Nature Medicine 2019; 25 (5): 734-737 (DOI: 10.1038/s41591-019-0403-9)
Post-print version

Distinct phenotype of CD4$^+$ T cells driving celiac disease identified in multiple autoimmune conditions

Asbjørn Christophersen[1,2,3,4], Eivind G Lund[1,2,3,14], Omri Snir[1,2,3,14], Elsa Solà[4,5], Chakravarthi Kanduri[1,6], Shiva Dahal-Koirala[1,2,3], Stephanie Zühlke[1,2,3], Øyvind Molberg[2,7], Paul J Utz[4], Mina Rohani-Pichavant[4], Julia F Simard[8], Cornelia L Dekker[9], Knut EA Lundin[1,2,10], Ludvig M Sollid[1,2,3,11,15] & Mark M Davis[4,12,13,15]

[1]KG Jebsen Coeliac Disease Research Centre, University of Oslo, Oslo, Norway.

[2]Institute of Clinical Medicine, University of Oslo, Norway.

[3]Department of Immunology, University of Oslo, Oslo, Norway.

[4]Institute for Immunity, Transplantation and Infection, Stanford University School of Medicine, Stanford, CA, USA.

[5]Liver Unit, Hospital Clínic Barcelona, University of Barcelona, IDIBAPS, Barcelona, Spain.

[6]Department of Informatics, University of Oslo, Oslo, Norway.

[7]Department of Rheumatology, Dermatology and Infectious Diseases, Oslo University Hospital, Oslo, Norway.

[8]Epidemiology, Health Research and Policy, Stanford School of Medicine, Stanford, CA, USA

[9]Department of Pediatrics, Stanford University School of Medicine, Stanford, CA, USA.

[10]Department of Gastroenterology, Oslo University Hospital, Oslo, Norway.

[11]Department of Immunology, Oslo University Hospital, Oslo, Norway.

[12]Department of Microbiology and Immunology, Stanford University School of Medicine, Stanford, CA, USA.

[13]The Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA, USA.

[14]These authors contributed equally to this work

[15]These authors contributed equally as last authors to this work and correspondence should be directed to them (l.m.sollid@medisin.uio.no and mmdavis@stanford.edu)

Nature Medicine 2019; 25 (5): 734-737 (DOI: 10.1038/s41591-019-0403-9)
Post-print version

ABSTRACT

Combining HLA-DQ:gluten tetramers with mass cytometry and RNA-seq analysis, we find that gluten-specific CD4$^+$ T cells in blood and intestines of celiac disease patients display a surprisingly rare phenotype. Cells with this phenotype are also elevated in patients with systemic sclerosis and systemic lupus erythematosus, suggesting a way to characterize CD4$^+$ T cells specific for disease-driving antigens in multiple autoimmune conditions.

MAIN TEXT

Celiac disease (CeD) is an HLA-DQ2/8-associated autoimmune enteropathy driven by the activation of gluten-specific CD4$^+$ T lymphocytes upon gluten consumption[1]. We combined gluten peptide-HLA-class II tetramers with a 43-parameter antibody panel for mass cytometry analysis (Extended Data Fig. 1-2, Supplementary Table 1). We found that cells binding these tetramers, representing five gluten peptides complexed to HLA-DQ2.5 (Supplementary Table 2), cluster within a surprisingly narrow subset of small intestinal CD4$^+$ T cells in HLA-DQ2.5$^+$ untreated CeD patients and comprise 0.3-1.5% of the total (Fig. 1a-b, participants: Supplementary Table 3). These gut T cells expressed multiple activation markers (CXCR3, CD38, CD161, CD28, HLA-DR, OX40) as well as CD39 and PD-1, suggestive of chronic activation, while being negative for the exhaustion marker KLRG1 (Fig. 1c-e, Supplementary Table 4, per donor in Extended Data Fig. 3). Importantly, the transcriptional profile of these tetramer positive CD4$^+$ gut T cells correlates highly with the surface marker expression (Fig. 1f, Extended Data Fig. 4).

Additionally, RNA-seq analysis demonstrated that CD200, CD84, CXCL13 and IL-21 are transcribed as well (Fig. 1g, complete list in Supplementary Table 5). These markers are

characteristic of follicular B-helper T (Tfh) cells, except that CXCR5 was not detectable on the surface of tetramer positive gut T cells (Fig. 1c-e), despite some transcription (Fig. 2f). Relevant to this, it was recently demonstrated that $CD4^+/PD\text{-}1^+/CXCR5^-$ cells, of unknown antigen specificity, are expanded in the synovium of seropositive rheumatoid arthritis (RA) patients and express a similar phenotype to what we report here, including expression of CD200, CXCL13, IL-21, PD-1, ICOS, OX40 and CD28[2]. The authors speculated that these cells induce plasma cell differentiation in the inflamed tissue. In general, T-cell induced plasma cell differentiation should show signs of proliferation and, with respect to the gluten-specific $CD4^+$ T cells analyzed here, the proliferation marker Ki-67 was expressed in blood (13-98%) but not in the gut (Fig. 1h-i). Conceivably, gluten-specific T cells in CeD can promote production of disease-specific antibodies to transglutaminase 2 and deamidated gluten peptides[3]. Our findings here, together with previous reports showing that these disease-specific gut plasma cells are negative for Ki-67[4,5], indicate that the disease-relevant T- and B cells initially interact and proliferate outside the celiac lesion. Once entering the gut, T cells may interact with plasma cells via the plasma-cell presentation of gluten T-cell epitopes[6] and influence the microenvironment. IL-21 is a key cytokine for plasma cells[7] and intraepithelial lymphocytes[8], both of which are increased in the celiac gut lesion[1,4].

While the relationship between lymphocytes in the blood versus those in tissues is frequently a question, here we find that gluten tetramer-binding T cells in blood of untreated CeD patients largely expresses the same pattern of markers as in the gut ($CXCR3^+$, $CD38^+$, $CD39^+$, $PD\text{-}1^+$, $HLA\text{-}DR^+$, $CD161^+$, $KLRG1^-$, $CD28^+$, $OX40^+$; Fig. 2a-e, per donor in Extended Data Fig. 5), except for being $CD69^-$. Further, despite Ki-67 expression (Fig. 1h-i), only a small fraction of the tetramer positive cells in blood expressed CXCR5 (confirmed by

FACS in Extended Data Fig. 6) and thus do not express a classical Tfh phenotype. As previously observed[9-11], the tetramer-binding cells were almost exclusively effector memory cells ($CD45RA^-$, $CD62L^-$), integrin-$\beta7^+$ and $CD38^+$.

It was recently reported that most gluten-specific cells express a Treg cell phenotype ($CD127^-/CD25^+/FoxP3^+$) after gluten exposure in vitro[12]. While confirming this finding (Extended Data Fig. 7a-b), our ex vivo analysis revealed that these cells are $CD137^{low}$, chiefly $CD25^-$ (Fig. 1 and 2) and negative for the Treg marker GARP (Extended Data Fig. 7c). And while some gluten-specific cells express FoxP3, these cells were $CD25^-$ (Extended Data Fig. 7d-g). Thus, gluten-specific T cells in vivo do not express a classical Treg phenotype.

We next asked whether antigenic stimulation drives these $CD4^+$ T cells. This involved a three-day oral gluten challenge in five CeD patients (previously on a gluten-free diet), which is known to mobilize preexisting clones of gluten-specific and gut homing T-cells into the blood on day six[10,13]. Upon challenge, these cells upregulated markers expressed by gluten-specific cells in the untreated celiac patients, including CD38, CD39, CXCR3, PD-1, ICOS, CD161, CCR5 and CD28 (Fig. 2f). These cells clustered in close proximity to tetramer-binding cells in untreated CeD, (Fig. 2g), differing chiefly by higher CCR5- and lower CD39-expression after the gluten challenge. Taken together, specific antigen-stimulation in vivo prompts gluten-specific T cells with an almost identical phenotype as those typical of untreated CeD.

To characterize the $CD4^+$ T-cells in patients with other autoimmune conditions, we performed mass cytometry analysis in PBMCs of patients with systemic sclerosis, systemic lupus erythematosus, together with CeD subjects and presumably healthy blood bank

donors (participants: Supplementary Table 6, antibody panel: Supplementary Table 7). We also included subjects suffering from acute influenza infection for comparison purposes. Unsupervised clustering of activated ($CD38^+$), memory ($CD45RA^-$) $CD4^+$ blood T cells showed that, unlike in CeD and the two other autoimmune conditions, the influenza-response was dominated by a $CD161^-/CD39^-$ subset (Fig. 2h), which faded with disease recovery and was very low in the other samples (Extended Data Fig. 8). We then tested whether an unbiased estimation (Extended Data Fig. 9) would report elevated levels of cells with the gluten-specific T-cell phenotype profile in these disease states. Strikingly, we found that 7/8 untreated CeD patients, 8/10 systemic sclerosis patients and 4/10 SLE patients had significantly elevated numbers of $CD4^+$ T cells with this phenotype compared to controls (Fig. 2i). Manual gating gave similar results (Extended Data Fig. 10), and we conclude that this subset is elevated in many patients with these types of autoimmunity. While 4/7 influenza infected individuals also showed elevated numbers of $CD4^+$ T cells with the phenotype displayed by gluten-specific cells, this was only a minor part of an influenza response (median <2% versus 20% constituted by the $CD161^-/CD39^-$ subset; Fig. 2h, Extended Data Fig. 8). It is nonetheless intriguing that $CD4^+$ T cells with the unique phenotype of gluten-specific cells are elevated not only in autoimmune conditions but also transiently during the acute phase of a viral infection. We speculate that these cells, unlike the $CD161^-/CD39^-$ cells, may represent self-antigen specific T cell clones that cross-react with influenza antigens, as suggested by the abundance of self-specific cells in healthy human beings[14] and their propensity for cross-reactivity[15].

In conclusion, CeD is the only human autoimmune disease in which the causative antigen is known, despite decades of effort in other systems. Here our results, combined with similar

findings in RA[2], strongly suggest that there is a distinct and relatively rare type of $CD4^+$ T lymphocytes that is common to multiple autoimmune disorders and transiently in at least one viral infectious disease. Since we know that most or all of the gluten-specific T cells are in this subset in CeD patients, it is reasonable to imagine that these cells might be the key disease-driving T cells in other autoimmune diseases as well.

ACKNOWLEDGEMENTS

We thank the patients participating in this study, S. Furholm, C. Hinrichs and M.H. Bakke for collecting patient material at the endoscopy unit (Oslo University Hospital – Rikshospitalet), the Stanford-LPCH Vaccine Program with S. Swope and S. Mackey for the study of influenza patients, A. Nau (Davis group), M. Leipold (The Human Immune Monitoring Center, Stanford University) and Brith Bergum (Flow Cytometry Core Facility, University of Bergen) for technical assistance with the Helios mass cytometers, B. Simonsen and S.R. Lund (Sollid group) for producing the biotinylated HLA-DQ2.5:gluten molecules, M.K. Johannesen (Sollid group) for lab technical assistance, K.J. Rolfsen (University of Oslo) for producing the cookies for gluten challenge, G.K. Sandve (University of Oslo) for critical inputs on RNA-seq data analysis in addition to V.K. Sarna (University of Oslo) and L. Chung (Stanford Medicine) for help in clinical assessment of patients. We also thank the Flow Cytometry Core Facility (Oslo University Hospital – Rikshospitalet and Radiumhospitalet) and the Stanford Shared FACS Facility for technical assistance. We express our gratitude to the funding bodies of this research: The University of Oslo Scientia Fellows program, co-funded by the University of Oslo World-leading research program on human immunology (WL-IMMUNOLOGY) (L.M.S) and by the EC FP7 Maria Sklodowska-Curie COFUND Programme (GA 609020) (A.C., L.M.S);

Stiftelsen KG Jebsen (project SKGJ-MED-017) (L.M.S., K.E.A.L.); The Unger-Vetlesen Medical Fund (A.C.); The U.S.-Norwegian Fulbright Foundation for educational exchange (A.C.); Fondsstiftelsen (Oslo University Hospital) (A.C.); The Howard Hughes Medical Institute (M.M.D.); The Simons Foundation (M.M.D.); The National Institutes of Health; U19-AI057229 (M.M.D.), U19-AI110491 (P.J.U.), UL1 TR001085 (C.L.D.) and R01 AI125197-01 (P.J.U.); The Donald E. and Delia B. Baxter Foundation (P.J.U.); The Henry Gustav Floren Trust (P.J.U.); and a gift from Elizabeth F. Adler (P.J.U.).

AUTHOR CONTRIBUTIONS

A.C., L.M.S. and M.M.D. conceptualized the study and drafted the manuscript with support from E.G.L. and O.S. A.C. developed the protocol for class II tetramer staining combined with mass cytometry, established the mass cytometry staining panels and performed the flow cytometry and most mass cytometry staining experiments. E.S. performed mass cytometry staining experiments on influenza samples and some autoimmune samples. O.S. and L.M.S. designed the RNA-seq study and O.S. prepared libraries for RNA seq. RNA-seq data was analyzed by E.G.L., C.K. and mass cytometry data by E.G.L., A.C., respectively. The CeD patient material was organized by K.E.A.L., S.D. and S.Z., and material from patients with autoimmune disorders other than CeD was organized by Ø.M., P.J.U., M.P. and J.F.S., material from patients during and after influenza infection was organized by C.L.D. Critical manuscript revisions were done by E.S., C.K., S.D., S.Z., Ø.M., P.J.U., M.P., J.F.S, C.L.D., K.E.A.L.

Nature Medicine 2019; 25 (5): 734-737 (DOI: 10.1038/s41591-019-0403-9)
Post-print version

COMPETING FINANCIAL INTERESTS

The authors have no financial conflicts of interest.

172 **REFERENCES (MAIN TEXT ONLY)**

173 1. Abadie, V.*, et al. Annu Rev Immunol* **29**, 493-525 (2011).

174 2. Rao, D.A.*, et al. Nature* **542**, 110-114 (2017).

175 3. Sollid, L.M.*, et al. Gut* **41**, 851-852 (1997).

176 4. Di Niro, R.*, et al. Nat Med* **18**, 441-445 (2012).

177 5. Steinsbo, O.*, et al. Nat Commun* **5**, 4041 (2014).

178 6. Hoydahl, L.S.*, et al. Gastroenterology* (2018).

179 7. Moens, L.*, et al. Front Immunol* **5**, 65 (2014).

180 8. Kooy-Winkelaar, Y.M.*, et al. Proc Natl Acad Sci U S A* **114**, E980-e989 (2017).

181 9. Christophersen, A.*, et al. United European Gastroenterol J* **2**, 268-278 (2014).

182 10. Raki, M.*, et al. Proc Natl Acad Sci U S A* **104**, 2831-2836 (2007).

183 11. du Pre, M.F.*, et al. Am J Gastroenterol* **106**, 1147-1159 (2011).

184 12. Cook, L.*, et al. J Allergy Clin Immunol* **140**, 1592-1603 (2017).

185 13. Risnes, L.F.*, et al. J Clin Invest* (2018).

186 14. Yu, W.*, et al. Immunity* **42**, 929-941 (2015).

187 15. Su, L.F.*, et al. Immunity* **38**, 373-383 (2013).

188

189

190

FIGURE LEGENDS (MAIN TEXT ONLY)

Figure 1. Distinct non-proliferative phenotype of gluten-specific CD4$^+$ gut T cells. (a) HLA-DQ2.5:gluten tetramer (tet)-staining in an untreated celiac disease (UCeD) patient and a control subject with mass cytometry (Fig. 1a-e: *n* = 6 UCeD patients, 7 controls, five experiments). **(b)** t-SNE plots of total CD4$^+$ gut T cells in a UCeD patient and a control subject. **(c)** Expression of proteins on tet$^+$/tet$^-$ CD4$^+$ gut T cells in an UCeD patient and **(d)** summarized for tet$^+$ cells in the 6 merged UCeD patients. **(e)** Mass cytometry-derived and **(f)** RNA seq-derived log2 fold-change expression of indicated markers in tet$^+$ versus tet$^-$ CD4$^+$ gut T cells of UCeD patients and compared to CD4$^+$ gut T cells of control subjects (Ctr.) (f-g: *n* = 5 UCeD patients, 4 control subjects, two experiments). **(g)** RNA-seq derived log2 fold-change expression of indicated markers in CD4$^+$ gut T cells, differentially expressed in tet$^+$ versus tet$^-$ and versus CD4$^+$ gut T cells in ctr. **(h)** Flow cytometry-derived Ki-67$^-$expression in tet$^+$/tet$^-$ CD4$^+$ blood and gut T cells of a gluten-challenged and untreated CeD patient, respectively **(i)** summarized for five gut samples, seven blood samples (four experiments).

Figure 2. Distinct, antigen-induced phenotype of gluten-specific CD4$^+$ blood T cells and occurrence of similar subset in other immune conditions. (a) HLA-DQ2.5:gluten tetramer (tet)-staining with mass cytometry for an untreated celiac disease (UCeD) patient and control subject (ctr.) (Fig. 2a-e: *n* = 7 UCeD patients, 10 ctr., nine experiments). **(b)** t-SNE plots with CD4$^+$ blood T cells of an UCeD patient and ctr. **(c)** Expression of proteins on tet$^+$/tet$^-$ CD4$^+$ blood T cells of an UCeD patient. **(d)** Heat map with absolute expression (staining intensity) of tet$^+$ cells and **(e)** log2 fold change for tet$^+$ versus pre tetramer-

enriched $CD4^+$ T cells in seven UCeD patients and versus CD4+ T cells of 10 ctr. **(f)** Log2 fold-change expression of indicated markers for tet^+ $CD4^+$ blood T cells after versus before gluten challenge of five treated CeD (TCeD)) patients (Fig. 2f-g, three experiments) and versus tet^+ cells of seven UCeD subjects (same UCeD in 2d, e). **(g)** t-SNE plot with tet^- and tet^+ cells in a TCeD subject before and following gluten challenge compared to tet^+ of an UCeD subject. **(h)** t-SNE plots and unsupervised clustering of activated ($CD38^+$) memory ($CD45RA^-$) $CD4^+$ blood T cells in indicated participant groups (*n* = 5 distinct samples in each group) and tet^+ cells of seven UCeD patients. Cluster 1, containing 75% of tet^+ cells from UCeD patients, and cluster 2, upregulated in subjects with influenza infection (Extended Data Fig. 8) are color-coded. **(i)** Unbiased prevalence estimate of tet^+ cell phenotype profile in UCeD patients among indicated diseases (19 experiments) using a supervised classification model (Extended Data Fig. 8). P-values calculated with unpaired, two-tailed t-test. Median frequency, interquartile range and max/min whiskers shown. Systemic lupus erythematosus (SLE).

ONLINE METHODS

Human material

All participants gave informed written consent. We obtained patient material from the endoscopy unit and the Rheumatology Department at Oslo University Hospital, from the Immunology and Rheumatology Division at the Department of Medicine and influenza patient material from the Emergency Department and the Express Outpatient Clinic at Stanford Hospital. All CeD patients were HLA-DQ2.5$^+$ (i.e., *DQA1**05 and *DQB1**02) or HLA-DQ8$^+$ (i.e., *DQA1*03* and *DQB1*03:02*) and diagnosed according to the guidelines of the British Society of Gastroenterology[16]. The studies on patient material obtained from subjects examined at Oslo University Hospital during routine follow-up were approved by the Regional Committee for Medical and Health Research Ethics South-East Norway (2010/2720). Treated CeD patients who were challenged with gluten received one in-house produced cookie containing 10g-enriched flour (Validus AS) each day for three days and blood samples were taken on day six after gluten challenge when a peak in the frequency of gluten-specific CD4$^+$ blood T cells was expected[10,17] (Regional Committee for Medical and Health Research Ethics South-East Norway, 2013/1237, Clinicaltrials.gov identifier NCT02464150). Blood samples from patients during and after influenza virus infection were obtained from a cohort of patients recruited from individuals with influenza-like symptoms attended at the Emergency Department or the Express Outpatient Clinic at Stanford Hospital. The study was approved by the Stanford University Administrative Panels on Human Subjects in Medical Research and covered by IRB 22442 (Immune Responses to Influenza-like Illness). Patients who tested positive for influenza A virus through a nasopharyngeal swab test (analyzed at the Virology Lab at Stanford Hospital) were included. All the included participants also tested negative with the same swab test for influenza B

virus, parainfluenza 1, 2 and 3 viruses, metapneumovirus and rhinovirus. Included participants were examined again 23 and 41 days after their initial medical examination and inclusion. One of the seven included patients did not donate blood at this second consultation. Influenza-associated symptoms of participants from the influenza cohort were documented on a patient diary and were evaluated by a research nurse at inclusion and during the follow-up visit. The definition of infection recovery was based on the resolution of influenza-like symptoms at the follow-up visit. Our study cohort of patients with autoimmune disorders other than CeD did not receive immunomodulating treatment at the time of blood draw and met classification criteria for systemic sclerosis[18] or systemic lupus erythematosus[19], respectively. The recruitment of these patients were covered by Regional Committee for Medical Research Ethics in South-East Norway (2016/119) and IRB 14734 (Stanford University Immunological and Rheumatic Disease Database: Disease Activity and Biomarker Study). Buffy coats were obtained from anonymous blood donors at the Stanford blood center or Oslo University Hospital (blood bank).

We isolated PBMCs through density gradient centrifugation (Lymphoprep; Axis Shield). Duodenal biopsies were treated 2x10 minutes with 2 mM EDTA + 2% fetal calf serum (FCS) in PBS at 37°C to remove epithelial layer prior to further digestion with collagenase (1 mg/ml) in 2% FCS in PBS at 37°C for 60 min. The samples were then homogenized using a 1.2 mm syringe and filtered through a 40 or 70 µm cell strainer to obtain single-cell suspensions. All samples were cryopreserved.

HLA-class II tetramer staining and mass cytometry

The protocol established here was partially derived from protocols on combination of HLA class-I tetramers and mass cytometry[20,21]. We thawed the frozen cell samples in 20% fetal calf serum (FCS) in RPMI and washed the cells in 10% FCS with Benzonase (Sigma-Aldrich/Merck, 1:10 000) in RPMI before resuspending and counting the cells in CyFACS buffer (0.1% bovine serum albumin, 2mM EDTA, 0.05% sodium azide in PBS). After 450g centrifugation, cells were treated with 1:10 diluted FcR block (Miltenyi Biotek), stained with anti-CD11c, anti-CD14 and 5 µg/ml purified anti-CD32 (clone FUN-2) to reduce nonspecific tetramer binding, and barcoded with anti-CD45 coupled with 89Y or 108Pd[22] in 200 µl CyFACS buffer. Only names and staining concentrations of monoclonal antibodies not listed in Supplementary Tables 1 and 7 are specified here. After one wash step, the samples from CeD patients were stained for 40 minutes at room temperature with HLA-DQ2.5:gluten tetramers representing the five different disease-relevant and immunodominant gluten T-cell epitopes[23] DQ2.5-glia-α1a, DQ2.5-glia-α2, DQ2.5-glia-ω1, DQ2.5-glia-ω2 and DQ2.5-hor3. (Supplementary Table 3) at 15 µg/ml each in 200 µl CyFACS buffer (BBMCs) or 100 µl (biopsy-derived single-cell suspensions). We also added tetramers representing HLA-DQ2.5:CLIP2 at a 20 µg/ml concentration in order to exclude tetramer background staining (Extended Data Fig. 1d). HLA-DQ2.5:gluten and HLA-DQ2.5:CLIP2 molecules were produced as previously described[24] and, two hours prior to cell staining, multimerized on PE-Cy7-coupled streptavidin or APC-Cy7-coupled streptavidin, respectively (Thermo Fisher Scientific). The cells were washed, and tetramer-binding cells were metal-tagged with 1.25 µl anti-PE and 1.25 µl anti-phycocyanin for 20 minutes on ice in 100 CyFACS buffer followed by another wash step. To facilitate tetramer enrichment, the PBMCs of CeD patients were resuspended in 50 µl anti-Cy7 metal beads with 150 µl CyFACS buffer and incubated for 20

minutes on ice (combined anti-Cy7 enrichment and anti-PE staining was established with the T-cell clone TCC1214P.A.27, derived from blood of CeD patient[9], and is visualized in Fig. 1c). The cells were washed and 2% of the PBMCs of the CeD patients (pre-tetramer enriched sample) were removed and added to one million CD45-barcoded carrier cells of a healthy donor to reduce cell loss and left on ice until further staining. The remaining PBMCs of CeD patients were enriched for tetramer-binding cells on a magnetized LS column (Miltenyi Biotec). We then added one million CD45-barcoded PBMCs from a healthy donor to the tetramer-enriched sample (and to the biopsy-derived single-cell suspensions that had not undergone tetramer enrichment), washed the cells x1 before all samples were stained for 20 minutes on ice with a panel of metal coupled antibodies (Supplementary Table 1 or in the case of participants included in Extended Data Fig. 10e-f; Supplementary Table 7). After one wash step, the cells were stained for five minutes at room temperature with Cisplatin (Fluidigm) at 1/1500 concentration and washed before overnight incubation at 4°C with 1:1000 diluted 125 µM DNA intercalator in Maxpar Fix and Perm Buffer (Fluidigm). The following day we washed the cells in CyFACS buffer, PBS and milli-Q water (1x each) before they were analyzed in milli-Q water at a Helios instrument (Fluidigm). Unlike in the gut samples analyzed here and in previous studies on gluten-specific cells in blood using flow cytometry[9,25], we have not specified the frequency of tetramer-binding cells in blood analyzed with mass cytometry as washing, resuspension in water and mass cytometer tubing considerably reduced the number of cells (including tetramer-binding cells) in the tetramer-enriched sample relative to the total number of $CD4^+$ T cells in the sample.

Prior to the establishment of the protocol, we also used fluorescein-coupled streptavidin (Biolegend) and anti-fluorescein 160Gd (Fluidigm) (Extended Data Fig. 1a) and the gluten-

specific T-cell clone TCC1030.43 derived from blood of CeD patients[9] to determine which fluorophore generated the best staining intensity through secondary metal-tagged antibody staining. In each experiment we stained a gluten-specific T-cell clone with the corresponding HLA-DQ2.5:gluten tetramer as a positive control for tetramer staining with mass cytometry.

Flow cytometric analysis

We prepared and stained CeD blood and biopsy material including T-cell clones with HLA-DQ2.5:gluten tetramers and surface markers according to protocols described elsewhere[9,13]. One CeD patient analyzed with flow cytometry was HLA-DQ8^{+}/HLA-DQ2.5^{-} and for this subject we used HLA-DQ8:gluten tetramers representing the two gluten epitopes HLA-DQ8-glia-α1 and HLA-DQ8-glia-γ1b[26]. Tetramer-sorted cells were cultured in vitro as previously described[27]. Staining for Ki-67 and FoxP3 was performed according to the manufacturer's protocol (Thermo Fischer Scientific's eBioscience FoxP3/Transcription factor staining buffer set). Antibodies used for flow cytometry staining are listed in Supplementary Table 8. The cells were analyzed with a LSR II instrument or sorted on a FACS Aria II instrument (BD Bioscience).

RNA seq analysis

Single cell suspension of duodenal biopsies from five CeD patients and four healthy subjects (Supplementary Table 3) were stained with PE-conjugated HLA-DQ2.5:gluten tetramers representing four immunodominant T-cell epitopes of gluten: DQ2.5-glia-α1a, DQ2.5-glia-α2, DQ2.5-glia-ω1 and DQ2.5-glia-ω2 (Supplementary Table 2)[26] as previously described[28]. Following tetramer staining, the cells were labeled with anti-CD3 BV570 (Biolegend), anti-CD4 APC-H7 (BD Biosciences), anti-CD14 Pacific Blue (Biolegend), anti-CD11c Horizon V450

(BD Bioscience), anti-CD27 PE-Cy7 (eBioscience), IgA FITC (Southern Biotech) and Live/Dead Fixable Violet Dead Cell Stain (Thermo Fischer Scientific). See also the Life Sciences Reporting Summary for more details on the antibodies used. We added anti-CD27 and anti-IgA due to a parallel study on a different cell subset. HLA-DQ2.5:gluten tetramer positive and tetramer negative $CD4^+$ T cells were sorted in two separate tubes using FACS Aria II (BD Bioscience). RNA was extracted using RNeasy micro kit (Qiagen) and quantified on 2100 Bioanalyzer using a RNA 6000 Pico kit (Agilent Technologies).

Approximately 90 ng of RNA was used for cDNA synthesis and amplification. cDNA synthesis was performed at 42°C for 90 min and 70°C for 10 min and followed by amplification 95°C, 1 min; [98°C, 10 sec; 65°C, 30 sec; and 68°C, 3 min] 15x cycles and 72°C, 10 min using SMARTer® Ultra® Low Input RNA Kit for Sequencing - v3 (Clontech Laboratories). Amplified cDNA was quantified using the High Sensitivity DNA Kit (Agilent Technologies). Tagmentation and adapter ligation were achieved using Nextera XT library preparation kit (Illumina, Inc). Amplicon libraries were sequenced on NextSeq500 (Illumina, Inc) at the Norwegian Sequencing Center (http://www.sequencing.uio.no).

Statistics and data analysis

Both mass cytometry and flow cytometry data were analyzed with *FlowJo* version 10.4 (FlowJo LLC) for visualization of data in two-parametric 2D-plots (Fig. 1a, 1c, Fig. 2a, 2c, 2h, Extended Data Fig. 1a-d, Extended Data Fig. 2, Extended Data Fig. 6, Extended Data Fig. 7a-b, 7d, 7f and Extended Data Fig. 10a) and for cell quantifications (Extended Data Fig. 7e, 7g, Extended Data Fig. 10b-f). We used the *GraphPad Prism 7* software (GraphPad Software, Inc) for statistical analysis and visualization of cell frequencies (Fig. 1i, Extended Data Fig. 7e,

g and Extended Data Fig. 10b-f). Here we applied an unpaired, two-tailed t-test (Extended Data Fig. 10b-c, 10e (median frequency and interquartile range indicated)), or a paired, two-tailed t-test (Extended Data Fig. 10d, 10f) to calculate statistical significance. We also used FlowJo to exclude cells that were not $CD4^+$ T cells (gating strategy in Extended Data Fig. 2) before exporting the fcs-files, containing only $CD4^+$ blood or gut T cells, for generation of t-SNE-plots (t-Distributed Stochastic Neighbor Embedding)[29] and all other analysis presented in Fig. 1d-g, 2d-i, Extended Data Fig. 3-5, and Extended Data Fig.8-9. The markers used to generate the t-SNE plots in Fig. 1b; Fig. 2b, 2g, Extended Data Fig. 3c and Extended Data Fig. 5c (31 markers both in gut and blood samples, which did not include the marker for tetramer staining (165Ho anti-Phycoerythrin)) are identified with one asterisk in Supplementary Table 1.

Mass cytometry data (the fcs-file containing only blood or gut $CD4^+$ T cells) was loaded into *R* using the *flowCORE* package. Here the aggregate marker intensity (Fig. 1d, Fig. 2d and Extended Data Fig. 8b) was computed as the grand mean of each donors mean marker intensity. Mass cytometry fold change (Fig. 1e, Fig. 2e, 2f, and Extended Data Fig. 3a, Extended Data Fig. 5a, Extended Data Fig. 7c, Extended Data Fig. 8c) was computed as the log2 fold change of the aggregate marker intensity. Heat maps, to visualize the aggregate marker intensity and the log2 fold change, were generated using *ggplot2*, and t-SNE plots were generated using the *Rtsne* package. For t-SNE plots, boxplots (Extended Data Fig. 3b, Extended Data Fig. 5b), supervised classification (Fig. 2i, Extended Data Fig. 9) and fold change significance testing (Supplementary Table 4), the raw mass cytometry intensity values were first transformed using the inverse hyperbolic sine, as described by *Nowicka et al*[30]. In the generated boxplots in Extended Data Fig. 3b and Extended Data 5b (generated

with *ggplot2)*, the Y-axis indicates arcsinh-transformed intensity and the boxes show median frequency and interquartile range. Whiskers show largest/smallest value below 1.5 times the interquartile range. In Supplementary Table 4, we used a paired, two-tailed t-test to calculate significant differences in mean marker intensity between tetramer positive and tetramer negative cells from CeD patients in blood and gut. In the same table, we performed an unpaired, two-tailed t-test for all other comparisons where the test conditions were from unmatched donors (e.g. CeD patients versus healthy controls). P-values were corrected for multiple testing using the Benjamini–Hochberg procedure.

In Figure 2h and Extended Data Fig. 8, we did unsupervised clustering of activated ($CD38^+$) memory ($CD45RA^-$) $CD4^+$ T cells using the *FlowSOM* and *ConsesusClusterPlus* packages. To avoid introducing bias to the clustering and the t-SNE visualization, we had a balanced number of cells and samples per disease group. Thus, we randomly selected five samples per disease, except for gluten challenge where we only had four samples with sufficient cells. Furthermore, we sampled at most 3707 activated cells per sample, which is the median number of cells per sample, and used these cells for clustering. For t-SNE visualization in figure 2h, we subsampled 807 cells per sample, which is the number of activated cells in the smallest sample. We visualized the prevalence of cells within the two clusters (cluster 1 and cluster 2) in a boxplot (Extended Data Fig. 8a) indicating median frequency and interquartile range. Here the whiskers show largest/smallest value below 1.5 times the interquartile range and single data points depict outliers. The markers used to generate the t-SNE plot in Fig. 2h and Extended Data Fig. 8 are listed in Extended Data Fig. 8b-c.

We trained a supervised classification model on tetramer positive and tetramer negative $CD4^+$ T cells from tetramer-enriched PBMC samples from untreated CeD patients (Fig. 2i, with a diagram illustrating the workflow in Extended Data Fig. 9). The model was subsequently used to obtain an unbiased prevalence estimate of $CD4^+$ T cells with a phenotype highly similar to the gluten specific $CD4^+$ T cells in all included blood samples analyzed with mass cytometry (excluding the tetramer-enriched samples that were only used to train the model). More specifically, we used 10-fold cross validation with three repeats to train a random forest model[31] using *caret* version 6.0-79. The optimal *mtry* parameter for the data was selected with a grid search between one and the total number of markers divided by three. Log loss was used as a metric to select the optimal model. The *doMC* package, version 1.3.5, was used to parallelize model training. We used the *GraphPad Prism 7* software to visualize the prevalence estimates in a boxplot (Fig. 2i), which shows median frequency and interquartile range, while the whiskers indicate max/min values. Here P-values (each participant group versus the group of healthy controls) were calculated using an unpaired, two-tailed t-test. The markers used to generate the prediction model in Fig. 2i (the 22 $CD4^+$ T-cell markers that were common to the two mass cytometry staining panels in Supplementary Table 1 and Supplementary Table 7) are identified with two asterisks in Supplementary table 7). The *importance* function of the *randomForest* package was used to extract the mean decrease Gini score from the final model. A high scoring parameter is important to the model and a low scoring value is less relevant. This Gini score is visualized in Extended Data Fig. 9b using *ggplot2*.

RNA-seq reads (76 bp paired end) were mapped to the human reference genome GRCh38.p7 containing alternative loci with gene annotations curated by *Ensembl* release 86

using *Salmon*[32] version 0.7.2 for mapping with parameters *-l UI --useVBOpt --numBootStraps 30 --seqBias --gcBias*. The quasi-mapping index in *Salmon* was built using a default k-mer length of 31. Read counts of transcripts (including those on alternative loci) were aggregated to gene-level. The raw sequencing data were processed on a secure computing platform; the TSD (Tjeneste for Sensitive Data) facilities owned by the University of Oslo, operated and developed by the TSD-service group. Further data processing was performed using *R* version 3.2 with the *Bioconductor* version 3.4 and the *Tidyverse* version 1.2.1 collection of packages. Estimated gene counts were loaded into *R* using *Tximport*. Gene differential expression analysis and log fold change estimation (Fig. 1f-g, Extended Data Fig. 7c) was computed using *DESeq2*[33] with a design formula controlling for sample donor. A full list of the differentially expressed genes is listed with adjusted P values in Supplementary Table 5. Here we used a significance threshold of 5e-3 after adjusting for multiple testing. Heat maps, to visualize the the log2 fold change, were generated using *ggplot2*, as with fold change expression in the mass cytometry data.

Further information on methods, statistics, data analysis is provided in the Life Sciences Reporting Summary.

Data availability

The raw sequences of the RNA-seq data are deposited at the EGA European Genome Phenome Archive (https://ega-archive.org) under accession number EGAS00001003017. All other data supporting the findings of this study are available from the authors upon request.

ONLINE METHODS-ONLY REFERENCES

16. Ludvigsson, J.F.*, et al. Gut* **63**, 1210-1228 (2014).

17. Anderson, R.P.*, et al. Nat Med* **6**, 337-342 (2000).

18. van den Hoogen, F.*, et al. Ann Rheum Dis* **72**, 1747-1755 (2013).

19. Hochberg, M.C. *Arthritis Rheum* **40**, 1725 (1997).

20. Newell, E.W.*, et al. Immunity* **36**, 142-152 (2012).

21. Newell, E.W.*, et al. Nat Biotechnol* **31**, 623-629 (2013).

22. Mei, H.E.*, et al. J Immunol* **194**, 2022-2031 (2015).

23. Shan, L.*, et al. Science* **297**, 2275-2279 (2002).

24. Quarsten, H.*, et al. J Immunol* **167**, 4861-4868 (2001).

25. Christophersen, A.*, et al. J Immunol* **196**, 2819-2826 (2016).

26. Sollid, L.M.*, et al. Immunogenetics* **64**, 455-460 (2012).

27. Molberg, Ø.*, et al. Methods Mol Med* **41**, 105-124 (2000).

28. Bodd, M.*, et al. Eur J Immunol* **43**, 2605-2612 (2013).

29. van der Maaten, L.*, et al. J Mach Learn Res* **9**, 2579-2605 (2008).

30. Nowicka, M.*, et al. F1000Res* **6**, 748 (2017).

31. Andy, L.*, et al. R News* **2**, 18-22 (2002).

32. Patro, R.*, et al. Nat Methods* **14**, 417-419 (2017).

33. Love, M.I.*, et al. Genome Biol* **15**, 550 (2014).

Figure 1

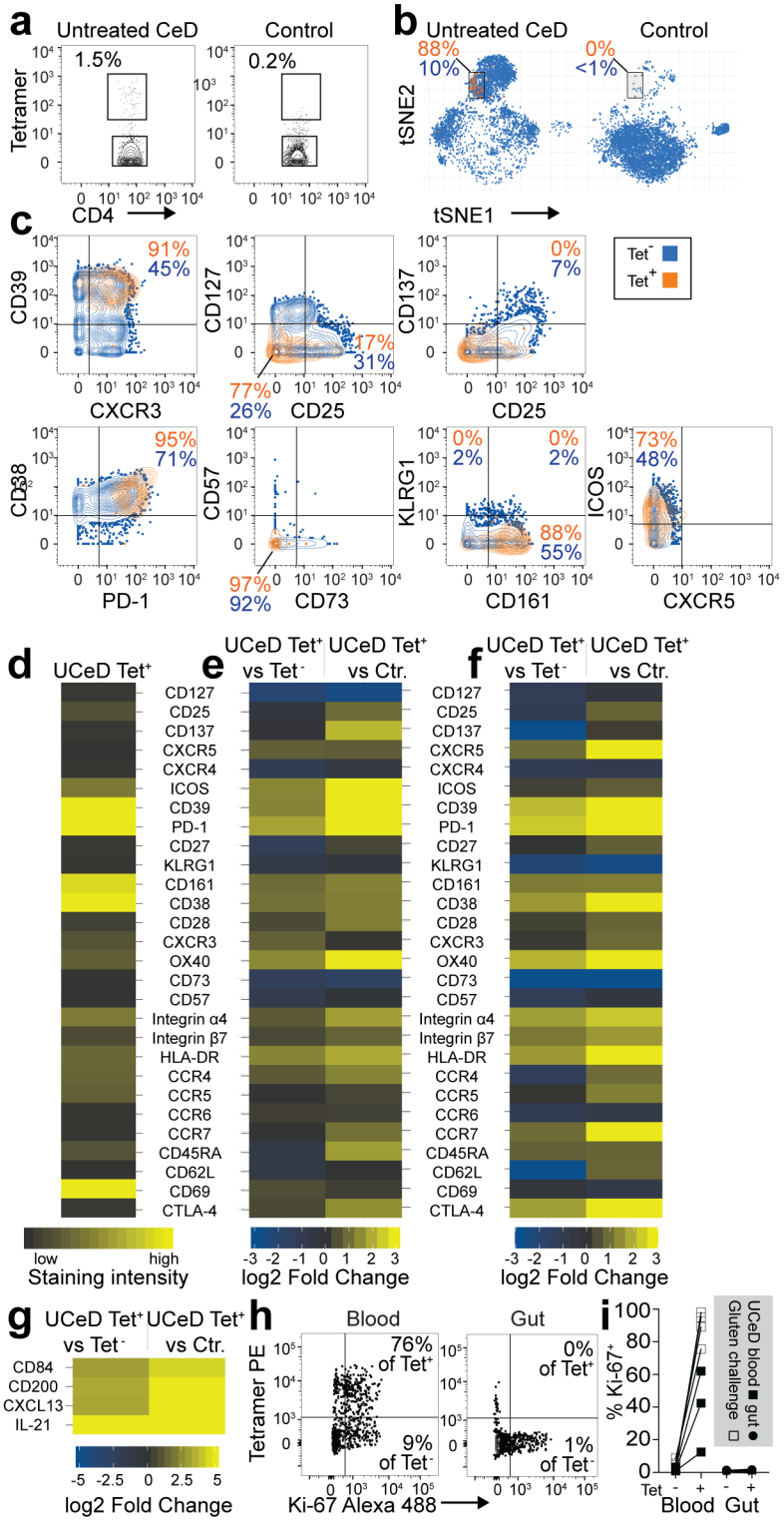
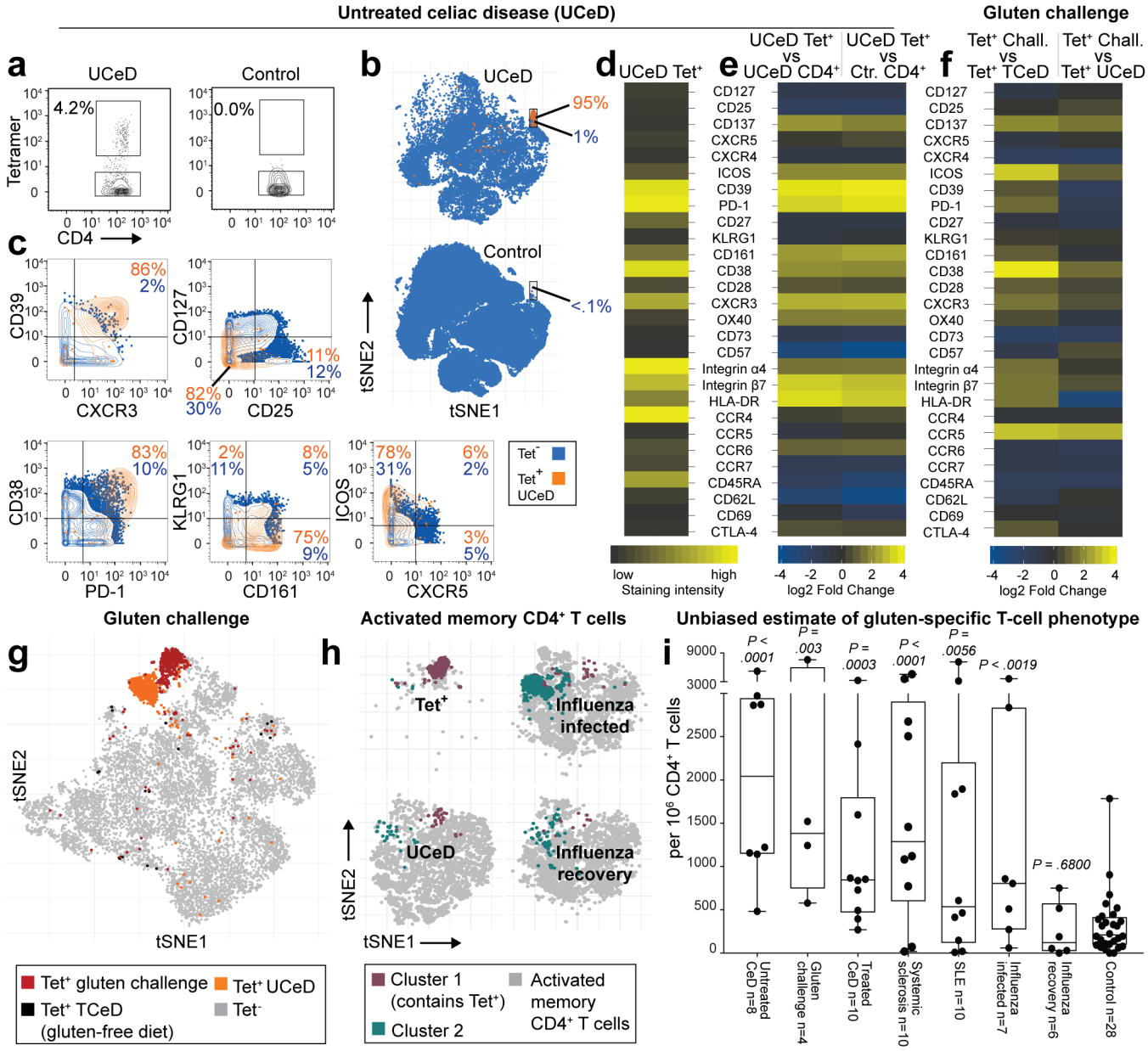
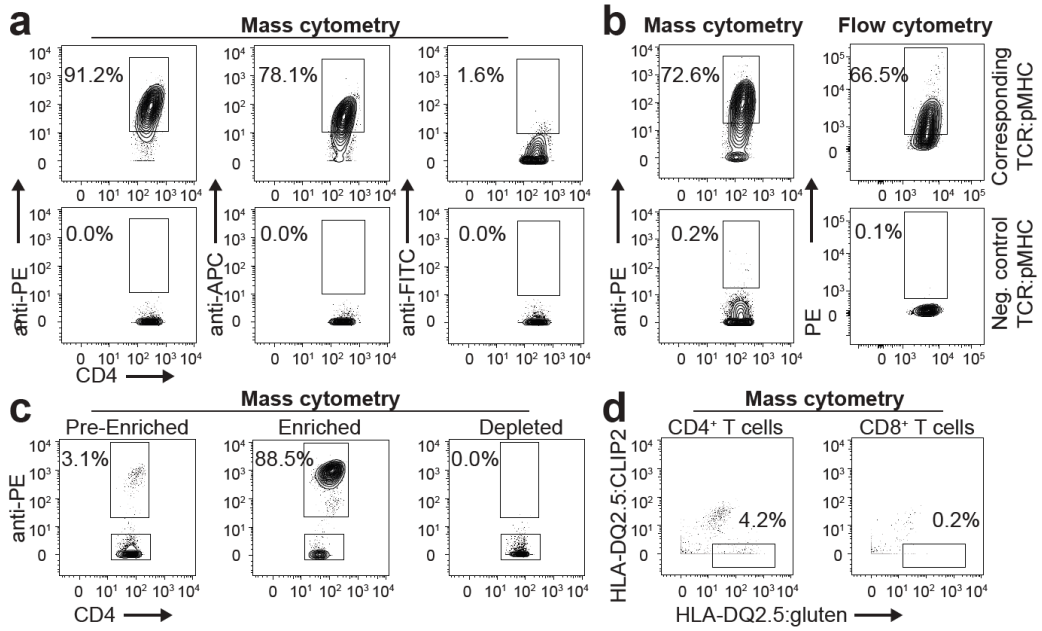

a
Untreated CeD
Control
Tetramer
1.5%
0.2%
CD4
b
Untreated CeD
Control
tSNE2
88%
10%
0%
<1%
tSNE1
c
CD39
CXCR3
91%
45%
CD127
CD25
17%
31%
77%
26%
CD137
CD25
0%
7%
Tet-
Tet+
CD38
PD-1
95%
71%
CD57
CD73
97%
92%
KLRG1
CD161
0%
2%
0%
2%
88%
55%
ICOS
CXCR5
73%
48%
d
UCeD Tet+
low
high
Staining intensity
e
UCeD Tet+ vs Tet-
UCeD Tet+ vs Ctr.
-3 -2 -1 0 1 2 3
log2 Fold Change
f
UCeD Tet+ vs Tet-
UCeD Tet+ vs Ctr.
-3 -2 -1 0 1 2 3
log2 Fold Change
CD127
CD25
CD137
CXCR5
CXCR4
ICOS
CD39
PD-1
CD27
KLRG1
CD161
CD38
CD28
CXCR3
OX40
CD73
CD57
Integrin α4
Integrin β7
HLA-DR
CCR4
CCR5
CCR6
CCR7
CD45RA
CD62L
CD69
CTLA-4
g
UCeD Tet+ vs Tet-
UCeD Tet+ vs Ctr.
CD84
CD200
CXCL13
IL-21
-5 -2.5 0 2.5 5
log2 Fold Change
h
Blood
Gut
Tetramer PE
76%
of Tet+
9%
of Tet-
0%
of Tet+
1%
of Tet-
Ki-67 Alexa 488
i
% Ki-67+
UCeD blood ■ gut ●
Gluten challenge □
Tet - + - +
Blood Gut

Figure 2


Untreated celiac disease (UCeD)
Gluten challenge
a
UCeD
Control
Tetramer
4.2%
0.0%
CD4
b
UCeD
95%
1%
Control
<.1%
tSNE2
tSNE1
c
CD39
CXCR3
86%
2%
CD127
CD25
11%
12%
82%
30%
CD38
PD-1
83%
10%
KLRG1
CD161
2%
11%
8%
5%
75%
9%
ICOS
CXCR5
78%
31%
6%
2%
3%
5%
Tet-
Tet+ UCeD
d
UCeD Tet+
e
UCeD Tet+ vs UCeD CD4+
UCeD Tet+ vs Ctr. CD4+
f
Tet+ Chall. vs Tet+ TCeD
Tet+ Chall. vs Tet+ UCeD
CD127
CD25
CD137
CXCR5
CXCR4
ICOS
CD39
PD-1
CD27
KLRG1
CD161
CD38
CD28
CXCR3
OX40
CD73
CD57
Integrin α4
Integrin β7
HLA-DR
CCR4
CCR5
CCR6
CCR7
CD45RA
CD62L
CD69
CTLA-4
low
high
Staining intensity
-4 -2 0 2 4
log2 Fold Change
Gluten challenge
Activated memory CD4+ T cells
Unbiased estimate of gluten-specific T-cell phenotype
g
tSNE2
tSNE1
Tet+ gluten challenge
Tet+ UCeD
Tet+ TCeD (gluten-free diet)
Tet-
h
Tet+
Influenza infected
UCeD
Influenza recovery
tSNE2
tSNE1
Cluster 1 (contains Tet+)
Cluster 2
Activated memory CD4+ T cells
i
per 10^6 CD4+ T cells
9000
3000
2500
2000
1500
1000
500
0
P < .0001
P = .003
P = .0003
P < .0001
P = .0056
P < .0019
P = .6800
Untreated CeD n=8
Gluten challenge n=4
Treated CeD n=10
Systemic sclerosis n=10
SLE n=10
Influenza infected n=7
Influenza recovery n=6
Control n=28

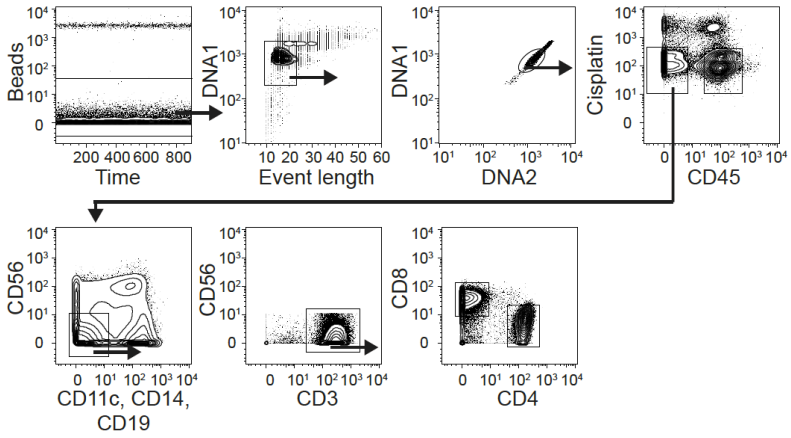
EXTENDED DATA FIGURES

To manuscript entitled "Gluten-specific CD4⁺ T cells in celiac disease have a rare phenotype shared with other autoimmune conditions"

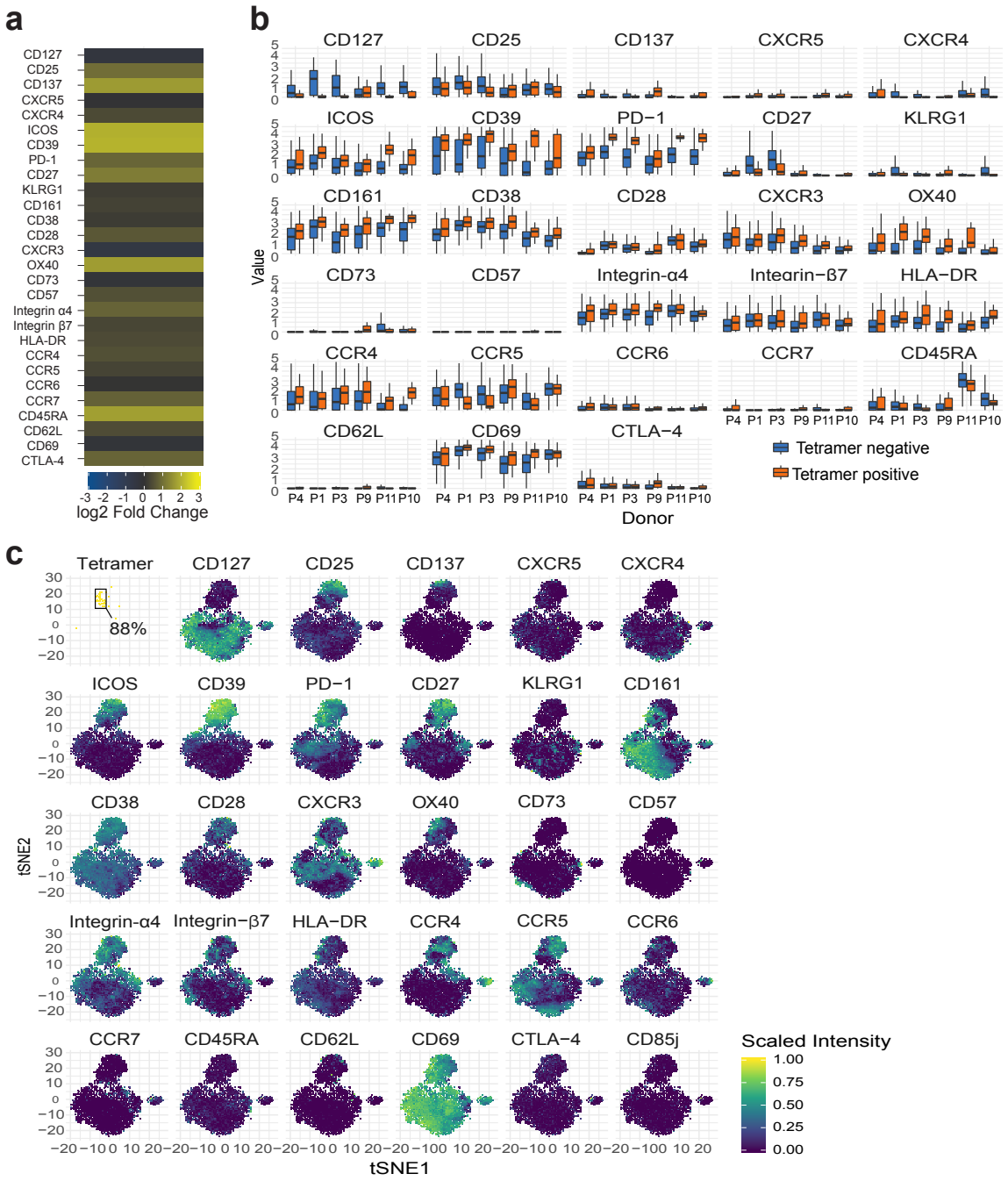


Extended Data Figure 1. Establishing HLA-class II tetramer staining with mass cytometry. (a) Gluten-specific T-cell clone binding a corresponding or negative control HLA-DQ2.5:gluten tetramer reagent metal-tagged with secondary binding to phycoerythrin (PE), allophycocyanin (APC) or fluorescein (FITC) (one T-cell clone in one experiment). **(b)** Comparison of tetramer-staining in mass cytometry and flow cytometry with a gluten-specific T-cell clone binding the corresponding or non-corresponding HLA-DQ2.5:gluten tetramer reagent (*n* = 8 T-cell clones in two mass cytometry and two flow cytometry experiments, respectively) **(c)** Tetramer-enrichment of a gluten-specific T-cell clone binding the corresponding PE-cyanine7 (PE-Cy7)-coupled HLA-DQ2.5:gluten tetramer reagent (one T-cell clone in one experiment). The T-cell

clone was spiked into PBMCs, enriched with anti-Cy7 beads and metal-tagged with anti-PE (one T-cell clone in one experiment). **(d)** Unspecific HLA-DQ2.5:gluten tetramer-binding was excluded with APC-Cy7-coupled HLA-DQ2.5:CLIP2 and metal-tagged anti-APC (*n* = 2 T-cell clones and 3 PBMC samples in two pilot experiments before established protocol).

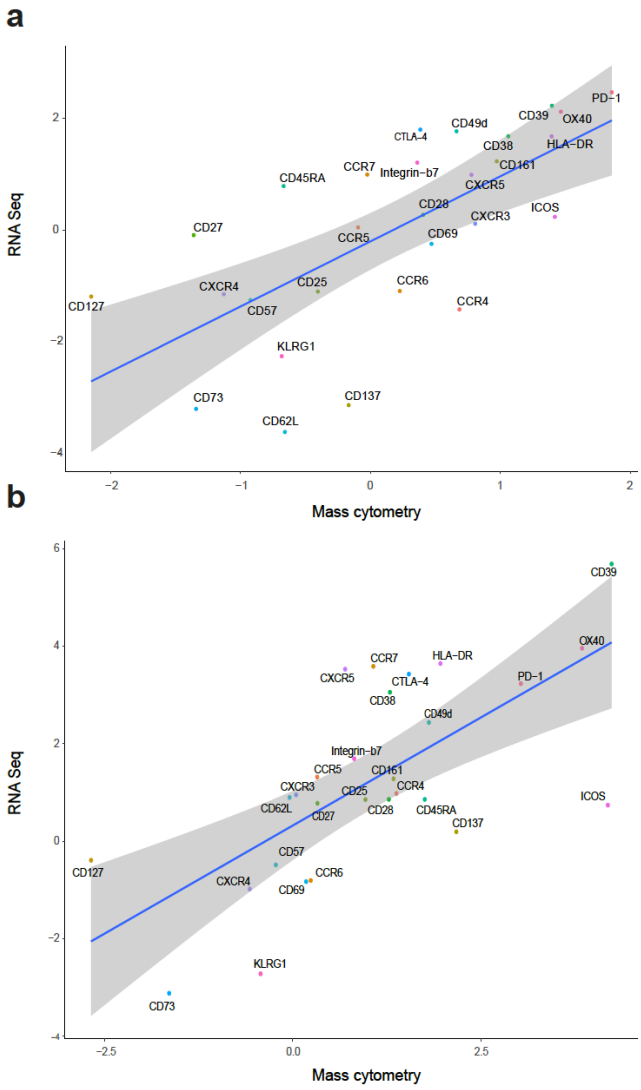


Extended Data Figure 2. Gating strategy for cells analyzed with mass cytometry. From initial plot to the plot and gate that encounters $CD4^+$ blood or gut T cells. Anti-CD45 coupled with 89Y or 108Pd was used for sample barcoding.

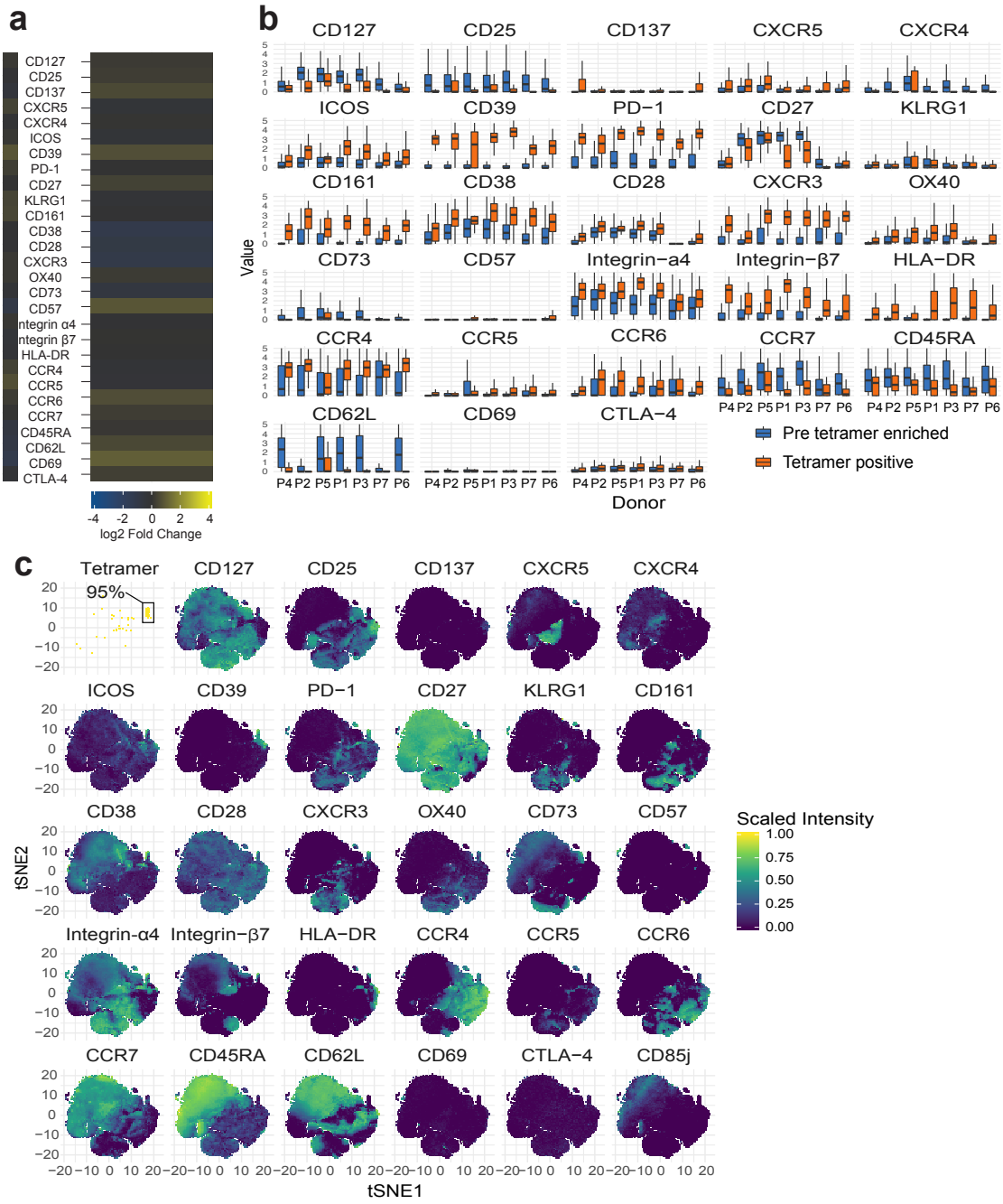


Extended Data Figure 3. On the CD4⁺ gut T cells analyzed with mass cytometry. (a) Heat map showing fold-change expression of indicated markers in CD4⁺ HLA-DQ2.5:gluten tetramer-negative gut T cells of untreated celiac disease (UCeD) patients (*n* = 6) versus CD4⁺ gut T cells of healthy controls (*n* = 7); Five experiments in total. **(b)** Expression level of mass cytometry panel

markers (Supplementary Table 1) in gluten tetramer positive and tetramer negative $CD4^+$ gut T cells in UCeD patients. Y-axis indicates arcsinh-transformed intensity values with cofactor 5. **(c)** t-SNE plots separately highlighting presence of cells expressing the markers in (a) and (b) in $CD4^+$ gut T cells merged from one UCeD patient and one healthy control. For comparison, the location of HLA-DQ2.5:gluten tetramer-binding cells of the same patient are visualized in the upper left plot.

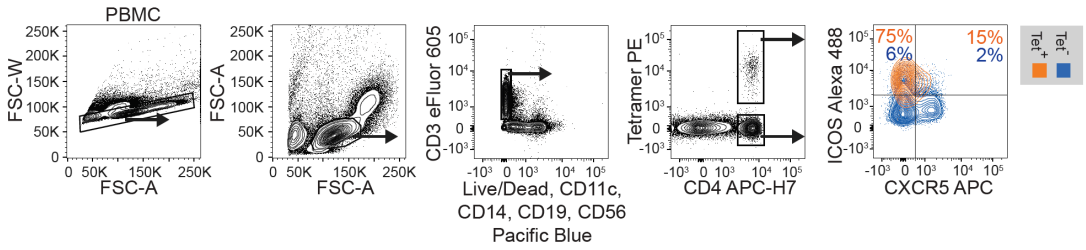


Extended Data Figure 4. Mass cytometry and RNA-seq data correlation. Correlation between RNA-seq and mass cytometry derived fold-change expression of HLA-DQ2.5:gluten tetramer positive **(a)** versus HLA-DQ2.5:gluten tetramer negative $CD4^+$ gut T cells in untreated celiac disease patients and **(b)** versus $CD4^+$ gut T cells in controls (corresponding data depicted as heat map in Figure 1e, f).

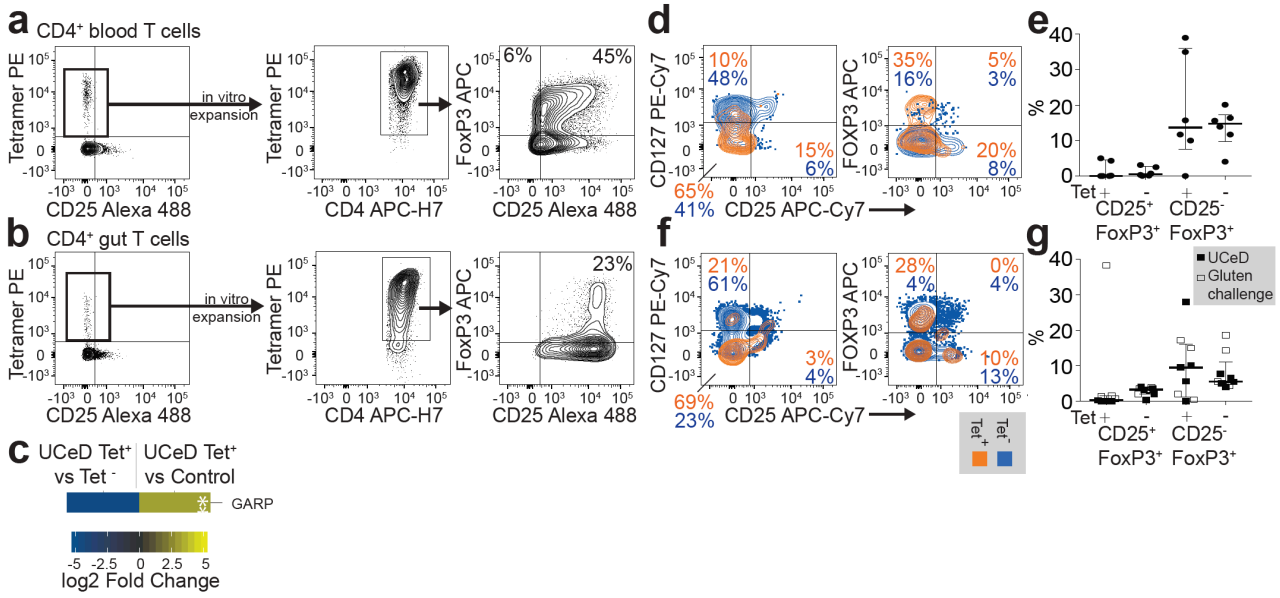


Extended Data Figure 5. On the CD4⁺ blood T cells analyzed with mass cytometry. (a) Heat map showing fold-change expression of indicated markers in CD4⁺ blood T cells of untreated celiac disease patients (pre tetramer-enriched sample, *n* = 7) versus healthy controls (*n* = 10). Untreated celiac disease (UCeD) patients and controls were analyzed with mass cytometry in

total nine experiments **(b)** Expression level of mass cytometry panel markers (Supplementary Table 1) in gluten tetramer positive and pre tetramer-enriched $CD4^+$ blood T cells in UCeD patients (*n* = 7). Y-axis indicates arcsinh-transformed intensity values with cofactor 5. **(c)** t-SNE plots separately highlighting presence of cells expressing the markers in (a) and (b) in $CD4^+$ blood T cells merged from one healthy control and one UCeD patient. For comparison, the location of HLA-DQ2.5:gluten tetramer-binding cells of an UCeD patient is visualized in the upper left plot.

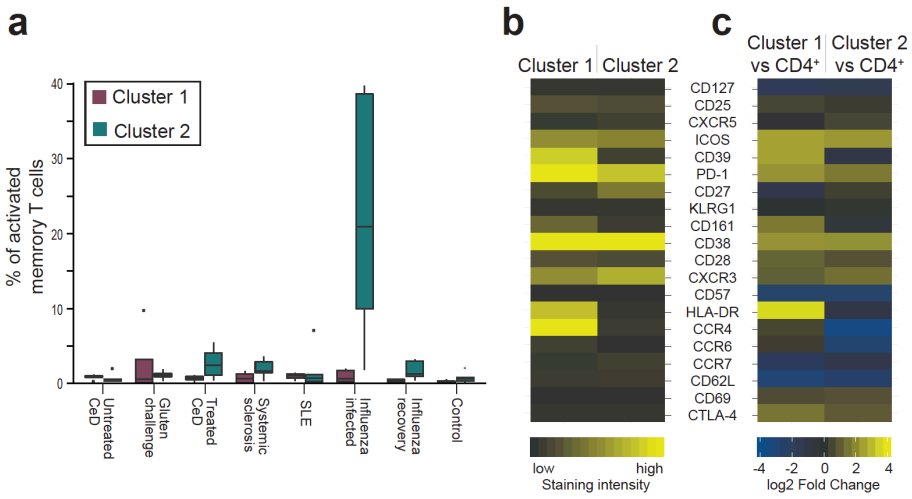


Extended Data Figure 6. Flow cytometry staining confirms CXCR5/ICOS-expression. General gating strategy for flow cytometry analysis of tetramer-binding cells including expression of CXCR5 and ICOS in tetramer (Tet) positive and negative ($^+/^-$) $CD4^+$ blood T cells in one untreated celiac disease patient (in one experiment).



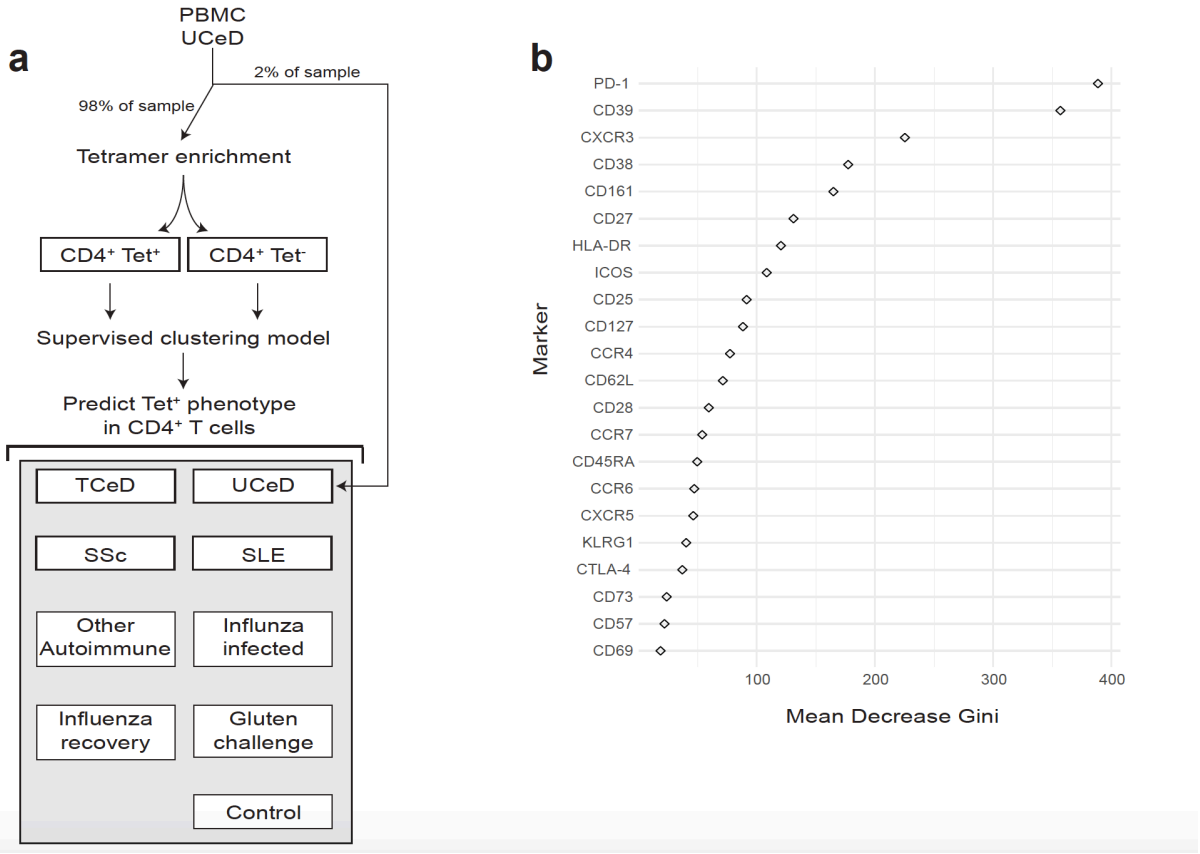
Extended Data Figure 7. Expression of regulatory T-cell-associated markers on gluten-specific CD4⁺ T cells in vitro and ex vivo. (a) CD4⁺ blood T cells of an untreated celiac disease (CeD) patient were HLA-DQ2.5:gluten tetramer (tet)-sorted ex vivo and cultured in vitro with phytohemmagglutinin and irradiated PBMCs for two weeks before re-staining with HLA-DQ2.5:gluten tetramers to analyze for expression of FoxP3 and CD25 (*n* = 2 in one experiment). **(b)** The same experiment as in (a), only with tetramer-sorted CD4⁺ gut T cells from the patient in (a) (*n* = 1 in one experiment). **(c)** RNA seq-derived fold-change expression of indicated marker in HLA-DQ2.5:gluten tet⁺ versus tet⁻ CD4⁺ gut T cells of untreated CeD patients (*n* = 5) and in tet⁺ of untreated CeD patients versus CD4⁺ gut T cells of control subjects (*n* = 4) calculated as the log2 fold change of the grand mean of donor marker intensity. GARP was differentially expressed in tet⁺ versus tet⁻ cells *but not differentially expressed when compared to CD4⁺ gut T cells in controls (complete list of differentially expressed genes in Supplementary Table 4). There were <2 GARP (Glycoprotein A repetitions predominant) transcripts per million in tet⁺ cells. **(d)** Ex vivo flow-cytometry staining of tet⁺/⁻ CD4⁺ gut T cells from an untreated CeD patient

with anti-CD127, anti-CD25 and anti-FoxP3 and **(e)** summarized CD25/FoxP3-staining in gut biopsies of five untreated HLA-DQ2.5^{+} and one HLA-DQ8^{+} CeD patients (in five experiments). **(f)** Tet $^{+}$/$^{-}$ CD4^{+} blood T cells from an untreated CeD patient with anti-CD127, anti-CD25 and anti-FoxP3 and **(g)** summarized CD25/FoxP3-staining in blood of five untreated and four gluten challenged CeD patients (in four experiments). Median frequency and interquartile range are indicated. Samples in a-b were stained with a different anti-CD25 antibody.



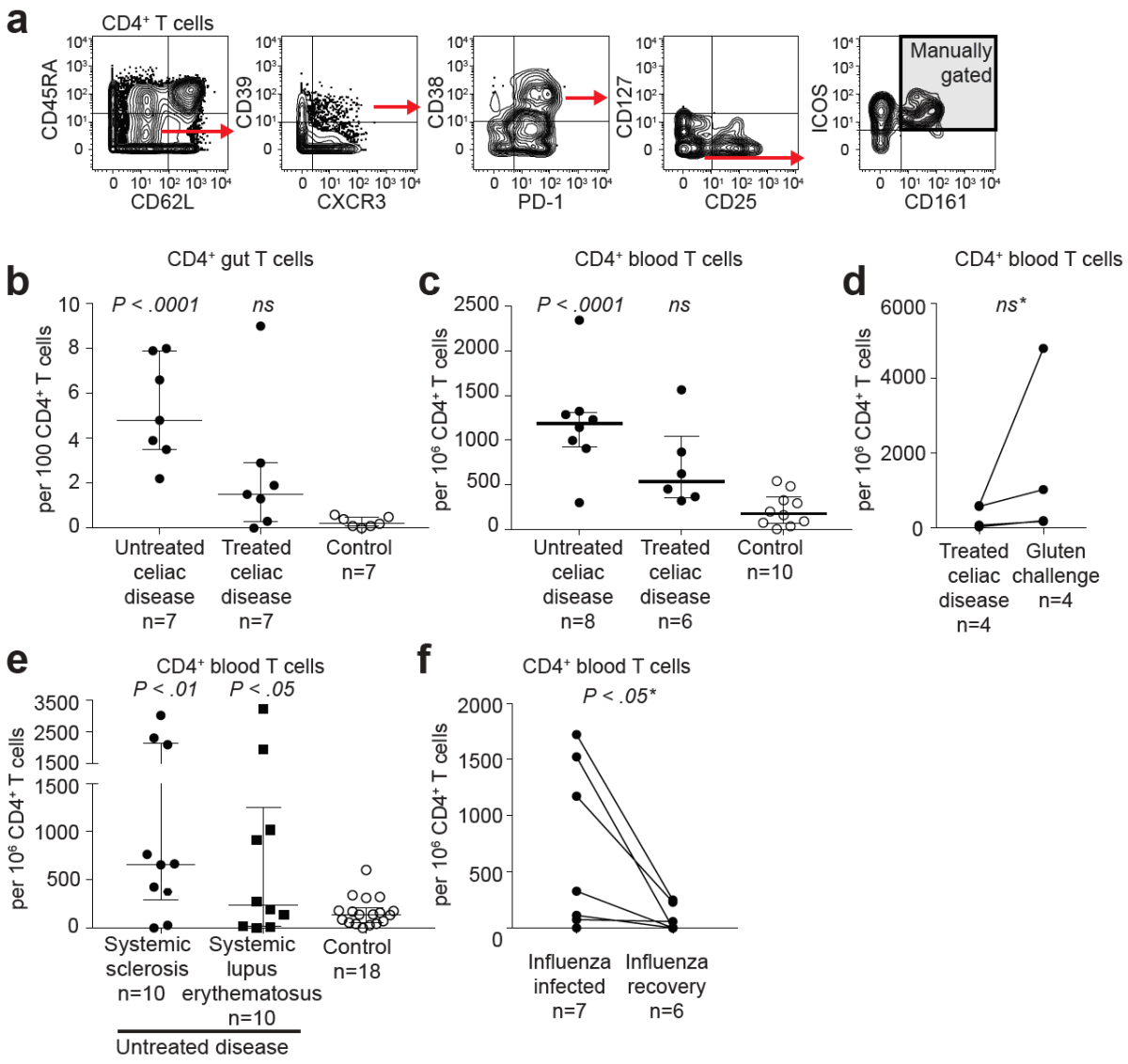
Extended Data Figure 8. Different pattern of activated CD4^{+} T cells in patients with autoimmune diseases versus influenza infection. (a) In Fig. 2h, t-SNE visualization and unsupervised clustering of activated (CD38^{+}) memory (CD45RA^{-}) CD4^{+} blood T cells in indicated participant groups and gluten tetramer positive (tet^{+}) cells of untreated celiac disease (UCeD) patients are shown. In Fig. 2h, one cluster containing 75% of tet^{+} cells (cluster 1) from seven UCeD patients and one cluster dramatically upregulated in subjects with influenza infection (cluster 2) are color-coded. **(a)** Prevalence of activated CD4^{+} memory T cells belonging to cluster 1 and cluster 2, respectively, for each indicated participant group. **(b)** Heat map of indicated

proteins in cluster 1 and cluster 2 with absolute expression (staining intensity) and **(c)** versus $CD4^+$ blood T cells depicted as the log2 fold change of the grand mean of donor marker intensity.



Extended Data Figure 9. Supervised clustering model predicting gluten-specific T-cell profile. **(a)** Diagram illustrating workflow for model training and prediction. PBMC samples from donors with untreated celiac disease are split in two parts as indicated. One part (right) is not tetramer enriched and later used for estimation of gluten-specific T-cell profile cell prevalence within the sample. The tetramer-enriched part (left) is used to train a random forest classification model using repeated K-fold cross-validation on the phenotype of the tetramer positive cells. **(b)** The

scatter plot of mean decrease in Gini score for each predictor provides information on how important the predictor variables are to the final model.



Extended Data Figure 10. Cells with profile of gluten-specific CD4⁺ T cells in celiac, autoimmune and viral disease identified with manual gating. (a) Manual gating strategy with markers giving a well-defined shift in staining intensity that define gluten-specific T cells, encompassing 41% and 48% of HLA-DQ2.5:gluten tetramer-binding CD4⁺ T cells in the gut and

$CD4^+$ effector memory T cells in blood, respectively, in untreated celiac disease (CeD) patients (while the gluten-specific cells were phenotypically similar, not all cells had a staining intensity for all ten markers above or below the manually set threshold, as visualized also in Fig. 1c, 2c). Here visualized in peripheral blood of an untreated CeD patient: $CD45RA^-$, $CD62L^-$, $CXCR3^+$, $CD39^+$, $CD38^+$, PD-1^+, $CD127^{low}$, $CD25^-$, $ICOS^+$, $CD161^+$ $CD4^+$ T cells. **(b)** Frequency of cells gated as in (a) in gut and **(c)** blood of untreated (gut *n* = 7, blood *n* = 8) and treated (gluten-free diet) CeD patients (gut *n* = 7, blood *n* = 6), healthy controls (gut *n* = 7, blood *n* = 10) **(d)** and in treated CeD patients prior to and following gluten challenge (*n = 4*) (differing from gating encountering gluten-specific cells in untreated CeD patients chiefly by lower CD39 expression as visualized also in Fig. 2f). Blood and gut samples analyzed in 12 and six experiments, respectively. Gluten challenge samples were analyzed in two experiments. **(e)** Frequency of cells gated as in (a) within patients with indicated autoimmune disorders and different set as in (b) of control subjects **(f)** and within a cohort during and after influenza infection (two experiments in total).

SUPPLEMENTARY TABLES

To manuscript entitled "Distinct phenotype of celiac disease-driving $CD4^+$ T cells identified in multiple autoimmune conditions"

Supplementary Tale 1. Mass cytometry antibody panel for celiac disease patients.

Label	Target	Clone	Supplier	Concentration
89Y	CD45	HI30	Fluidigm	2:100
108Pd	CD45	HI30	Biolegend	8 µg/ml
115In	CD57*, **	HCD57	Biolegend	1.5 µg/ml
139La	CD28*, **	CD28.2	Biolegend	4 µg/ml
141Pr	Intebrin-α4/CD49d*	9F10	Fluidigm	1:100
142Nd	KLRG1*, **	13F12F2	Thermo Fischer S.	3 µg/ml
143Nd	CD278/ICOS*, **	C398.4A	Fluidigm	0.5:100
144Nd	CD38*, **	HIT2	Fluidigm	1.5:100
145Nd	CD4	RPA-T4	Fluidigm	0.5:100
146Nd	CD8a	RPA-T8	Fluidigm	0.6:100
147Sm	CD137 (41BB)*	4-1BB	R&D systems	12 µg/ml
148Nd	CD27*, **	O323	Biolegend	1 µg/ml
149Sm	CD56 (NCAM)	NCAM16.2	Fluidigm	0.5:100
150Nd	CD127*, **	A019D5	Biolegend	1 µg/ml
151Eu	CD11c	Bu15	Biolegend	2 µg/ml
151Eu	CD19	HIB19	Biolegend	1 µg/ml
151Eu	CD14	M5E2	Fluidigm	1:100
152Sm	CD244*	2B4	R&D systems	4 µg/ml
153Eu	CD62L*, **	DREG-56	Fluidigm	0.5:100
154Sm	CD3	UCHT1	Fluidigm	0.8:100
155Gd	CD279 (PD-1)*, **	EH12.2H7	Fluidigm	1.8:100
156Gd	CD195 (CCR5)*	NP-6G4	Fluidigm	4:100
158Gd	CD194 (CCR4)*, **	L291H4	Fluidigm	0.5:100
159Tb	CD161*, **	HP-3G10	Fluidigm	0.5:100
160Gd	CD39*, **	A1	Fluidigm	1:100
161Dy	CD152 (CTLA-4)*, **	14D3	Fluidigm	5:100
162Dy	Integrin-β7*	FIB504	Fluidigm	0.5:100
163Dy	CD183 (CXCR3)*, **	G025H7	Fluidigm	0.75:100
164Dy	OX40 (CD134)*	Ber-ACT35	Biolegend	8 µg/ml
165Ho	Phycoerythrin	PE001	Fluidigm	1.25:100
166Er	CD85j/ILT2*	GHI/75	Fluidigm	1:100
167Er	CD197 (CCR7)*	G043H7	Fluidigm	1:100
168Er	CD73*, **	AD2	Fluidigm	1:100
169Tm	CD25 (IL-2R)*, **	2A3	Fluidigm	0.6:100
170Er	CD45RA*, **	HI100	Fluidigm	0.1:100
171Yb	CD185 (CXCR5)*, **	RF8B2	Fluidigm	0.75:100
172Yb	CD69*, **	FN50	Biolegend	1 µg/ml
173Yb	HLA-DR*, **	L243	Fluidigm	0.75:100
174Yb	CD196 (CCR6)*, **	G034E3	Biolegend	1 µg/ml
175Lu	CD184 (CXCR4)*	12G5	Fluidigm	1.25:100
176Yb	Allophycocyanin	APC003	Fluidigm	1.25:100
191Ir/193Ir	Nucleateed cells		Fluidigm	1:1000
195Pt	Dead cells		Fluidigm	1:1500
209Bi	CD11b*	ICRF44	Fluidigm	0.4:100

Supplementary Table 1. Mass cytometry antibody panel for celiac disease patients. Antibody panel for mass cytometry staining of HLA-DQ2.5:gluten tetramer-stained peripheral blood and single-cell suspensions of gut biopsies from celiac disease patients and controls subjects. The panel includes metal tags for sample barcoding (anti-CD45), secondary staining of phycoerythrin for identification of HLA-DQ2.5:gluten tetramer-binding cells and secondary staining of allophycocyanin for exclusion of non-HLA-DQ:gluten-specific HLA-DQ2.5:CLIP2 tetramer binding in addition to viability staining (195Pt) and nucleated cell staining (191/193Ir). One asterisk identifies markers included in the t-SNE plots in Fig. 1b, 2b, g and Extended Data Fig. 3c, 5c. Final concentrations are stated in µg/ml when using self-conjugated antibodies or per volume 100 when the concentration was not available from the manufacturer.

Suuplementary Table 2. Epitopes representad by HLA-DQ2.5 and HLA-DQ8 tetramers.

HLA-DQ2.5 epitope	Peptide sequence with underlined 9-mer core
DQ2.5-glia-α1a	QLQPFPQPELPY
DQ2.5-glia-α2	PQPELPYPQPE
DQ2.5-glia-ω1	QQPFPQPEQPFP
DQ2.5-glia-ω2	FPQPEQPFPWQP
DQ2.5-hor-3	PIPEQPQPYPQ
DQ2.5-CLIP2	MATPLLMQALPMGAL
HLA-DQ8 epitope	
DQ8-glia-α1a	SGEGSFQPSQENPQ
DQ2.5-glia-γ1b	FPEQPEQPYPEQ

Supplementary Table 2. Epitopes represented by HLA-DQ2.5 and HLA-DQ8 tetramers. We used soluble biotinylated HLA-DQ2.5 (i.e., *DQA1*05* and *DQB1*02*) or HLA-DQ8 (i.e., *DQA1*03* and *DQB1*03*) molecules covalently linked with the here listed gluten-derived $CD4^+$ T-cell epitopes (9-mer core sequence indicated in red).

Supplementary Table 3. Participant list.

Participant	Category	Sex	HLA type	Marsh Score	Anti-TG2	Anti-DGP	Matieral	Method
P1	UCeD	F	DQ2.5	3B-C	17	14	PBMC, SCS	Mass cytometry
P2	UCeD	F	DQ2.5/DQ2.2	3B-C	4.1	91	PBMC, SCS	Mass cytometry
P3	UCeD	F	DQ2.5	3B-C	>100	>100	PBMC, SCS	Mass cytometry
P4	UCeD	F	DQ2.5	3C	1	>100	PBMC, SCS	Mass cytometry
P5	UCeD	F	DQ2.5	3B	42	59	PBMC	Mass cytometry
P6	UCeD	M	DQ2.5	3A	100	37	PBMC	Mass cytometry
P7	UCeD	M	DQ2.5	3B	25	>100	PBMC	Mass cytometry
							SCS	Flow cytometry
P8	UCeD	M	DQ2.5	3C	not determined	not determined	PBMC	Mass cytometry
P9	UCeD	F	DQ2.5	3A	24.9	42	SCS	Mass cytometry
P10	UCeD	M	DQ2.5	3C	128	not determined	SCS	Mass cytometry
P11	UCeD	F	DQ2.5	3A	32	<5	SCS	Mass cytometry
P12	UCeD	M	DQ2.5	3b	>100	36	PBMC	Flow cytometry
P13	UCeD	M	DQ2.5	3C	>100	>100	PBMC, SCS	Flow cytometry
P14	UCeD	M	DQ2.5	3B-C	>100	41	PBMC, SCS	Flow cytometry
P15	UCeD	F	DQ2.5	3B	2.6	13	PBMC	Flow cytometry
P16	UCeD	F	DQ2.5	3A-B	20.8	64	PBMC	Flow cytometry
P17	UCeD	F	DQ2.5	3A	not determined	not determined	PBMC, SCS	Flow cytometry
P18	UCeD	M	DQ2.5	3B-C	>100	>100	SCS	Flow cytometry
P19	UCeD	F	DQ8	3B	77	not determined	SCS	Flow cytometry
P20	UCeD	F	DQ2.5	3B	27.3	>100	SCS	RNA Seq
P21	UCeD	F	DQ2.5	3C	>100	94	SCS	RNA Seq
P22	UCeD	F	DQ2.5	3A	4.2	18	SCS	RNA Seq
P23	UCeD	F	DQ2.5	3B	>100	>100	SCS	RNA Seq
P24	UCeD	F	DQ2.5	3C	>100	>100	SCS	RNA Seq
P25	TCeD	F	DQ2.5	not determined	<1	<5	PBMC	Mass cytometry
	Challenge	F	DQ2.5	not determined	<1	<5	PBMC	Flow & mass cytometry
P26	TCeD	F	DQ2.5	not determined	<1	<5	PBMC	Mass cytometry
	Challenge	F	DQ2.5	not determined	<1	<5	PBMC	Flow & mass cytometry
P27	Challenge	M	DQ2.5	not determined	<1	<5	PBMC	Flow cytometry
P28	Challenge	M		not determined	<1	<5	PBMC	Flow cytometry
P29	TCeD	F	DQ2.5	not determined	<1	7	PBMC	Mass cytometry
	Challenge				1.1	8		
P30	TCeD	F	DQ2.5	not determined	<1	<5	PBMC	Mass cytometry
	Challenge				<1	<5		
P31	TCeD	F	DQ2.5	not determined	not determined	not determined	PBMC	Mass cytometry
	Challenge				1.1	6		
P32	TCeD	F	DQ2.5	0	<1	<5	PBMC	Mass cytometry
P33	TCeD	F	DQ2.5	0	2.2	18	PBMC, SCS	Mass cytometry
P34	TCeD	M	DQ2.5	0	2	<5	PBMC, SCS	Mass cytometry
P35	TCeD	F	DQ2.5	0	<1	<5	PBMC	Mass cytometry
P36	TCeD	F	DQ2.5	0	1.1	13	PBMC	Mass cytometry
P37	TCeD	F	DQ2.5	0	<1	13	PBMC	Mass cytometry
P38	TCeD	F	DQ2.5	3A	<1	<5	PBMC	Mass cytometry
P39	TCeD	M	DQ2.5	1	<1	<5	SCS	Mass cytometry
P40	TCeD	F	DQ2.5	3A	2	5	SCS	Mass cytometry
P41	TCeD	F	DQ2.5	0	<1	<5	SCS	Mass cytometry
P42	TCeD	F	DQ2.5	3B	42.1	80	SCS	Mass cytometry
P43	Control	F	DQ2.5	0	<1	6	PBMC, SCS	Mass cytometry
P44	Control	F	DQ2.5	0	<1	<5	PBMC, SCS	Mass cytometry
P45	Control	F	DQ2.5	0	not determined	not determined	PBMC, SCS	Mass cytometry
P46	Control	M	DQ2.5	0	<1	<5	PBMC,SCS	Mass cytometry
P47	Control	Unknown	not determined	not determined	not determined	not determined	PBMC	Mass cytometry
P48	Control	Unknown	not determined	not determined	not determined	not determined	PBMC	Mass cytometry
P49	Control	Unknown	not determined	not determined	not determined	not determined	PBMC	Mass cytometry
P50	Control	Unknown	not determined	not determined	not determined	not determined	PBMC	Mass cytometry
P51	Control	Unknown	not determined	not determined	not determined	not determined	PBMC	Mass cytometry
P52	Control	Unknown	not determined	not determined	not determined	not determined	PBMC	Mass cytometry
P53	Control	F	DQ2.5	0	<1	<5	SCS	Mass cytometry
P54	Control	F	DQ2.5	0	<1	<5	SCS	Mass cytometry
P55	Control	M	DQ2.5	0	<1	14	SCS	Mass cytometry
P56	Control	M	DQ2.5	0	<1	<5	SCS	RNA Seq

P57	Control	M	DQ8	0	<1	<5	SCS	RNA Seq
P58	Control	F	DQ2.5	0	<1	<5	SCS	RNA Seq
P59	Control	F	DQ8	not determined	<1	<5	SCS	RNA Seq

Supplementary Table 3. Participant list. Untreated and treated celiac disease (UCeD and TCeD, respectively) patients and controls (for participants with other autoimmune diseases, influenza infection and controls, see Supplementary Table 6). The histological appearance in the duodenal mucosa was graded according to the Marsh score; Normal mucosa (Marsh score 0), increased number of intraepithelial lymphocytes (Marsh score 1), hyperplastic lesion and crypt hyperplasia (Marsh score 2) and various degree of villous atrophy (Marsh score 3A-C)[17,18]. Reference range anti-transglutaminase 2 IgA antibodies (Anti-TG2) <3U/mL, anti-deamidated gliadin peptide IgG antibodies (ant-DGP) < 20 Units/mL. Analyzed material: Peripheral blood mononuclear cells (PBMC), single-cell suspension (SCS) from duodenal biopsies.

Supplementary Table 4. Mass cytometry-derived fold change, P values with false discovery rate per marker.

UCeD tetramer pos vs neg CD4+ blood T cells			
Variable	Fold chang	P value	FDR
PD-1	3.37446489	6.2625E-06	0.00018161
CD161	2.17292987	3.0457E-05	0.00044163
CD39	3.66724386	4.7212E-05	0.00045638
CD45RA	-1.8726527	9.614E-05	0.00055962
CXCR3	2.67875437	9.6486E-05	0.00055962
CD25	-2.0599651	0.0003219	0.00084863
CD38	2.00321055	0.0002185	0.00084863
CTLA-4	0.70632218	0.00022526	0.00084863
HLA-DR	3.54694316	0.0002385	0.00084863
ICOS	1.98031909	0.00029958	0.00084863
Integrin-β7	3.35498746	0.00031919	0.00084863
CD49d	1.52339882	0.00039355	0.00095108
CXCR4	-0.8217108	0.00054343	0.00121227
CCR6	0.94168994	0.0008466	0.00165236
CCR7	-1.5620493	0.00085467	0.00165236
CD28	1.03068158	0.00126272	0.00218587
OX40	1.57652604	0.00128137	0.00218587
CD62L	-3.0189471	0.00169502	0.00273086
CD73	-1.5208278	0.00189519	0.00289265
CXCR5	0.16313598	0.00250674	0.00363478
CCR4	0.20448483	0.00312444	0.00431471
CD127	-1.4573202	0.0038376	0.00505865
CD137	2.20559632	0.06803669	0.08578539
CD85j	-0.5616357	0.07339271	0.08868286
CD27	-0.9823509	0.11666971	0.13533686
CCR5	-0.6718246	0.16377833	0.18267583
CD69	-0.4542505	0.31769572	0.34122874
KLRG1	-0.543436	0.82730049	0.85684694
CD57	-2.4448589	0.97096702	0.97096702

UCeD tetramer pos vs control CD4+ blood T cells			
Variable	Fold chang	P value	FDR
PD-1	2.15156943	1.1733E-07	3.4026E-06
CD161	1.67125133	6.7931E-06	0.00019021
CD62L	-1.538977	7.0713E-06	0.00019093
CD39	2.83912916	1.7926E-05	0.00046607
CCR6	0.72083885	2.1544E-05	0.0005386
CXCR3	1.64067935	3.9765E-05	0.00095437
CD25	-0.9838858	6.2388E-05	0.00143493
CD45RA	-1.3467149	0.00012626	0.00277762
HLA-DR	2.19361245	0.00013194	0.00277762
CCR4	0.29149137	0.00017006	0.00340113
CD73	-0.7563181	0.00032984	0.00626701
CXCR4	-0.2944311	0.00070688	0.01272393
CD38	0.99668453	0.00096018	0.01626858
Integrin-β7	1.83268317	0.00095698	0.01626858
CD49d	0.81238742	0.00153617	0.02304254
ICOS	1.16189932	0.0016852	0.02359284
CD127	-0.2952447	0.00483982	0.0629177
CD69	-1.2463682	0.0063521	0.07622523
CCR7	-0.4932464	0.00819167	0.08676084
CXCR5	0.50363021	0.00788735	0.08676084
OX40	1.1529639	0.01037375	0.09336372
CTLA-4	0.23881263	0.05660291	0.4528233
CD85j	-0.7565234	0.0690963	0.4836741
CD137	0.82384276	0.08622295	0.51733769
CCR5	0.18679908	0.27049022	1
CD27	-0.4337247	0.21267916	1
CD28	0.09696722	0.46916725	1
CD57	-1.4186848	0.33503479	1
KLRG1	0.23814081	0.82236883	1

UCeD tetramer pos vs neg CD4+ gut T cells			
Variable	Fold chang	P value	FDR
CD161	0.97220416	0.00032835	0.00952204
CXCR3	0.80636811	0.00108693	0.01208599
HLA-DR	1.39450801	0.00125028	0.01208599
CD39	1.39717471	0.00233828	0.01520002
OX40	1.46577946	0.00268195	0.01520002
PD-1	1.85744183	0.00314483	0.01520002
CD38	1.06108479	0.00368047	0.01524766
CCR4	0.68483148	0.0069221	0.02007409
CXCR5	0.77724598	0.00662812	0.02007409
ICOS	1.41940434	0.00618273	0.02007409
CD49d	0.66196686	0.01584541	0.04177427
CD127	-2.1500238	0.01862315	0.04500594
CCR6	0.2253228	0.02394554	0.05239352
Integrin-β7	0.35987293	0.02529342	0.05239352
CD69	0.46978026	0.02812773	0.05438027
CD85j	1.19527263	0.03840194	0.06960352
CD28	0.40522775	0.07587382	0.1294318
CTLA-4	0.38342996	0.13391463	0.21575135
CD57	-0.9246062	0.14609522	0.21576885

UCeD tetramer pos vs control CD4+ gut T cells			
Variable	Fold chang	P value	FDR
CD39	2.76037179	3.1303E-06	9.078E-05
CD49d	0.04227848	1.1633E-05	0.00032572
CCR4	1.17800479	0.00010815	0.00292013
CD127	-1.1591164	0.0001516	0.0039416
ICOS	3.41784645	0.00031783	0.00794571
OX40	2.03091136	0.00150522	0.0361252
PD-1	2.07209776	0.00168659	0.03879168
Integrin-β7	0.34402756	0.00268442	0.05905725
HLA-DR	0.38887431	0.00339612	0.0713185
CD161	0.85093757	0.00784807	0.15696138
CD85j	-0.3621977	0.01661189	0.315626
CD38	-0.8229729	0.03193237	0.57478262
CCR7	-0.1859834	0.036309	0.61725301
CTLA-4	0.26910706	0.05595624	0.89529979
CD137	-0.2766683	0.06308645	0.93138189
CD28	0.6410895	0.06209213	0.93138189
CD62L	-1.672258	0.07571616	0.98431007
CCR5	0.02622354	0.75333356	1
CCR6	-0.1799132	0.28255317	1

CXCR4	-1.1292641	0.1488061	0.21576885
CD25	-0.4054953	0.18510135	0.25561614
CD45RA	-0.6689098	0.23821767	0.3140142
CD27	-1.3612242	0.3278513	0.41337773
CCR5	-0.0952532	0.34324992	0.41476032
CD73	-1.3430901	0.50025216	0.5802925
KLRG1	-0.6835401	0.53751303	0.59953377
CCR7	-0.024925	0.67654213	0.70070434
CD137	-0.1678129	0.65930509	0.70070434
CD62L	-0.6591668	0.9103162	0.9103162

CD25	1.17686872	0.9910158	1
CD27	0.71823074	0.95712371	1
CD45RA	0.64616154	0.24243876	1
CD57	-1.6656093	0.13109569	1
CD69	-0.1206862	0.54090404	1
CD73	-0.7077205	0.09137822	1
CXCR3	-0.62058	0.73659431	1
CXCR4	-0.5520921	0.21856755	1
CXCR5	0.33254566	0.10640317	1
KLRG1	-0.9329318	0.46390184	1

Supplementary Table 4. Fold change, P values with false discovery rate per marker. Mass cytometry-derived fold change (> 1.5 highlighted) of indicated markers (visualized as heat map in Figure 1e (gut) and 2e (blood). P values (< 0.05 highlighted) and false discovery rate (FDR) (<0.05 highlighted) are also shown. The fold change is calculated as the log2 fold change of the grand mean of donor marker intensity for tetramer positive versus tetramer pre-enriched (blood, n = 7, upper left) or tetramer negative (gut, n = 6, lower left) $CD4^+$ T cells in untreated celiac disease (UCeD) patients. Fold change, P values and FDR are also shown for tetramer positive blood T cells in UCD patients versus $CD4^+$ blood T cells in controls (n = 10 controls, upper right) and for tetramer positive gut T cells of UCeD patients versus $CD4^+$ gut T cells of controls (n = 7 controls, lower right). P values and FDRs were calculated using an unpaired, two-tailed t-test and the Benjamini-Hochberg procedure, respectively.

Supplementary Table 5. List of differentially expressed genes. See separate excel-file.

Supplementary Table 6. Participants (P) with autoimmune disorders, influenza and controls in Fig. 2h-i, Extended Data Fig. 8 and 10e-f

Participant	Category	Sex	Disease	Organ involvement and other information	Elevated autantibodies or relevant test
P60	Untreated	F	Systemic sclerosis	Pulmonary arterial hypertension, digital ulcers, sclerodactily, oesophagal dysmotility	ANA, anti-centromere, anti-Ro/SSA
P61	Untreated	F	Systemic sclerosis	Worsening skin thickening	ANA, anti-Scl-70
P62	Untreated	F	Systemic sclerosis	Pulmonary fibrosis, Stable skin and lung disease	ANA, anti-RNA polymerase III
P63	Untreated	F	Systemic sclerosis	Renaud syndrom, active digital ulcers, osteomyelitis	ANA, anti-centromere, anti-Ro/SSA
P64	Untreated	F	Systemic sclerosis	Raynaud, interstitial lung disease, oesophagal dysmotility	ANA
P65	Untreated	M	Systemic sclerosis	Raynaud sondrom, sclerodactily, subcutaneous calcinosis, oesophagal dysmotility	ANA, anti-centromere (CENP-B), anti-Ro/SSA
P66	Untreated	F	Systemic sclerosis	Raynaud, digital ulcers, oesophagal dysmotility	ANA, anti-centromere
P67	Untreated	M	Systemic sclerosis	Raynaud, sclerodactily, renal crisis, Interstitial lung disease	ANA, anti-RNA polymerase III
P68	Untreated	F	Systemic sclerosis	Worsening skin thickening	ANA, anti-Scl-70
P69	Untreated	F	Systemic sclerosis	Stable disease, sclerodactyly Stable disease Sclerodactyly Stable disease, sclerodactyly	ANA, Scl-70
P70	Untreated	F	Systemic lupus erythematosus	Flare of malar rash, fatigue, arthralgia	ANA, anti-dsDNA, anti-RO, anti-U1-snRNP, anti-Sm
P71	Untreated	F	Systemic lupus erythematosus	Nephritis (LN III A/C), arthritis	ANA, anti-dsDNA
P72	Untreated	F	Systemic lupus erythematosus	Lupus nephritis, arthitis	anti-Ro/SSA, anti-RNP, anti-Ku
P73	Untreated	F	Systemic lupus erythematosus	UV-sensitive rash, arthritis	ANA, anti-dsDNA, anti-beta2-glycoprotein 1
P74	Untreated	F	Systemic lupus erythematosus	In remission	ANA
P75	Untreated	F	Systemic lupus erythematosus	Raynaud's disease, arthritis, telangiectasias Arthritis	ANA, anti-RNP

P76	Untreated	F	Systemic lupus erythematosus	Dry eyes and mouth, children with neonatal systemic lupus erythemotosus	ANA, anti-dsDNA, anti-Ro/SSA, anti-La/SSB
P77	Untreated	F	Systemic lupus erythematosus	Sicca symptoms, skin flare	ANA, anti-Ro/SSA, anti-La/SSB
P78	Untreated	F	Systemic lupus erythematosus	Stalbe disease	ANA
P79	Untreated	F	Systemic lupus erythematosus	New rash and headache	ANA, anti-Ro/SSA
P80	Control	Unknown	Unknown	Blood bank donor	Not determined
P81	Control	Unknown	Unknown	Blood bank donor	Not determined
P82	Control	Unknown	Unknown	Blood bank donor	Not determined
P83	Control	Unknown	Unknown	Blood bank donor	Not determined
P84	Control	Unknown	Unknown	Blood bank donor	Not determined
P85	Control	Unknown	Unknown	Blood bank donor	Not determined
P86	Control	Unknown	Unknown	Blood bank donor	Not determined
P87	Control	Unknown	Unknown	Blood bank donor	Not determined
P88	Control	Unknown	Unknown	Blood bank donor	Not determined
P89	Control	Unknown	Unknown	Blood bank donor	Not determined
P90	Control	Unknown	Unknown	Blood bank donor	Not determined
P91	Control	Unknown	Unknown	Blood bank donor	Not determined
P92	Control	Unknown	Unknown	Blood bank donor	Not determined
P93	Control	Unknown	Unknown	Blood bank donor	Not determined
P94	Control	Unknown	Unknown	Blood bank donor	Not determined
P95	Control	Unknown	Unknown	Blood bank donor	Not determined
P96	Control	Unknown	Unknown	Blood bank donor	Not determined
P97	Control	Unknown	Unknown	Blood bank donor	Not determined
P98	Untreated	M	Influenza	Fever, cough, soar throat, runny nose, myalgia	NP Swab: Influenza A^+
			Recovery, 41 days	No influenza-related symptoms	Not determined
P99	Untreated	F	Influenza	Fever, cough, soar throat, runny nose, myalgia	NP Swab: Influenza A^+
P100	Untreated	F	Influenza	Fever, cough, soar throat, runny nose, myalgia	NP Swab: Influenza A^+
			Recovery, 27 days	No influenza-related symptoms	Not determined
P101	Untreated	F	Influenza	Fever, cough, soar throat, runny nose, myalgia	NP Swab: Influenza A^+
			Recovery, 23 days	No influenza-related symptoms	Not determined
P102	Untreated	F	Influenza	Fever, cough, soar throat, runny nose, myalgia	NP Swab: Influenza A^+

			Recovery, 30 days	No influenza-related symptoms	Not determined
P103	Untreated	F	Influenza	Fever, cough, soar throat, runny nose, myalgia	NP Swab: Influenza A^+
			Recovery, 30 days	No influenza-related symptoms	Not determined
P104	Untreated	M	Influenza	Fever, cough, soar throat, runny nose, myalgia	NP Swab: Influenza A^+
			Recovery, 28 days	No influenza-related symptoms	Not determined

Supplementary Table 6. Participants (P) with autoimmune disorders, influenza and controls in Fig. 2h-i, Extended Data Fig. 8 and 10e-f. For participants (P) with autoimmune disorders, autoantibodies that were measured above the upper limit of normal at the time-point of blood draw for this study are listed. For participants included before and after influenza infection, positive nasopharengeal (NP) swab test results for influenza A virus is indicated. The patients are listed as untreated as none of them were treated with steroids or other immunomodulating drugs at the time point of blood draw. However, P99, P100, P101, P103 and P104 were treated with the antiviral drug Oseltamivir between the first and second consultation.

Supplementary Table 7. Mass cytometry antibody panel for autoimmune disorders.

Label	Target	Clone	Supplier	Concentration
89Y	CD45	HI30	Fluidigm	2:100
108Pd	CD45	HI30	Biolegend	8 µg/ml
115In	CD57**	HCD57	Biolegend	1.5 µg/ml
139La	CD28**	CD28.2	Biolegend	4 µg/ml
141Pr	Intebrin-α4/CD49d	9F10	Fluidigm	1:100
142Nd	KLRG1**	13F12F2	Thermo Fischer S.	3 µg/ml
143Nd	CD278/ICOS**	C398.4A	Fluidigm	0.5:100
144Nd	CD38**	HIT2	Fluidigm	1.5:100
145Nd	CD4	RPA-T4	Fluidigm	0.5:100
146Nd	CD8a	RPA-T8	Fluidigm	0.6:100
147Sm	TIGIT	372702	Biolegend	8 µg/ml
148Nd	CD27**	O323	Biolegend	1 µg/ml
149Sm	CD56 (NCAM)	NCAM16.2	Fluidigm	0.5:100
150Nd	CD127**	A019D5	Biolegend	1 µg/ml
151Eu	CD11c	Bu15	Biolegend	2 µg/ml
151Eu	CD19	HIB19	Biolegend	1 µg/ml
151Eu	CD14	M5E2	Fluidigm	1:100
152Sm	TCRg/d	11F2	Fluidigm	1:100
153Eu	CD62L**	DREG-56	Fluidigm	0.5:100
154Sm	CD3	UCHT1	Fluidigm	0.8:100
155Gd	CD279 (PD-1)**	EH12.2H7	Fluidigm	1.8:100
156Gd	CD29	TS2/16	Biolegend	1.6 µg/ml
158Gd	CD194 (CCR4)**	L291H4	Fluidigm	0.5:100
159Tb	CD161**	HP-3G10	Fluidigm	0.5:100
160Gd	CD39**	A1	Fluidigm	1:100
161Dy	CD152 (CTLA-4)**	14D3	Fluidigm	5:100
162Dy	Integrin-β7	FIB504	Fluidigm	0.5:100
163Dy	CD183 (CXCR3)**	G025H7	Fluidigm	0.75:100
164Dy	CD200	OX-104	Biolegend	8 µg/ml
165Ho	CD103	B-Ly7	Thermo Fischer S.	0.5:100
166Er	CCR2	K036C2	Biolegend	2 µg/ml
167Er	CD197 (CCR7)**	G043H7	Fluidigm	1:100
168Er	CD73**	AD2	Fluidigm	1:100
169Tm	CD25 (IL-2R)**	2A3	Fluidigm	0.6:100
170Er	CD45RA**	HI100	Fluidigm	0.1:100
171Yb	CD185 (CXCR5)**	RF8B2	Fluidigm	0.75:100
172Yb	CD69**	FN50	Biolegend	1 µg/ml
173Yb	HLA-DR**	L243	Fluidigm	0.75:100
174Yb	CD196 (CCR6)**	G034E3	Biolegend	1 µg/ml
175Lu	CX3CR1	2A9-1	Biolegend	6 µg/ml
176Yb	TCRα/β	IP26	Fluidigm	1.5:100
191Ir/193Ir	Nucleateed cells		Fluidigm	1:1000
195Pt	Dead cells		Fluidigm	1:1500
209Bi	CD11b	ICRF44	Fluidigm	0.4:100

Supplementary Table 7. Mass cytometry antibody panel for autoimmune disorders. Antibody panel for mass cytometry staining of peripheral blood from participants with autoimmune disease, participants during and after influenza infection and control subjects (Figure 2h-i and Extended Data Fig. 8). Two asterisks identify the markers used to generate Fig. 2i (22 $CD4^+$ T-cell markers common to mass cytometry staining panel in Supplementary table 1). Final concentrations are stated in µg/ml when using self-conjugated antibodies or per volume 100 when the concentration was not available from the manufacturer.

Supplementary Table 8. Antibodies used in flow cytometry experiments.

Label	Target	Clone	Supplier	Concentration
APC	FoxP3	PCH101	Thermo Fischer S.	5 µg/ml
Alexa 700	CD4	A161A1	Biolegend	3:100
APC-Cy7	CD25	BC96	Biolegend	4:100
PE-Cy5	CD45RA	HI100	Biolegend	1:100
PE-Cy7	CD127	A019D5	Biolegend	5:100
Alexa 488	Ki-67	Ki-67	Biolegend	5 µg/ml
Pacific Blue	CD11c	3.9	Biolegend	1.5:100
Pacific Blue	CD56	5.1H11	Biolegend	1.5:101
Pacific Blue	CD14	HCD14	Biolegend	1.5:102
Pacific Blue	CD19	6D5	Biolegend	1.5:103
Aqua/510	Dead cells		Biolegend	1:100
BV605	CD3	UCHT1	Biolegend	4:100
BV650	Integrin β7	FIB504	BD Bisciences	5:100
Alexa 488	CD278/ICOS	C398.4A	Biolegend	1 µg/ml
PerCP	CD62L	DREG-56	Biolegend	3:100
Pe-Cy7	CD45RA	HI100	Thermo Fischer S.	2.5:100
APC	CXCR4	12G5	Biolegend	1:100
APC	CXCR5	J252D4	Biolegend	1:100
APC-H7	CD4	SK3	BD Bisciences	0.36 µg/ml
Fixable Violet/405	Dead cells		Thermo Fischer S.	1:100
BV605	CD3	OKT3	Biolegend	3:100
Alexa 488	CD25	14101	Thermo Fischer S.	4:100

Supplementary Table 8. Antibodies used in flow cytometry experiments. Antibodies used for flow cytometry staining. Final concentrations are stated in µg/ml when information available or per volume 100 when the concentration was not available from the manufacturer.