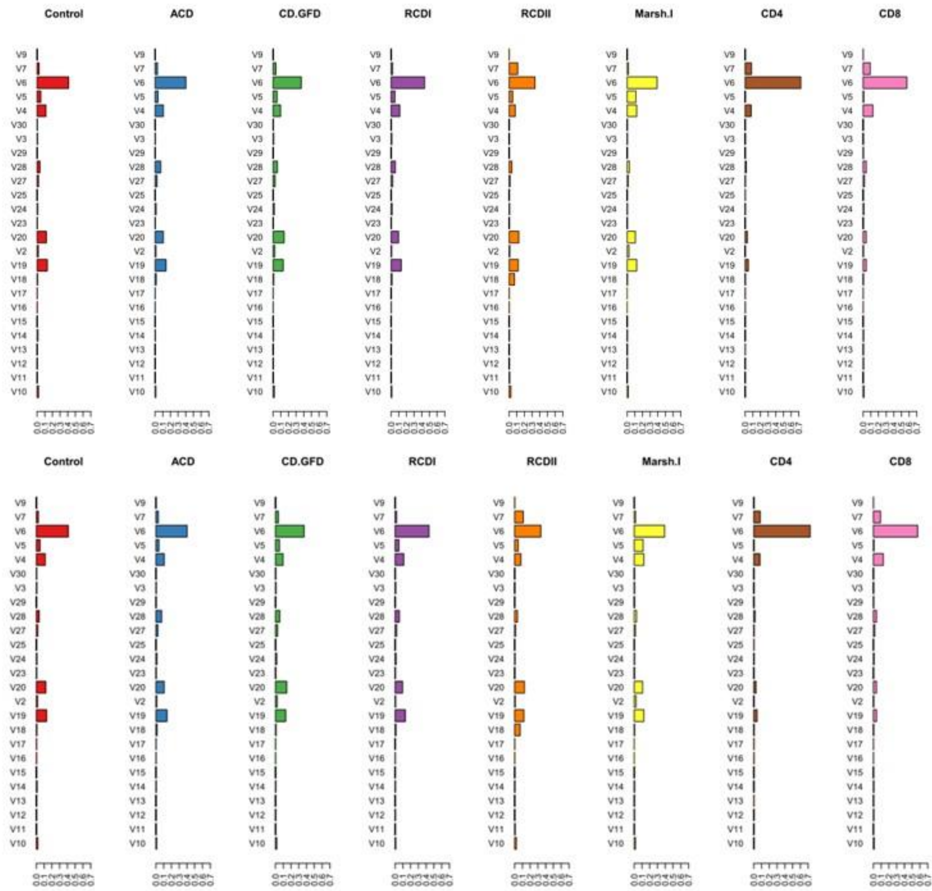



**Figure S2.** Correlation analysis of percentage of CD3-positive cells and number of clonotypes found in the individual samples. In this analysis we included data of 32 patients, distributed along all groups. The number of clonotypes found in the individual sample appears not to correlate with the percentage of CD3-positive T-cells in the sample (p=0.20).

A




B

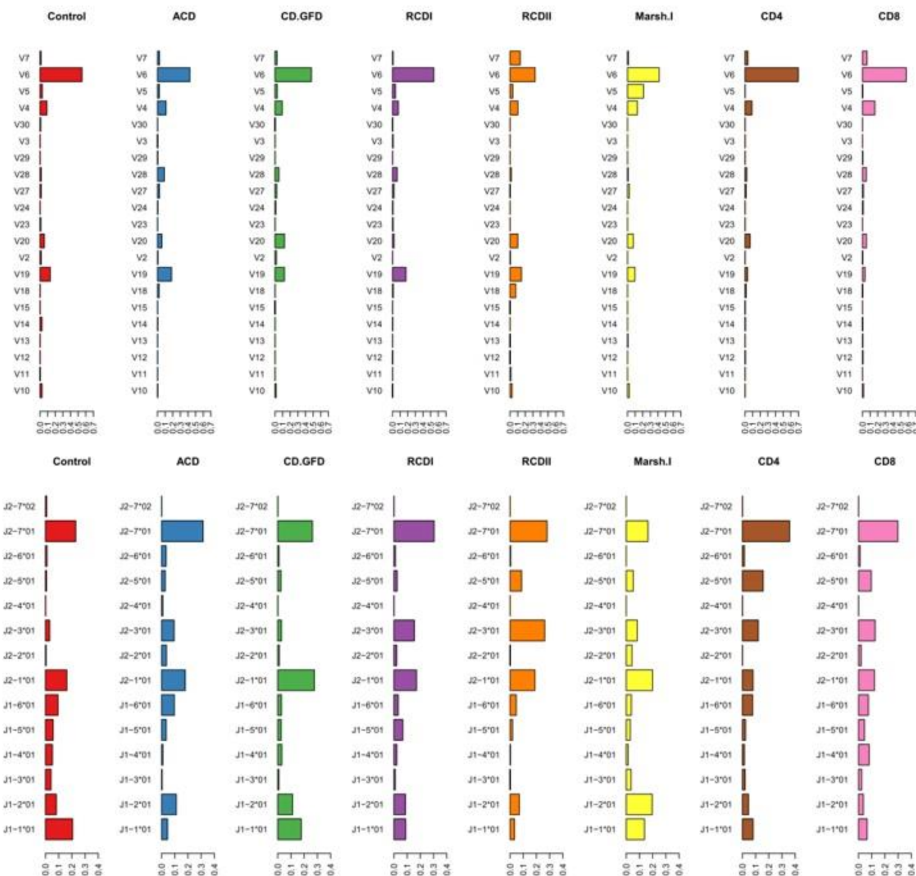


**Figure S3.** Analysis of V- and J-segment usage of all patients assigned to their respective CD groups. Firstly, (**A**) the read-percentages of all detected clonotypes identified by NGS were summed up for their respective V- and J-segments and were normalized. Then, (**B**) only read-percentages of the twenty most common clonotypes per patient were analyzed. Over all groups no different V- or J-segment usage was found applying multiple testing (Wilcoxon exact test and Bonferroni-Holm correction).

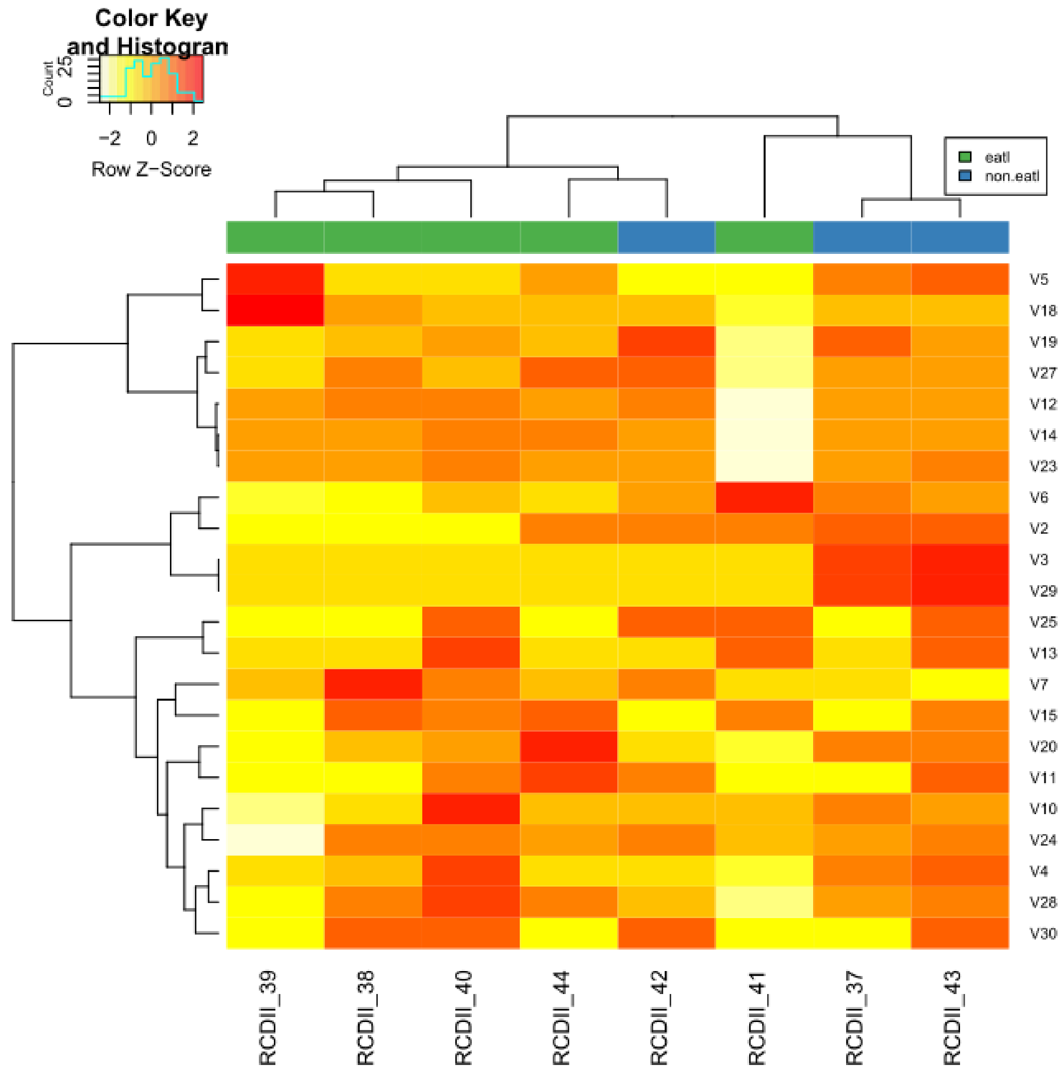



**Figure S4.** Heatmap of V-segment usage in patients with RCDII and RCDII/EATL. A hierarchical clustering using Ward´s algorithm grouped patients and V-Segments based on the similarity of their percentage of use. The read-percentages for every patient were summed up for the respective V-segments and were normalized to 1. The intensity of usage is color-coded, ranging from white (low usage) to red (high usage). The color key decodes the respective values. The histogram reveals the distribution of values.

Most common clonotype [%]

100, 80, 60, 40, 20, 0

0, 5, 10, 15, 20, 25

Tg-IgA titer

(measured value/upper limit of normal)

$R^2 = 0.0273$
$p = 0.2466$

**Figure S5.** Correlation analysis of serum transglutaminase-IgA titer (expressed as the measured Tg-IgA-titer divided by the upper limit of normal) and the frequency of the most dominant CDR3 clonotype. The frequency of the most common CDR3 clonotype did not correlate with the Tg-IgA-titer (p=0.25).

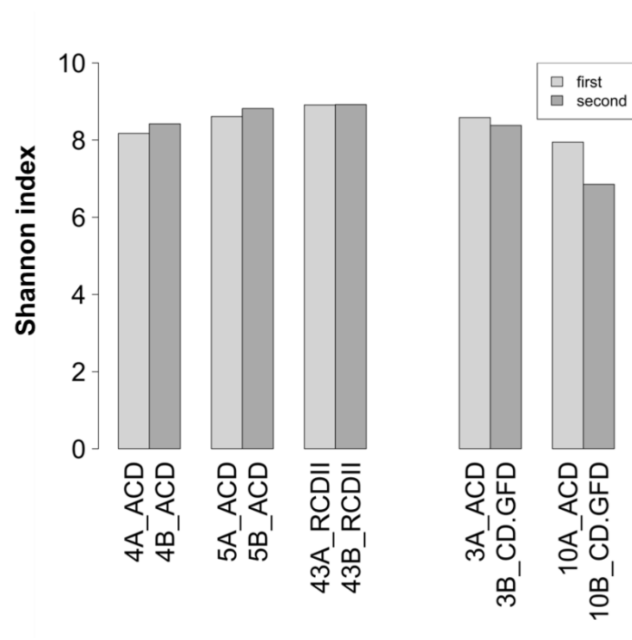



**Figure S6.** Stability of HTS-determined TCR variability over time. Shannon indices of TCR diversity as determined by CDR3-TCRβ-HTS in small intestinal mucosae in five patients who were examined at two independent points in time ("first" [lighter grey] and "second" [darker grey] examination).

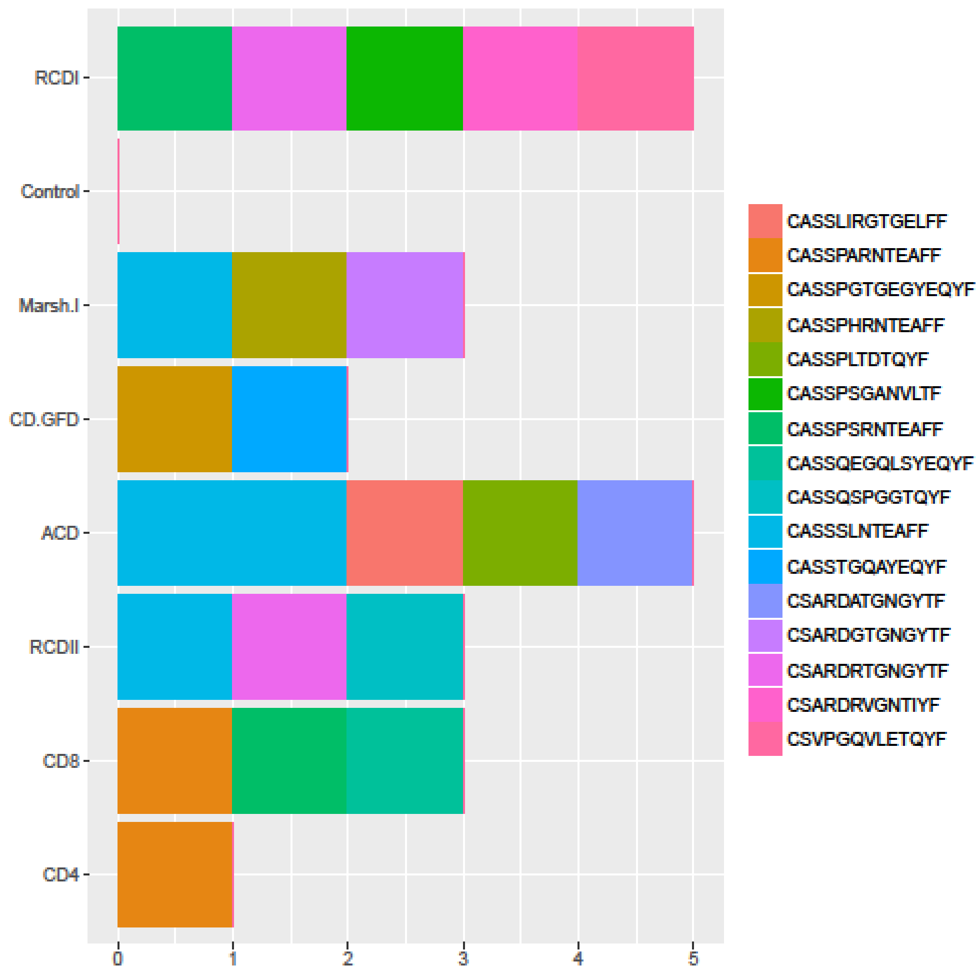



**Figure S7.** Public CMV and EBV TCRβ sequences in the small intestinal T-cells and peripheral blood (CD4+ and CD8+) T-cells. Alignment analysis of the CDR3 peptide sequence revealed homology with previously published public, EBV- and CMV-specific CDR3 sequences.[10, 11, 12] The numbers of public sequences are plotted group-specifically. Of 341 CDR3 sequences that were evaluated, 16 were identified in our data set, some of which occurred in multiple subjects (number of individuals with confirmed public TCR-CDR3 sequence: control=0, ACD=4, CD GFD=1, RCD-I=3, RCD-II=2, Marsh I=3, $CD4^+$ T-cells=1, $CD8^+$ T-cells =2). A detailed overview of the detected public virus-specific TCRβ sequences and their individual occurrence are given in Table S2.

**Table S1**

| Group | Patient-# | Clonotype | Percentage |
|---|---|---|---|
| ACD | 1 | ASSIRHTDTQY | 0.01171 |
| | | ASSLRSTDTQY | 0.00314 |
| | 2 | ASSLRSTDTQY | 0.00791 |
| | 5A | ASSLRATDTQY | 0.00691 |
| | | ASSIRSTDTQY | 0.00414 |
| | | ASSVRFTDTQY | 0.00412 |
| | | ASSIRFTDTQY | 0.00325 |
| | 5B | ASSIRFTDTQY | 0.00940 |
| | | ASSIRATDTQY | 0.00314 |
| | | ASSFRSTDTQY | 0.00275 |
| | | ASSVRFTDTQY | 0.00259 |
| | 7 | ASSIRSTDTQY | 0.00988 |
| | | ASSFRSTDTQY | 0.00848 |
| | 9 | ASSLRSTDTQY | 0.00936 |
| CD GFD | 11 | ASSFRSTDTQY | 0.00653 |
| | | ASSIRSTDTQY | 0.00306 |
| | | ASSIRATDTQY | 0.00194 |
| | 15 | ASSLRSTDTQY | 0.01553 |
| | | ASSFRSTDTQY | 0.00423 |
| | 17 | ASSLRFTDTQY | 0.03514 |
| | | ASSLRYTDTQY | 0.01249 |
| | | ASSLRATDTQY | 0.01224 |
| | 19 | ASSLRFTDTQY | 0.00148 |
| RCD I | 30 | ASSLRFTDTQY | 0.00587 |
| RCD II | 38 | ASSFRSTDTQY | 0.00689 |
| | 40 | ASSIRSTDTQY | 0.01638 |
| | | ASSLRTTDTQY | 0.00310 |
| | | ASSLRFTDTQY | 0.00229 |
| | 41 | ASSLRYTDTQY | 0.00343 |
| | 43A | ASSLRATDTQY | 0.00365 |
| | | ASSIRSTDTQY | 0.00334 |
| | 43B | ASSIRSTDTQY | 0.03698 |
| | | ASSRRSTDTQY | 0.01362 |
| | | ASSFRSTDTQY | 0.01291 |
| | | ASSIRATDTQY | 0.00135 |
| Marsh I | 45 | ASSLRSTDTQY | 0.00699 |
| $CD4^+$ T cells | P6 | ASSFRSTDTQY | 0.01184 |
| $CD8^+$ T cells | P2 | ASSLRTSDTQY | 0.00836 |
| | P4 | ASSLRATDTQY | 0.00269 |
| | P5 | ASSFRSTDTQY | 0.00340 |

**Table S2**

| Group | Patient-# | Clonotype | Percentage | Virus |
|---|---|---|---|---|
| ACD | 3A | CASSPLTDTQYF | 0.06567 | EBV |
| | 4B | CASSLIRGTGELFF | 0.09986 | EBV |
| | 5A | CASSSLNTEAFF | 0.05075 | EBV |
| | 5B | CASSSLNTEAFF | 0.03084 | EBV |
| | 6 | CSARDATGNGYTF | 0.05460 | EBV |
| CD GFD | 11 | CASSPGTGEGYEQYF | 0.18740 | EBV |
| | | CASSTGQAYEQYF | 0.03411 | EBV |
| RCD I | 34 | CSARDRVGNTIYF | 0.06824 | EBV |
| | 35 | CASSPSGANVLTF | 0.08804 | EBV |
| | | CASSPSRNTEAFF | 0.04952 | CMV |
| | 36 | CSVPGQVLETQYF | 0.07027 | EBV |
| | | CSARDRTGNGYTF | 0.05129 | EBV |
| RCD II | 37 | CASSSLNTEAFF | 0.05265 | EBV |
| | 40 | CASSQSPGGTQYF | 0.35366 | EBV |
| | | CSARDRTGNGYTF | 0.04558 | EBV |
| Marsh I | 45 | CASSPHRNTEAFF | 0.28147 | CMV |
| | 46 | CSARDGTGNGYTF | 0.03306 | EBV |
| | 47 | CASSSLNTEAFF | 0.04430 | EBV |
| CD4+ T cells | P2 | CASSPARNTEAFF | 0.04546 | CMV |
| CD8+ T cells | P2 | CASSPSRNTEAFF | 1.14998 | CMV |
| | | CASSPARNTEAFF | 0.21502 | CMV |
| | P6 | CASSQEGQLSYEQYF | 0.03127 | EBV |

**REFERENCES of Supplementary material**


1 Norton AJ, Jordan S, Yeomans P. Brief, high-temperature heat denaturation (pressure cooking): a simple and effective method of antigen retrieval for routinely processed tissues. The Journal of pathology 1994;**173**:371-9.

2 Marsh MN. Clinical and pathological spectrum of coeliac disease. Gut 1993;**34**:1740; author reply 1.

3 Oberhuber G, Caspary WF, Kirchner T, Borchard F, Stolte M. [Study Group of Gastroenterological Pathology of the German Society of Pathology. Recommendations for celiac disease/sprue diagnosis]. Zeitschrift fur Gastroenterologie 2001;**39**:157-66.

4 van Dongen JJ, Langerak AW, Bruggemann M, Evans PA, Hummel M, Lavender FL*, et al.* Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. Leukemia 2003;**17**:2257-317.

5 Ritter J, Seitz V, Balzer H, Gary R, Lenze D, Moi S*, et al.* Donor CD4 T Cell Diversity Determines Virus Reactivation in Patients After HLA-Matched Allogeneic Stem Cell Transplantation. American journal of transplantation : official journal of the American Society of Transplantation and the American Society of Transplant Surgeons 2015.

6 Broughton SE, Petersen J, Theodossis A, Scally SW, Loh KL, Thompson 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.

7 Qiao SW, Christophersen A, Lundin KE, Sollid LM. Biased usage and preferred pairing of alpha- and beta-chains of TCRs specific for an immunodominant gluten epitope in coeliac disease. International immunology 2014;**26**:13-9.

8 Trajanoski D, Fidler SJ. HLA typing using bead-based methods. Methods Mol Biol 2012;**882**:47-65.

9 Shannon CE. The mathematical theory of communication. 1963. MD computing : computers in medical practice 1997;**14**:306-17.

10 Miconnet I, Coste I, Beermann F, Haeuw JF, Cerottini JC, Bonnefoy JY*, et al.* Cancer vaccine design: a novel bacterial adjuvant for peptide-specific CTL induction. J Immunol 2001;**166**:4612-9.

11 Price DA, Brenchley JM, Ruff LE, Betts MR, Hill BJ, Roederer M*, et al.* Avidity for antigen shapes clonal dominance in CD8+ T cell populations specific for persistent DNA viruses. J Exp Med 2005;**202**:1349-61.

12 Lim A, Trautmann L, Peyrat MA, Couedel C, Davodeau F, Romagne F*, et al.* Frequent contribution of T cell clonotypes with public TCR features to the chronic response against a dominant EBV-derived epitope: application to direct detection of their molecular imprint on the human peripheral T cell repertoire. J Immunol 2000;**165**:2001-11.