BRIEF COMMUNICATION
https://doi.org/10.1038/s41591-019-0403-9
nature medicine

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 E. A. Lundin[1,2,10], Ludvig M. Sollid[1,2,3,11,15]* and Mark M. Davis[4,12,13,15]*

Combining HLA-DQ–gluten tetramers with mass cytometry and RNA sequencing analysis, we find that gluten-specific CD4+ T cells in the blood and intestines of patients with celiac disease 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.

Celiac disease (CeD) is a human leukocyte antigen (HLA)-DQ2/HLA-DQ8-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 Figs. 1 and 2 and Supplementary Table 1). We found that cells binding to 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+ patients with untreated CeD and comprise 0.3–1.5% of the total (Fig. 1a,b and participants: Supplementary Table 3). These gut T cells expressed multiple activation markers (C-X-C chemokine receptor type 3 (CXCR3), CD38, CD161, CD28, HLA-DR and OX40 (also known as TNFRSF4)) as well as CD39 and programmed cell death protein 1 (PD-1), suggestive of chronic activation, whereas are negative for the exhaustion marker killer cell lectin-like receptor subfamily G member 1 (KLRG1) (Fig. 1c–e, Supplementary Table 4 and 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 and Extended Data Fig. 4).

In addition, RNA sequencing (RNA-seq) analysis demonstrated that CD200, CD84, C-X-C motif chemokine 13 (CXCL13) and interleukin-21 (IL-21) are transcribed as well (Fig. 1g and the complete list in Supplementary Table 5). These markers are characteristic of follicular B helper T 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-1+/CXCR5− cells, of unknown antigen specificity, are expanded in the synovium of seropositive patients with rheumatoid arthritis and express a similar phenotype to what we report here, including expression of CD200, CXCL13, IL-21, PD-1, inducible T-cell costimulator (ICOS), OX40 and CD28 (ref. [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 the blood (13–98%) but not in the gut (Fig. 1h,i). Conceivably, gluten-specific T cells in CeD can promote the 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 (refs. [4,5]), indicate that the disease-relevant T cells 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].

Although 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 the blood of patients with untreated CeD largely expresses the same pattern of markers as in the gut (CXCR3+, CD38+, CD39+, PD-1+, HLA-DR+, CD161+, KLRG1−, CD28+ and OX40+; Fig. 2a–e and per donor in Extended Data Fig. 5), except for being CD69−. Furthermore, despite Ki-67 expression (Fig. 1h,i), only a small fraction of the tetramer-positive cells in the blood expressed CXCR5 (confirmed by FACS in Extended Data Fig. 6) and thus do not express a classical follicular T helper cell phenotype. As previously observed[9–11], the tetramer-binding cells were almost exclusively effector memory cells (CD45RA− and CD62L−), integrin β_7+ and CD38+. It was recently reported that most gluten-specific cells express

[1]KG Jebsen Coeliac Disease Research Centre, University of Oslo, Oslo, Norway. [2]Institute of Clinical Medicine, University of Oslo, 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: Eivind G. Lund, Omri Snir. [15]These authors jointly directed this work: Ludvig M. Sollid, Mark M. Davis. *e-mail: l.m.sollid@medisin.uio.no; mmdavis@stanford.edu

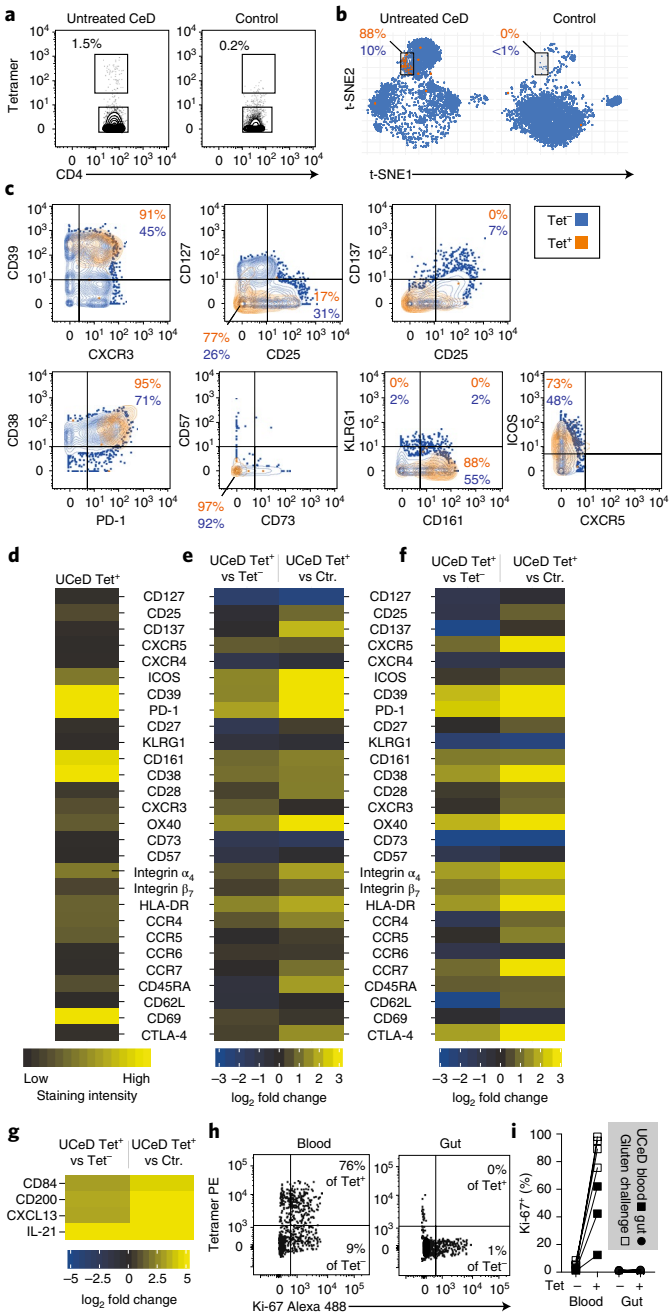


Fig. 1 | Distinct non-proliferative phenotype of gluten-specific CD4⁺ gut T cells. a, HLA-DQ2.5–gluten tetramer (tet) staining in a patient with untreated CeD (UCeD) and a control subject with mass cytometry . **b**, t-SNE plots of total CD4⁺ gut T cells in an UCeD patient and a control subject. **c**,**d**, Expression of proteins on tet⁺/tet⁻ CD4⁺ gut T cells in an UCeD patient (**c**) and summarized for tet⁺ cells in the six merged patients with UCeD (**d**). CTLA-4, cytotoxic T lymphocyte protein 4. **e**,**f**, Mass cytometry-derived (**e**) and RNA-seq-derived (**f**) $\log_2$ fold-change expression of the indicated markers in tet⁺ versus tet⁻ CD4⁺ gut T cells of patients with UCeD and compared to CD4⁺ gut T cells of control subjects (Ctr.). $n = 6$ patients with UCeD, 7 controls, 5 experiments (**a**–**e**). **g**, RNA-seq-derived $\log_2$ fold-change expression of the indicated markers in CD4⁺ gut T cells, differentially expressed in tet⁺ versus tet⁻ and versus CD4⁺ gut T cells in control subjects. $n = 5$ patients with UCeD, 4 control subjects, 2 experiments (**f**,**g**). **h**,**i**, Flow cytometry-derived Ki-67 expression in tet⁺/tet⁻ CD4⁺ blood and gut T cells of a gluten-challenged and UCeD patient, respectively (**h**), summarized for five gut samples, seven blood samples (four experiments) (**i**).

a regulatory T 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⁻ (Figs. 1 and 2), and negative for the regulatory T cell marker GARP (Extended Data Fig. 7c). In addition, although 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 regulatory T cell phenotype.

We next asked whether antigenic stimulation drives these CD4⁺ T cells. This involved a 3-day oral gluten challenge in five patients with CeD (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 6 (refs. [10,13]). Upon challenge, these cells upregulated markers expressed by gluten-specific cells in the patients with untreated celiac, including CD38, CD39, CXCR3, PD-1, ICOS, CD161, C-C chemokine receptor type 5 (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 peripheral blood mononuclear cells (PBMCs) of patients with systemic sclerosis and systemic lupus erythematosus, together with CeD subjects and presumably healthy blood bank donors (participants: Supplementary Table 6 and 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 seven out of eight patients with untreated CeD, eight out of ten patients with systemic sclerosis and four out of ten patients with systemic lupus erythematosus 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. Although four out of seven 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 and 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 humans[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 rheumatoid arthritis[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. As we know that most or all of the gluten-specific T cells are in this subset in patients with CeD, it is reasonable to imagine that these cells might also be the key disease-driving T cells in other autoimmune diseases.

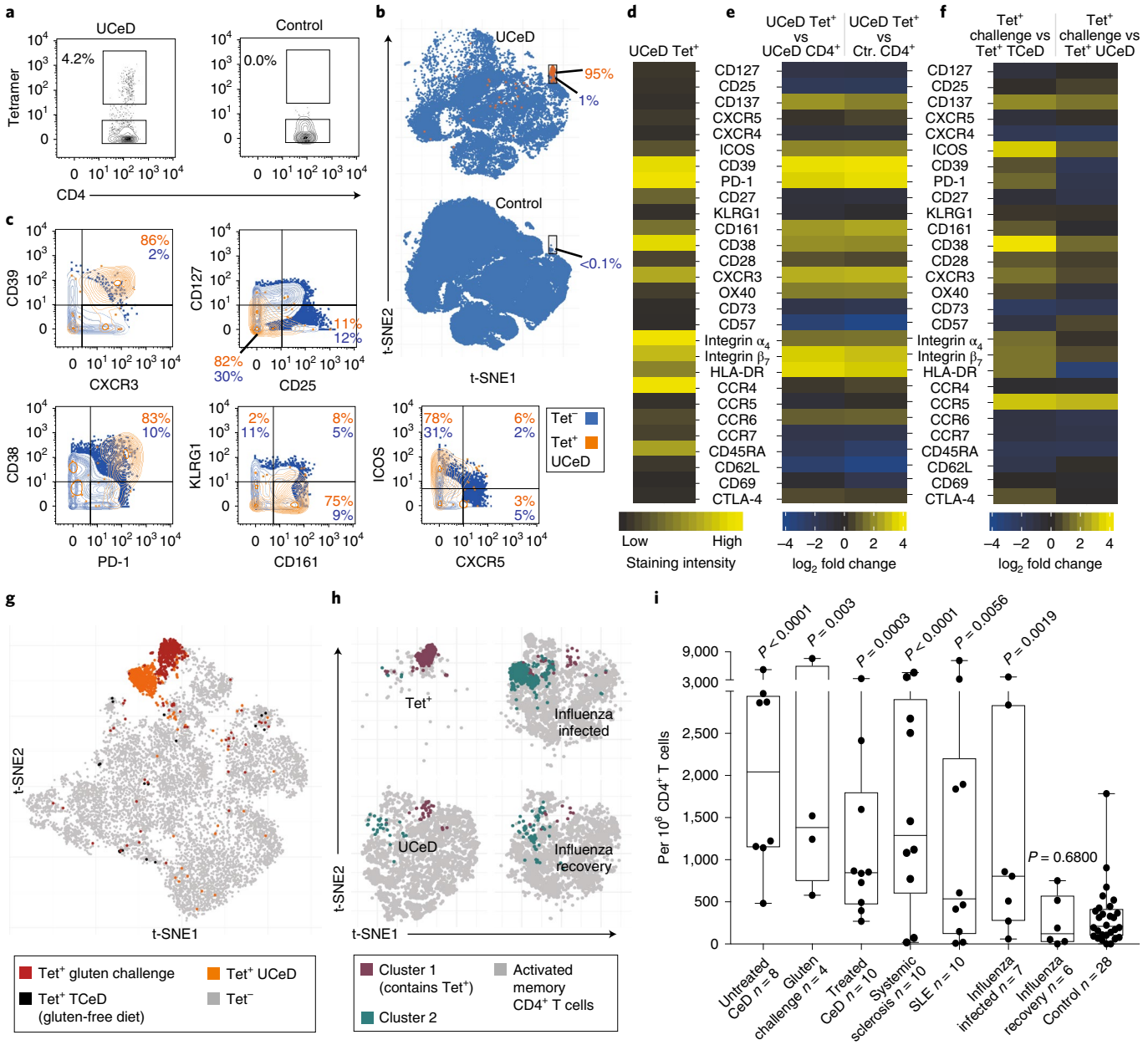


Fig. 2 | Distinct, antigen-induced phenotype of gluten-specific CD4⁺ blood T cells and occurrence of a similar subset in other immune conditions. **a**, HLA-DQ2.5:gluten tet staining with mass cytometry for a patient with UCeD and control subject. **b**, t-SNE plots with CD4⁺ blood T cells of a patient with UCeD and control subject. **c**, Expression of proteins on tet⁺/tet⁻ CD4⁺ blood T cells of a patient with UCeD. **d**, Heat map with absolute expression (staining intensity) of tet⁺ cells. **e**, $\log_2$ fold change for tet⁺ versus pre-tet-enriched CD4⁺ T cells. $n = 7$ patients with UCeD, 10 control subjects, 9 experiments (**a**–**e**). **f**, $\log_2$ fold-change expression of indicated markers for tet⁺ CD4⁺ blood T cells before versus following gluten challenge of five patients with treated CeD (TCeD) and versus tet⁺ cells of seven patients with UCeD (same patients with UCeD in **d** and **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. Three representative experiments are shown in **f** and **g**. **h**, t-SNE plots and unsupervised clustering of activated (CD38⁺) memory (CD45RA⁻) CD4⁺ blood T cells in the indicated participant groups ($n = 5$ distinct samples in each group) and tet⁺ cells of seven patients with UCeD. Cluster 1, containing 75% of tet⁺ cells from patients with UCeD, and cluster 2, upregulated in subjects with influenza infection (Extended Data Fig. 8), are color coded. **i**, Unbiased prevalence estimate among the indicated participant groups (19 experiments) for cells with the same phenotype profile as tet⁺ cells in patients with UCeD, using a supervised classification model (Extended Data Fig. 8). *P* values were calculated with an unpaired, two-tailed *t*-test. The median frequency, interquartile range and maximum/minimum whiskers are shown for the box plots. SLE, systemic lupus erythematosus.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of data availability and associated accession codes are available at https://doi.org/10.1038/s41591-019-0403-9.

Received: 15 December 2017; Accepted: 19 February 2019;
Published online: 25 March 2019

References

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

2. Rao, D. A. et al. *Nature* **542**, 110–114 (2017).
3. Sollid, L. M. et al. *Gut* **41**, 851–852 (1997).
4. Di Niro, R. et al. *Nat. Med.* **18**, 441–445 (2012).
5. Steinsbo, O. et al. *Nat. Commun.* **5**, 4041 (2014).
6. Hoydahl, L. S., et al. *Gastroenterology* https://doi.org/10.1053/j.gastro.2018.12.013 (2018).
7. Moens, L. et al. *Front. Immunol.* **5**, 65 (2014).
8. Kooy-Winkelaar, Y. M. et al. *Proc. Natl Acad. Sci. USA* **114**, E980–E989 (2017).
9. Christophersen, A. et al. *United European Gastroenterol. J.* **2**, 268–278 (2014).
10. Raki, M. et al. *Proc. Natl Acad. Sci. USA* **104**, 2831–2836 (2007).
11. Du Pre, M. F. et al. *Am. J. Gastroenterol.* **106**, 1147–1159 (2011).
12. Cook, L. et al. *J. Allergy Clin. Immunol.* **140**, 1592–1603 (2017).
13. Risnes, L. F. et al. *J. Clin. Invest.* **128**, 2642–2650 (2018).
14. Yu, W. et al. *Immunity* **42**, 929–941 (2015).
15. Su, L. F. et al. *Immunity* **38**, 373–383 (2013).

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 patients with influenza virus infection, 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 laboratory 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 with the 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. and L.M.S.); Stiftelsen KG Jebsen (project SKGJ-MED-017) (L.M.S. and 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 the 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 the libraries for RNA-seq. RNA-seq data were analyzed by E.G.L. and C.K. The mass cytometry data were analyzed by E.G.L. and A.C. The CeD patient material was organized by K.E.A.L., S.D.-K. and S.Z. Material from patients with autoimmune disorders other than CeD was organized by Ø.M., P.J.U., M.R.-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.-K., S.Z., Ø.M., P.J.U., M.R.-P., J.F.S, C.L.D. and K.E.A.L.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41591-019-0403-9.

Supplementary information is available for this paper at https://doi.org/10.1038/s41591-019-0403-9.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to L.M.S. or M.M.D.

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 (Oslo, Norway), 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 (Stanford, CA, USA). All patients with CeD were HLA-DQ2.5$^+$ (that is, *DQA1*05* and *DQB1*02*) or HLA-DQ8$^+$ (that is, *DQA1*03* and *DQB1*03:02*) and diagnosed according to the guidelines of the British Society of Gastroenterology[16]. 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]. 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). Patients with treated CeD who were challenged with gluten received one in-house-produced cookie containing 10 g of enriched flour (Validus AS) each day for 3 days and blood samples were taken on day 6 after gluten challenge when a peak in the frequency of gluten-specific CD4$^+$ blood T cells was expected[10,19] (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 to 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[20] or systemic lupus erythematosus[21], respectively. The recruitment of these patients was covered by the 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 2 × 10 min with 2 mM EDTA + 2% FCS in PBS at 37 °C to remove the epithelial layer prior to further digestion with collagenase (1 mg ml^{-1}) 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-µm 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 a combination of HLA class-I tetramers and mass cytometry[22,23]. We thawed the frozen cell samples in 20% 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% BSA, 2 mM EDTA and 0.05% sodium azide in PBS). After 450*g* centrifugation, cells were treated with 1:10 diluted FcR block (Miltenyi Biotek), stained with anti-CD11c, anti-CD14 and 5 µg ml^{-1} purified anti-CD32 (clone FUN-2) to reduce nonspecific tetramer binding, and barcoded with anti-CD45 coupled with 89Y or 108Pd[24] in 200 µl CyFACS buffer. Only names and staining concentrations of monoclonal antibodies that are not listed in Supplementary Tables 1 and 7 are specified here. After one wash step, the samples from patients with CeD were stained for 40 min at room temperature with HLA-DQ2.5:gluten tetramers representing the five different disease-relevant and immunodominant gluten T cell epitopes[25]—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^{-1} each in 200 µl CyFACS buffer, in the case of peripheral blood mononuclear cells (PBMCs), or 100 µl CyFACS buffer, in the case of biopsy-derived single-cell suspensions. We also added tetramers representing HLA-DQ2.5:CLIP2 at a concentration of 20 µg ml^{-1} to exclude tetramer background staining (Extended Data Fig. 1d). HLA-DQ2.5:gluten and HLA-DQ2.5:CLIP2 molecules were produced as previously described[26] and, 2 h before cell staining, multimerized on phycoerythrin (PE)-cyanine7 (Cy7)-coupled streptavidin or allophycocyanin (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 min on ice in 100 µl CyFACS buffer followed by another wash step. To facilitate tetramer enrichment, the PBMCs of patients with CeD were resuspended in 50 µl anti-Cy7 metal beads with 150 µl CyFACS buffer and incubated for 20 min on ice (combined anti-Cy7 enrichment and anti-PE staining was established with the T cell clone TCC1214P.A.27, derived from the blood of patients with CeD[9], and is visualized in Fig. 1c). The cells were washed and 2% of the PBMCs of patients with CeD (pre-tetramer-enriched sample) were removed and added to 1 million CD45-barcoded carrier cells of a healthy donor to reduce cell loss and left on ice until further staining. The remaining PBMCs of patients with CeD were enriched for tetramer-binding cells on a magnetized LS column (Miltenyi Biotec). We then added 1 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), and washed the cells once before all samples were stained for 20 min 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 5 min at room temperature with cisplatin (Fluidigm) at 1/1,500 concentration and washed before overnight incubation at 4 °C with 1:1,000 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 (1× each) before they were analyzed in milli-Q water using a Helios instrument (Fluidigm). Unlike in the gut samples analyzed here and in previous studies on gluten-specific cells in the blood using flow cytometry[9,27], we have not specified the frequency of tetramer-binding cells in the 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 the blood of patients with CeD[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. The T cell clones used to establish and validate the HLA class II tetramer staining with mass cytometry can be provided on request after completion of a material transfer agreement.

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 patient with CeD 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[28]. Tetramer-sorted cells were cultured in vitro as previously described[29]. Staining for Ki-67 and Foxp3 was performed according to the manufacturer's protocol (Thermo Fisher 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 patients with CeD 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 (ref. [28]) (Supplementary Table 2), as previously described[30]. 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 Fisher Scientific). See also the Nature Research Reporting Summary for more details on the antibodies used. We added anti-CD27 and anti-IgA owing 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 a 2100 Bioanalyzer using a RNA 6000 Pico Kit (Agilent Technologies).

Approximately 90 ng RNA was used for cDNA synthesis and amplification. cDNA synthesis was performed at 42 °C for 90 min and 70 °C for 10 min. cDNA was amplified with these PCR conditions: 95 °C for 1 min; followed by 15 cycles (98 °C for 10 s; 65 °C for 30 s; and 68 °C for 3 min) and 72 °C for 10 min using the 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 the Nextera XT library preparation kit (Illumina). Amplicon libraries were sequenced on NextSeq500 (Illumina) 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 (Figs. 1a,c and 2a,c,h and Extended Data Figs. 1a–d, 2, 6, 7a,b,d,f and 10a) and for cell quantifications (Extended Data Figs. 7e,g and 10b–f).

We used the GraphPad Prism 7 software (GraphPad Software) for statistical analysis and visualization of cell frequencies (Fig. 1i and Extended Data Figs. 7e,g and 10b–f). Here, we applied an unpaired, two-tailed *t*-test (Extended Data Fig. 10b,c,e (the median frequency and interquartile range indicated)) or a paired, two-tailed *t*-test (Extended Data Fig. 10d,f) 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 the generation of t-SNE (*t*-distributed stochastic neighbor embedding) plots [31] and all other analysis presented in Figs. 1d–g and 2d–i and Extended Data Figs. 3–5, 8 and 9. The markers used to generate the t-SNE plots in Figs. 1b and 2b,g and Extended Data Figs. 3c and 5c (31 markers in both gut and blood samples, which did not include the marker for tetramer staining (165Ho anti-PE)) are identified with one asterisk in Supplementary Table 1.

Mass cytometry data (the fcs file containing only blood or gut $CD4^+$ T cells) were loaded into R using the flowCORE package. Here, the aggregate marker intensity (Figs. 1d and 2d and Extended Data Fig. 8b) was computed as the grand mean of each donors mean marker intensity. Mass cytometry fold change (Figs. 1e and 2e,f and Extended Data Figs. 3a, 5a, 7c and 8c) was computed as the $\log_2$ fold change of the aggregate marker intensity. Heat maps, to visualize the aggregate marker intensity and the $\log_2$ fold change, were generated using ggplot2, and the t-SNE plots were generated using the Rtsne package. For t-SNE plots, box plots (Extended Data Figs. 3b and 5b), supervised classification (Fig. 2i and 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.[32]. In the generated box plots in Extended Data Figs. 3b and 5b (generated with ggplot2), the *y* axis indicates ArcSinh-transformed intensity and the boxes show the median frequency and interquartile range. Whiskers show the largest and smallest values below 1.5 times the interquartile range. In Supplementary Table 4, we used a paired, two-tailed *t*-test to calculate significant differences in the mean marker intensity between tetramer-positive and tetramer-negative cells from patients with CeD in the blood and the gut. In the same table, we performed an unpaired, two-tailed *t*-test for all other comparisons in which the test conditions were from unmatched donors (for example, patients with CeD versus healthy controls). *P* values were corrected for multiple testing using the Benjamini–Hochberg procedure.

In Fig. 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 in which we only had four samples with sufficient cells. Furthermore, we sampled at most 3,707 activated cells per sample, which is the median number of cells per sample, and used these cells for clustering. For t-SNE visualization in Fig. 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 box plot (Extended Data Fig. 8a), indicating the median frequency and interquartile range. Here, the whiskers show the largest and smallest values 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 patients with untreated CeD (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[33] 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 box plot (Fig. 2i), which shows the median frequency and interquartile range, whereas the whiskers indicate the maximum and minimum 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 Tables 1 and 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[34] version 0.7.2 for mapping with parameters -l IU—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 and Extended Data Fig. 7c) were computed using DESeq2 (ref. [35]) 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 5×10^{-3} after adjusting for multiple testing. Heat maps, to visualize the $\log_2$ fold change, were generated using ggplot2, as with fold-change expression in the mass cytometry data.

Further information on methods, statistics and data analysis is provided in the Nature Research Reporting Summary.

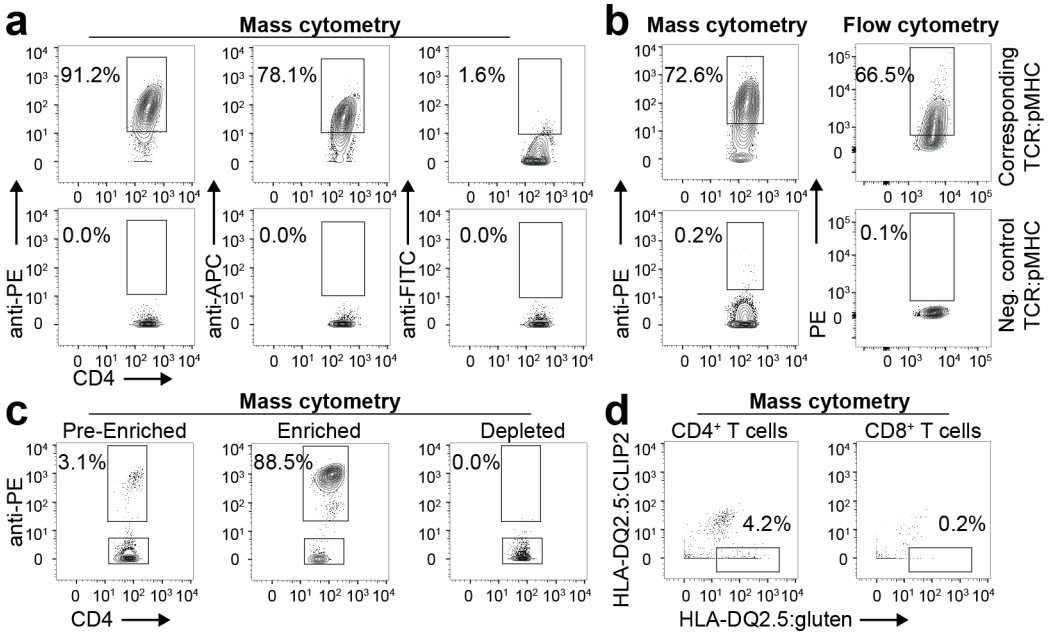
Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data and code availability

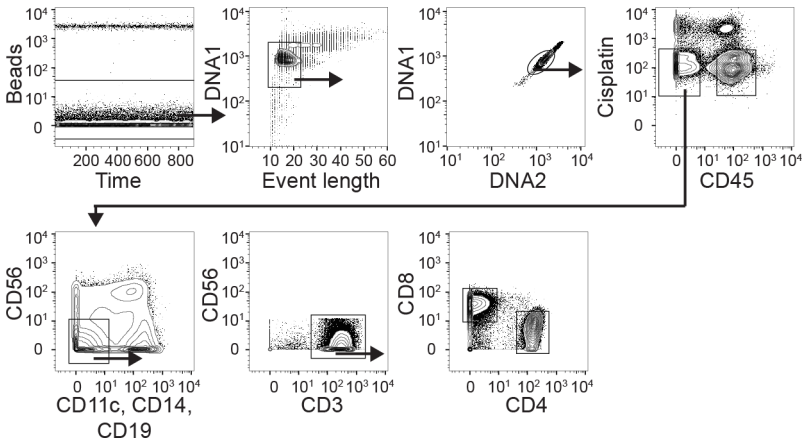
The raw sequences of the RNA-seq data are deposited at the EGA European Genome Phenome Archive (https://ega-archive.org/studies/EGAS00001003017). The source code is available on https://github.com/eivindgl/natmed_gluten_tcell_mass_cytometry. All other data supporting the findings of this study are available from the corresponding authors on request.

References

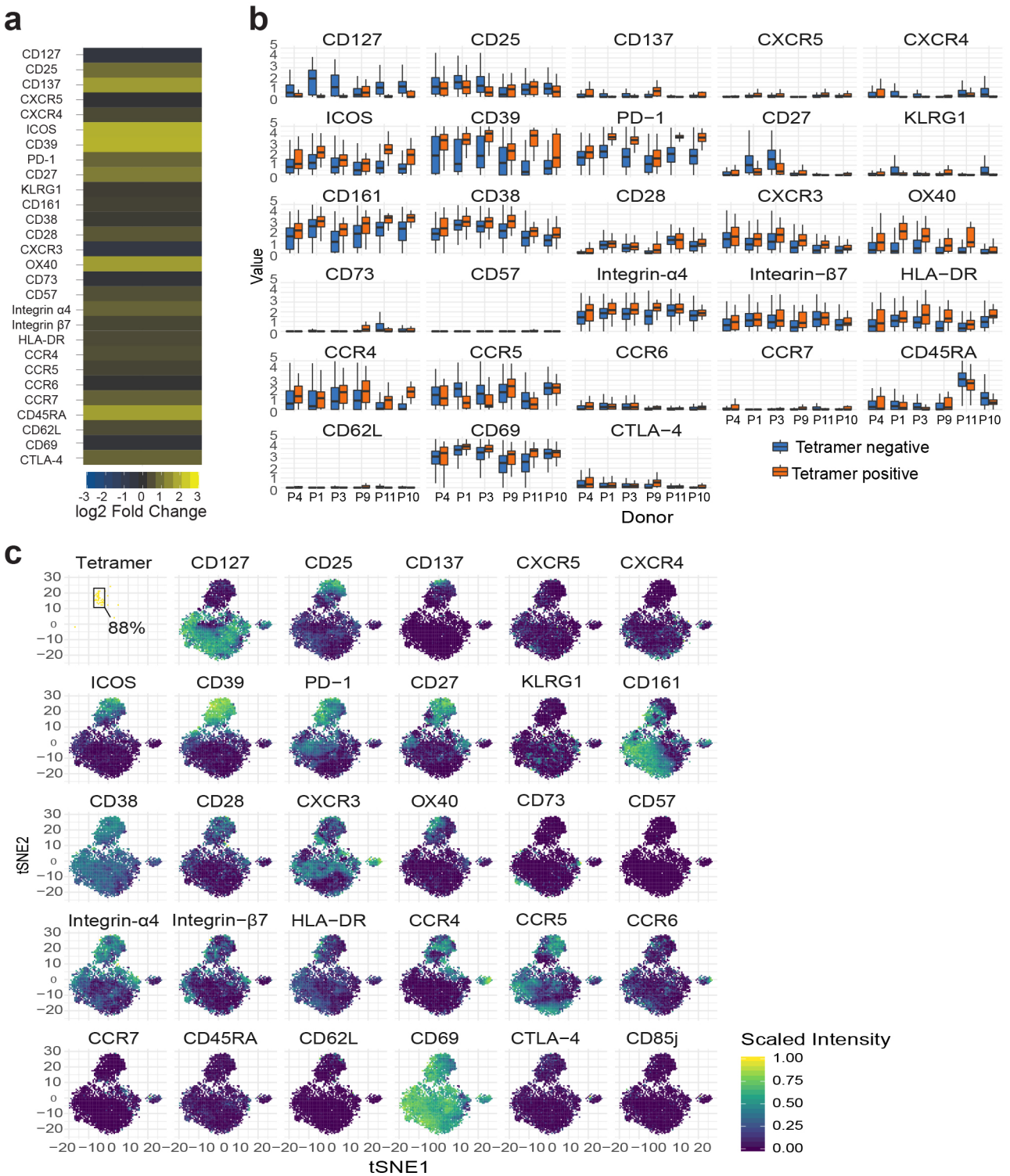
16. Ludvigsson, J. F. et al. *Gut* **63**, 1210–1228 (2014).
17. Marsh, M. N. et al. *Baillieres Clin. Gastroenterol.* **9**, 273–293 (1995).
18. Oberhuber, G. et al. *Eur. J. Gastroenterol. Hepatol.* **11**, 1185–1194 (1999).
19. Anderson, R. P. et al. *Nat. Med.* **6**, 337–342 (2000).
20. van den Hoogen, F. et al. *Ann. Rheum. Dis.* **72**, 1747–1755 (2013).
21. Hochberg, M. C. *Arthritis Rheum.* **40**, 1725 (1997).
22. Newell, E. W. et al. *Immunity* **36**, 142–152 (2012).
23. Newell, E. W. et al. *Nat. Biotechnol.* **31**, 623–629 (2013).
24. Mei, H. E. et al. *J. Immunol.* **194**, 2022–2031 (2015).
25. Shan, L. et al. *Science* **297**, 2275–2279 (2002).
26. Quarsten, H. et al. *J. Immunol.* **167**, 4861–4868 (2001).
27. Christophersen, A. et al. *J. Immunol.* **196**, 2819–2826 (2016).
28. Sollid, L. M. et al. *Immunogenetics* **64**, 455–460 (2012).
29. Molberg, Ø. et al. *Methods Mol. Med.* **41**, 105–124 (2000).
30. Bodd, M. et al. *Eur. J. Immunol.* **43**, 2605–2612 (2013).
31. van der Maaten, L. et al. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
32. Nowicka, M. et al. *F1000Res.* **6**, 748 (2017).
33. Andy, L. et al. *R News* **2**, 18–22 (2002).
34. Patro, R. et al. *Nat. Methods* **14**, 417–419 (2017).
35. Love, M. I. et al. *Genome Biol.* **15**, 550 (2014).



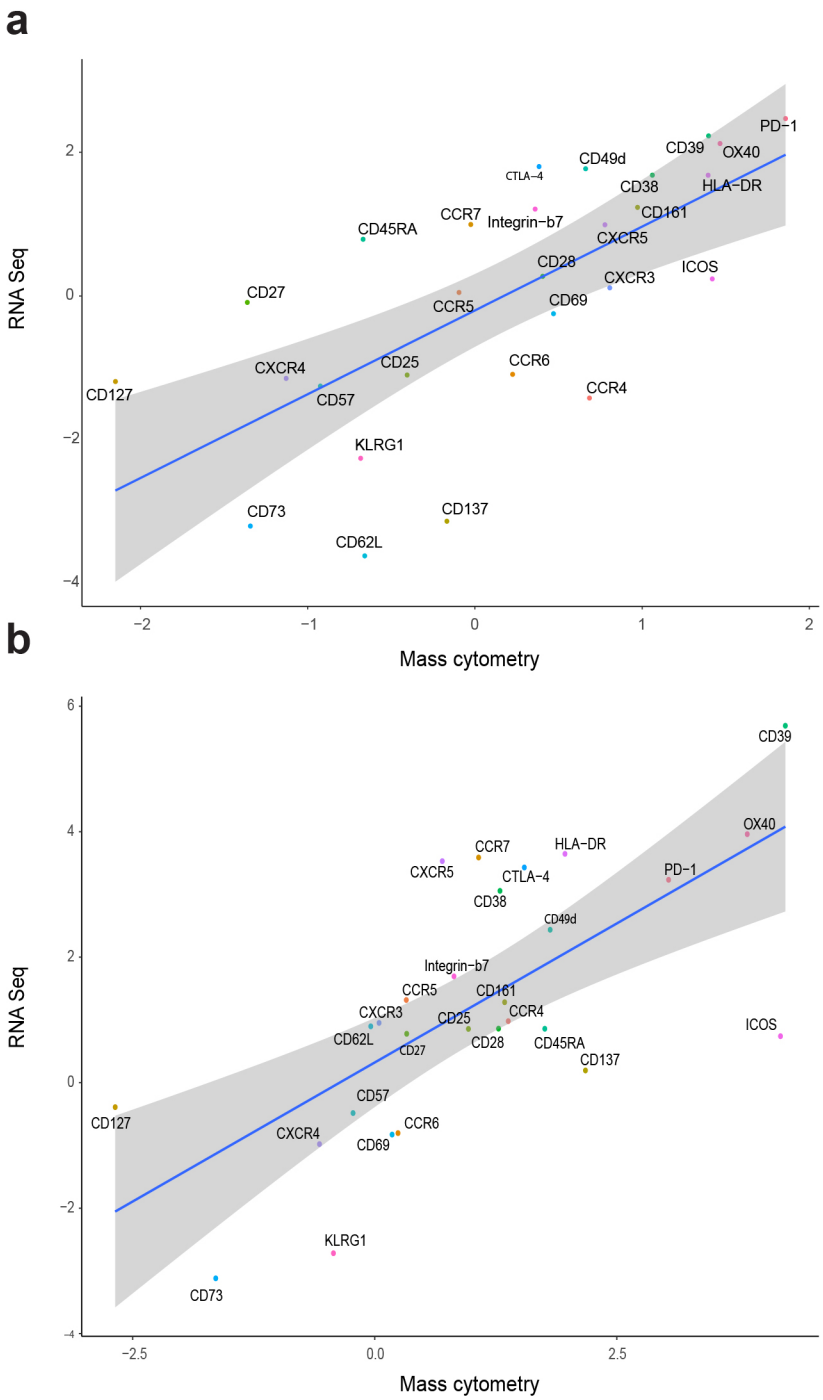
Extended Data Fig. 1 | Establishing HLA class II tetramer staining with mass cytometry. a, Gluten-specific T cell clone binding a corresponding or negative control HLA-DQ2.5:gluten tet reagent metal tagged with secondary binding to PE, APC or FITC (one T cell clone in one experiment). **b**, Comparison of tet staining in mass cytometry and flow cytometry with a gluten-specific T cell clone binding the corresponding or non-corresponding HLA-DQ2.5:gluten tet reagent ($n = 8$ distinct T cell clones in two mass cytometry and two flow cytometry experiments, respectively). **c**, Tet enrichment of a gluten-specific T cell clone binding the corresponding PE-Cy7-coupled HLA-DQ2.5:gluten tet 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 tet binding was excluded with APC-Cy7-coupled HLA-DQ2.5:CLIP2 and metal-tagged anti-APC ($n = 7$ patients with UCeD and 10 controls in 9 experiments. Here, two distinct T cell clones and three PBMC samples were analyzed in two pilot experiments before the tet staining approach was established).



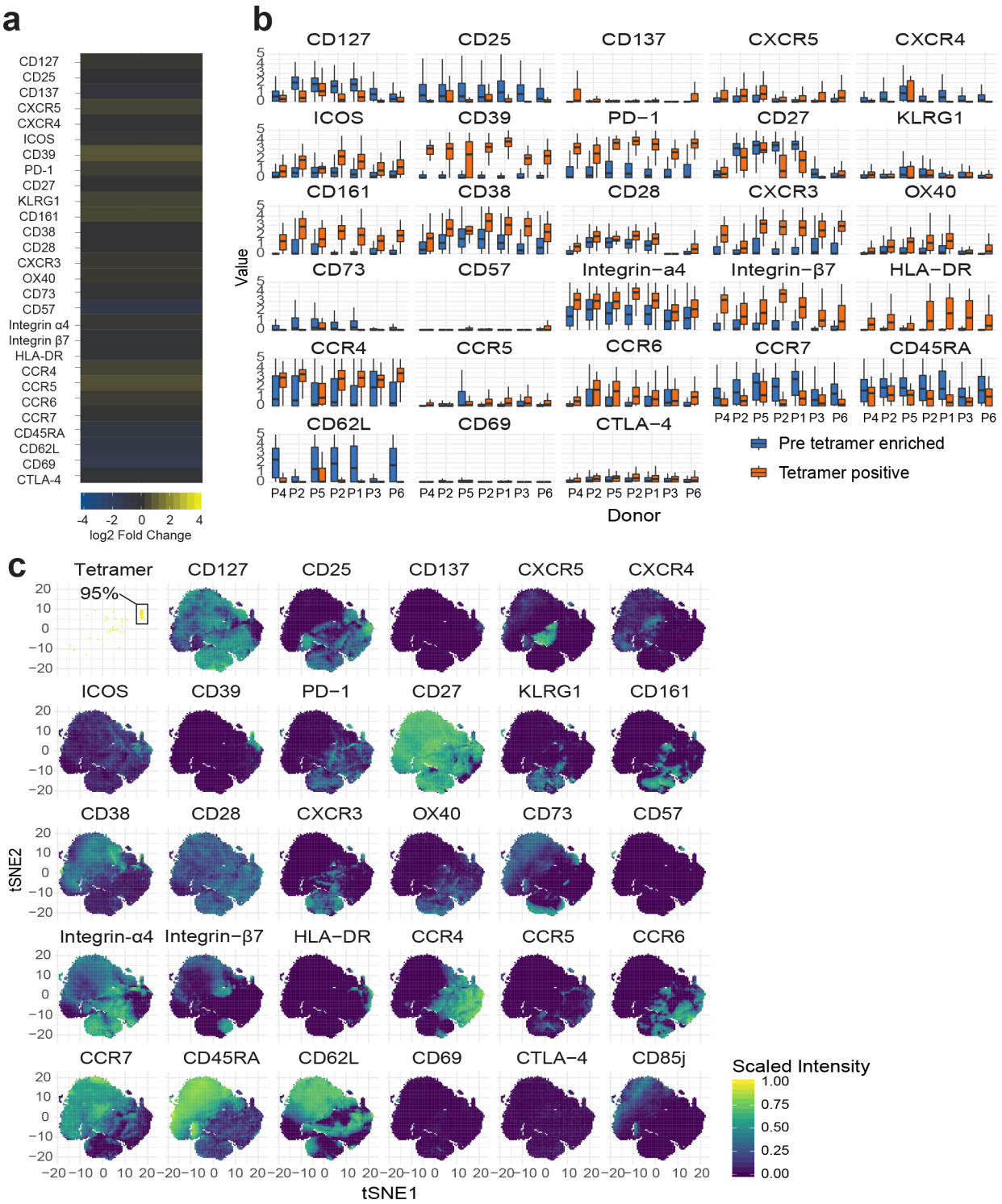
Extended Data Fig. 2 | Gating strategy for cells analyzed with mass cytometry. From the initial plot to the plot and gate that encounters $CD4^+$ blood or gut T cells. Anti-CD45 coupled with 89Y or 108Pd was used for sample barcoding.



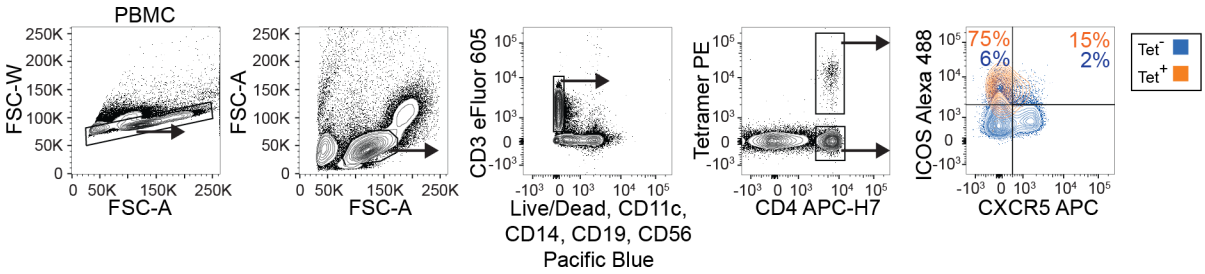
Extended Data Fig. 3 | On the CD4⁺ gut T cells analyzed with mass cytometry. a, Heat map showing the fold-change expression of indicated markers in CD4⁺ HLA-DQ2.5:gluten tet-negative gut T cells of six patients with UCeD versus CD4⁺ gut T cells of seven healthy controls (five experiments in total). **b**, Expression level of mass cytometry panel markers (Supplementary Table 1) in gluten tet-positive and tet-negative CD4⁺ gut T cells in six patients with UCeD (five experiments). The *y* axis indicates the ArcSinh-transformed intensity. Box plots show the median frequency, the interquartile range and the whiskers show largest and smallest values below 1.5 times the interquartile range. **c**, t-SNE plots separately highlighting the presence of cells expressing the markers in **a** and **b** in CD4⁺ gut T cells merged from 1 patient with UCeD and 1 control subject (1 of $n = 6$ patients with UCeD and 1 of $n = 7$ controls, 5 experiments). For comparison, the location of HLA-DQ2.5:gluten tet-binding CD4⁺ gut T cells of the same patient is visualized in the upper left plot.



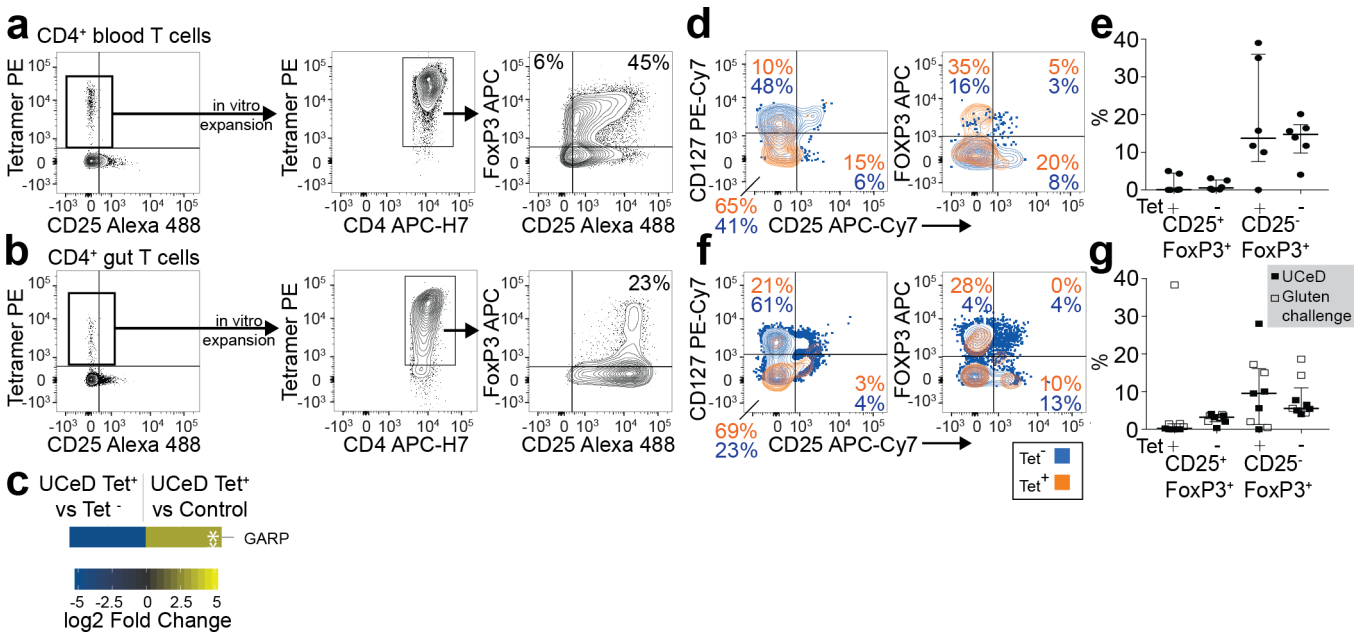
Extended Data Fig. 4 | Mass cytometry and RNA-seq data correlation. a,**b**, Correlation between RNA-seq-derived and mass cytometry-derived fold-change expression of HLA-DQ2.5:gluten tet-positive versus HLA-DQ2.5:gluten tet-negative CD4⁺ gut T cells in patients with UCeD (**a**) and versus CD4⁺ gut T cells in controls (corresponding data depicted as a heat map in Fig. 1e,f) (**b**). For RNA-seq data: $n = 5$ patients with UCeD, 4 control subjects in 2 experiments. For mass cytometry data: $n = 6$ patients with UCeD, 7 controls in 5 experiments. The shaded region indicates the 95% confidence interval around the regression line.



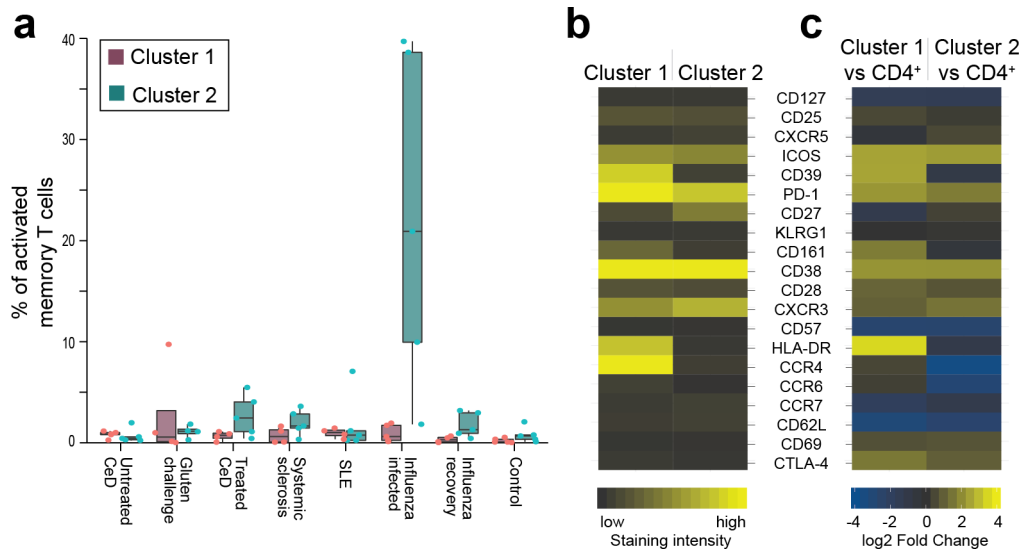
Extended Data Fig. 5 | On the CD4⁺ blood T cells analyzed with mass cytometry. a, Heat map showing the $\log_2$ fold-change expression of indicated markers in CD4⁺ blood T cells of seven patients with UCeD (pre-tet-enriched sample) versus ten healthy controls in nine experiments. **b**, Expression level of mass cytometry panel markers (Supplementary Table 1) in gluten tet-positive and pre-tet-enriched CD4⁺ blood T cells in seven patients with UCeD (six experiments). The *y* axis indicates the ArcSinh-transformed intensity values. The box plots show the median frequency, the interquartile range and the whiskers show the largest and smallest values below 1.5 times the interquartile range. **c**, t-SNE plots separately highlighting the presence of cells expressing the markers in **a** and **b** in CD4⁺ blood T cells merged from 1 healthy control and 1 patient with UCeD (1 of $n=7$ patients with UCeD and 1 of $n=10$ controls in 9 experiments). For comparison, the location of HLA-DQ2.5:gluten tet-binding CD4⁺ blood T cells of the same patient with UCeD is visualized in the upper left plot.



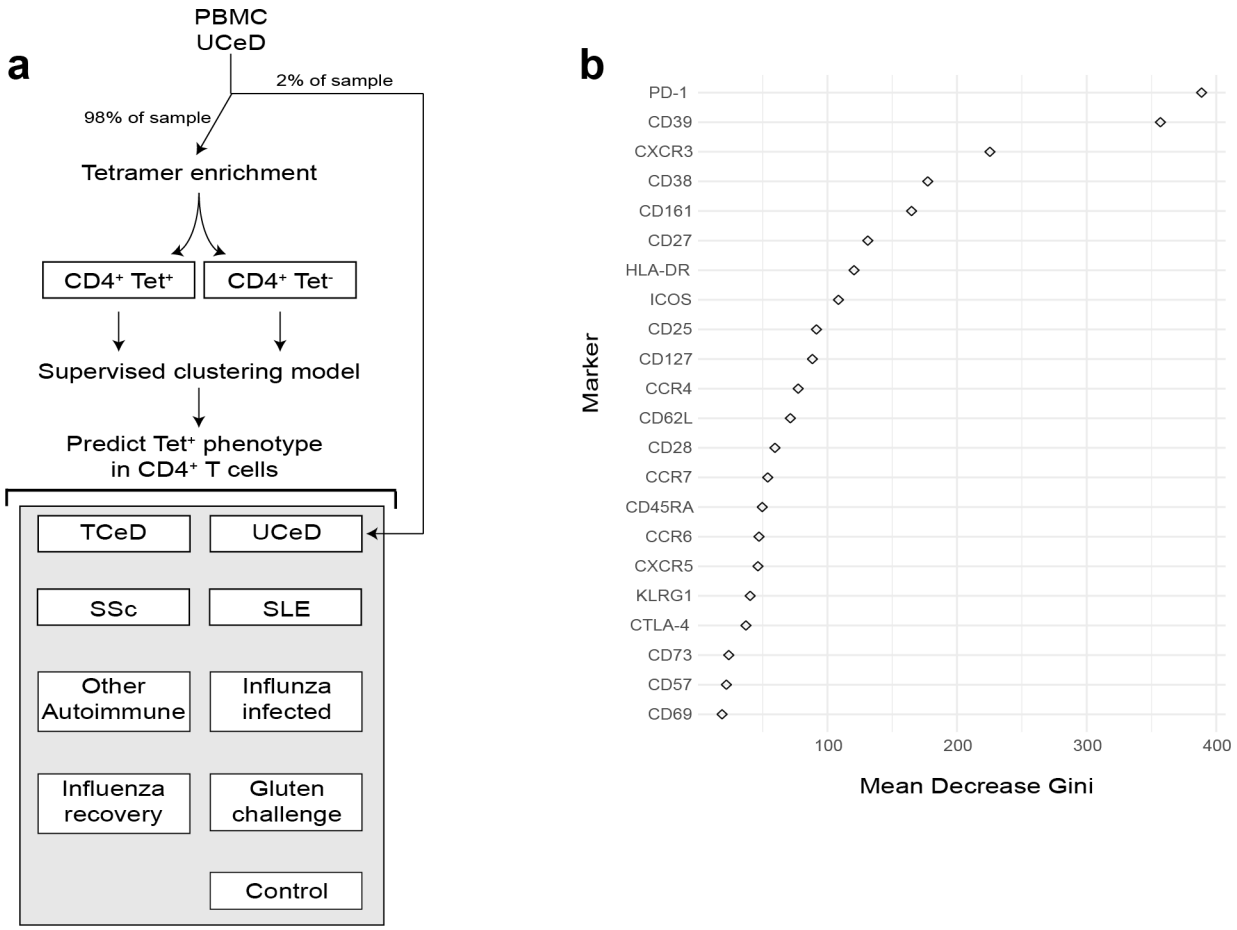
Extended Data Fig. 6 | Flow cytometry staining confirms CXCR5/ICOS expression. General gating strategy for flow cytometry analysis of tet-binding cells, including the expression of CXCR5 and ICOS in tet-positive and tet-negative ($^{+}/^{-}$) $CD4^{+}$ blood T cells in one patient with UCeD (in one experiment).



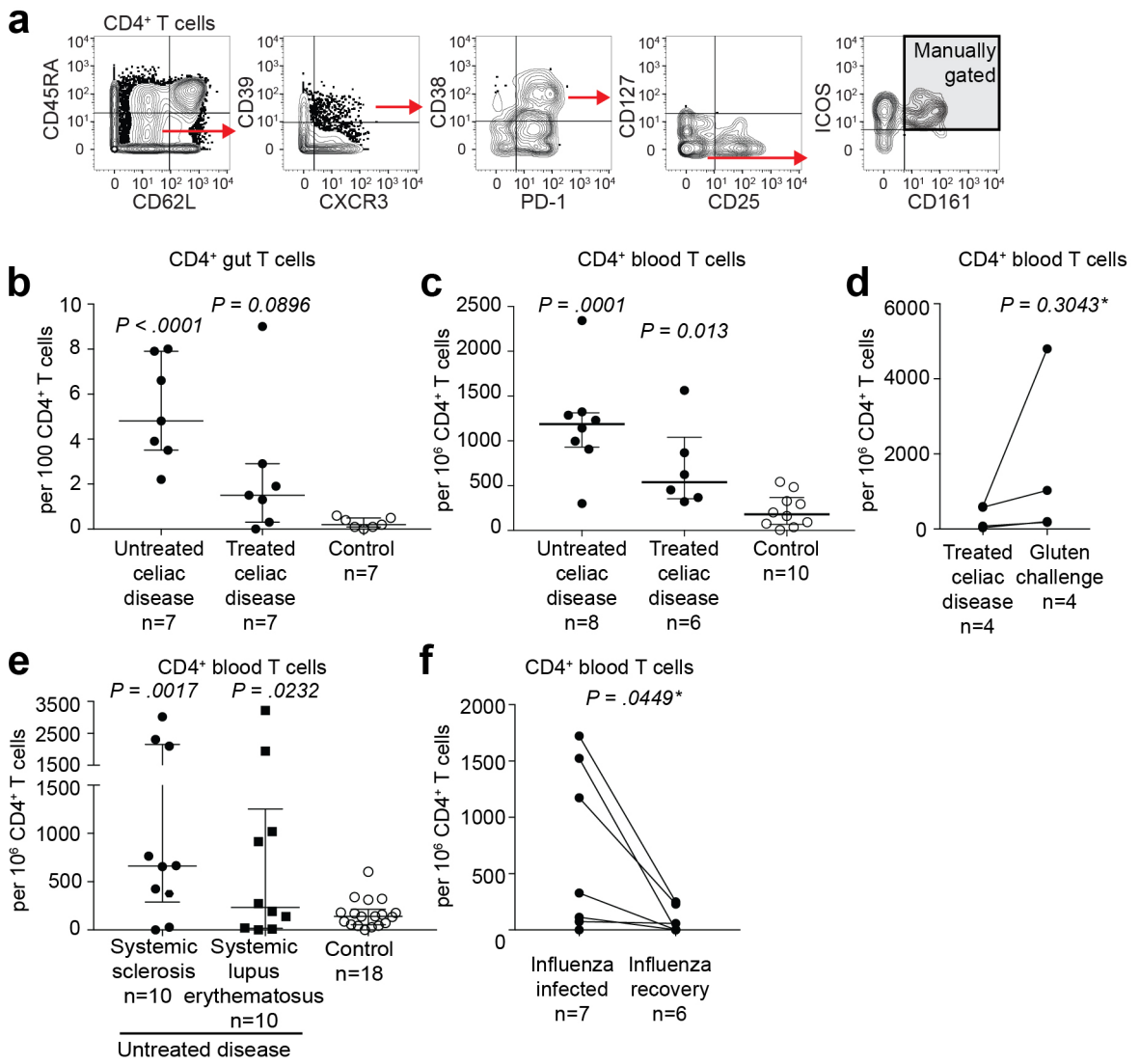
Extended Data Fig. 7 | Expression of regulatory T cell-associated markers on gluten-specific CD4⁺ T cells in vitro and ex vivo. a, CD4⁺ blood T cells of a patient with UCeD were HLA-DQ2.5:gluten tet-sorted ex vivo and cultured in vitro with phytohemmagglutinin and irradiated PBMCs for 2 weeks before re-staining with HLA-DQ2.5:gluten tets to analyze for the expression of Foxp3 and CD25 ($n = 2$, 1 experiment). **b**, The same experiment as in **a**, only with tet-sorted CD4⁺ gut T cells from the patient in **a** ($n = 1$, 1 experiment). **c**, RNA-seq-derived $\log_2$ fold-change expression of glycoprotein A repetitions predominant (GARP) in tet⁺ (<2 GARP transcripts per million) versus tet⁻ CD4⁺ gut T cells of five patients with UCeD and in tet⁺ of five patients with UCeD versus CD4⁺ gut T cells of four control subjects. GARP was differentially expressed in tet⁺ versus tet⁻ cells, but not when compared to CD4⁺ gut T cells in controls (indicated by asterisks) (differentially expressed genes in Supplementary Table 5). **d**,**e**, Ex vivo flow cytometry staining of tet⁺/⁻ CD4⁺ gut T cells from a patient with UCeD with anti-CD127, anti-CD25 and anti-Foxp3 (**d**) and summarized CD25/Foxp3 staining in gut biopsies of five patients with HLA-DQ2.5⁺ UCeD and one patient with UCeD HLA-DQ8⁺ (five experiments) (**e**). The median frequency and interquartile range are indicated. **f**,**g**, Tet⁺/⁻ CD4⁺ blood T cells from a patient with UCeD with anti-CD127, anti-CD25 and anti-Foxp3 (**f**) and summarized CD25/FoxP3 staining in blood of five patients with UCeD and four gluten-challenged patients with UCeD (four experiments) (**g**). The median frequency and interquartile range are indicated. Samples in **a** and **b** were stained with a different anti-CD25 antibody than samples in **d**, **e**, **f**, and **g**.



Extended Data Fig. 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 tet-positive (tet⁺) cells of patients with UCeD are shown. One cluster containing 75% of tet⁺ cells (cluster 1) from seven patients with UCeD and one cluster dramatically upregulated in subjects with influenza infection (cluster 2) are color coded in Fig. 2h. The prevalence of activated CD4⁺ memory T cells belonging to cluster 1 and cluster 2, respectively, for each indicated participant group is shown. Here, we randomly selected $n = 5$ distinct samples per participant group, except for the gluten challenge group for which we only had $n = 4$ distinct samples with sufficient cells. The median frequency and the interquartile range are shown, as well as the whiskers, showing the largest and smallest values below 1.5 times the interquartile range. The single data points depict outliers. **b**,**c**, Heat map of the indicated proteins in cluster 1 and cluster 2 with absolute expression (staining intensity) (**b**) and versus CD4⁺ blood T cells depicted as the $\log_2$ fold change of the grand mean of donor marker intensity (**c**).



Extended Data Fig. 9 | Supervised clustering model predicting the gluten-specific T cell profile. a, Diagram illustrating the workflow for model training and prediction. PBMC samples from donors with UCeD are split into two parts as indicated. One part (right) is not tet enriched and is later used for estimation of gluten-specific T cell profile cell prevalence within the sample. The tet-enriched part (left) is used to train a random forest classification model using repeated K-fold cross-validation on the phenotype of the tet-positive cells. **b**, Scatter plot of the mean decrease in the Gini score for each predictor provides information on how important the predictor variables are to the final model.



Extended Data Fig. 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 tet-binding CD4⁺ T cells in the gut and CD4⁺ effector memory T cells in the blood, respectively, in patients with UCeD (although the gluten-specific cells were phenotypically similar, not all of the cells had a staining intensity for all ten markers above or below the manually set threshold, as also visualized in Figs. 1c and 2c). Here, visualized in the peripheral blood of a patient with UCeD. **b**–**d**, Frequency of cells gated as in **a** in the gut (**b**) and in the blood (**c**) (pre-tet-enriched sample) of patients with UCeD, TCeD (gluten-free diet) and healthy controls, and in patients with TCeD before and following gluten challenge (differing from gating encountering gluten-specific cells in patients with UCeD chiefly by lower CD39 expression, as also visualized in Fig. 2f) (**d**). Blood and gut samples analyzed in 12 and six experiments, respectively. Gluten challenge samples were analyzed in two experiments. **e**,**f**, Frequency of cells gated as in **a** within patients with the indicated autoimmune disorders and different set as in **b** of control subjects (seven experiments in total) (**e**), and within a cohort during and after influenza infection (**f**) (three experiments in total). Statistics: an unpaired, two-tailed *t*-test was used and the median frequency and interquartile range are indicated in **b**, **c** and **e**. *A paired, two-tailed *t*-test was used and the lines indicate paired samples in **d** and **f**.

nature research

Corresponding author(s): Ludvig M Sollid
Mark M Davis

Last updated by author(s): Jan 31, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Flow cytometry data were collected on a LSR II instrument or a FACS Aria II cell sorter with BD FACSDiva Software, version 8. Mass cytometry data were acquired on a Helios instrument with CyTOF Software, version 6.7. RNA-seq data, amplicon libraries were sequenced on NextSeq500 (Illumina, Inc) at the Norwegian Sequencing Center (http://www.sequencing.uio.no).
Data analysis	We analyzed flow cytometry and mass cytometry data using FlowJo, version 10.4. We used the GraphPad Prism 7 software (GraphPad Software, Inc) for statistical analysis and visualization of cell frequencies. Further, mass cytometry data (fcs-files from Flowjo analysis containing only the CD4+ blood or gut T cells) were loaded into R using the flowCORE package, and visualized in heat maps, boxplots or a cell marker importance function, using ggplot2, and in t-SNE plots, using the Rtsne package. We applied the the FlowSOM and ConsesusClusterPlus packages for unsupervised clustering of activated memory CD4+ T cells. We trained a random forest model, using caret version 6.0-79, to obtain an unbiased prevalence estimate of CD4+ T cells with a phenotype highly similar to the gluten specific CD4 + T cells and the doMC package, version 1.3.5, to parallelize model training. We mapped RNA-seq reads (76 bp paired end) to the human reference genome GRCh38.p7 with gene annotations curated by Ensembl release 86, using Salmon version 0.7.2. The raw sequencing data were processed on the TSD-platform, owned by the University of Oslo. 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. We used we used DESeq2 to call differentially expressed genes and estimate fold change and ggplot2 for visualization.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on 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 (flow cytometry and mass cytometry) are available from the authors upon request. All figures have associated raw data, except for the illustration in Extended Data Fig. 9a.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Access to blood and tissue samples with a sufficient number of live cells from celiac patients, control subjects and patients with other autoimmune conditions is limited. To be able to determine test significance we chose sample size of four or more when calculating differences in cell frequencies.
Data exclusions	We excluded one individual from the mass cytometry analysis and two individuals from the RNA seq analysis from the group of celiac disease patients. We recruited patients before their final diagnosis were made, and we decided to exclude these three subjects early on in the study. Two of the subjects (one for mass cytometry and one for RNA seq) were classified to have potential celiac disease (having slightly positive disease-specific serology but normal gut histology) which according to the established diagnostic guidelines is not celiac disease. We excluded one subject with celiac disease from the RNA seq analysis. On the diagnostic workup, this subject was found to have inflammation in bulbus duodeni only and not in pars descendens duodeni where the biopsies for RNA seq analysis were taken. We deemed it was important to analyze only gut biopsies of untreated celiac patients that were classified to have inflammation. We excluded a group of five patients with other autoimmune diseases (type I diabetes, hashimoto thyroeoiditis, polymyositis and mixed connective tissue disease) included in a previous version of the manuscript as per request by reviewer. The reviewer argued that this group was too small and heterogeneous to allow a meaningful interpretation. We also excluded one patient in the cohort of influenza-infected participants due to comorbidities including a clonal T-cell population consistent with large granular lymphocytic leukemia, chronic joint pains, gastritis and a small-cell B-cell lymphoma.
Replication	After establishment of the protocol, all attempts to replicate tetramer staining with mass cytometry in untreated celiac disease patients, T-cell lines and T-cell clones were successful. This was also the case for additional samples included in the revised version of the manuscript. All mass cytometry experiments were carefully controlled. A gluten-specific T-cell clone was regularly used as positive control and PBMCs from a healthy donor as a negative control and as control for the antibody panel. Pre-tetramer enriched, and tetramer-negative samples served as additional internal control samples. The number of tetramer-binding cells was lower or absent in treated celiac disease patients and controls, hence in agreement with published data using flow cytometry. In addition, results from mass cytometry, RNA seq and flow cytometry gave almost identical results, using different sets of patients and controls and finally, mass cytometry and flow cytometry results obtained at the core facilities at Stanford University could be replicated (with different patient samples), giving highly similar results at the flow core facilities at the University of Bergen and the University of Oslo, respectively.
Randomization	For mass cytometry and flow cytometry analysis, we included at the same point of analysis at least one patient sample and one control sample. Potential covariates were controlled for by including the same healthy control sample to mass cytometry stainings. The RNA-seq data was acquired in two rounds. In the first round, four potential candidates for celiac disease were included (two later excluded, see above). In the second round, three celiac disease patients and four control subjects were included.
Blinding	The researchers were not blinded during data acquisition or analysis. However, in the case of celiac patients, only the serology parameters prior to referral to the hospital was known and diagnosis was given weeks after acquisition of data with mass cytometry or RNA seq.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☐	☒ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

Target - Label - Clone - Supplier - Catalogue number - Final concentration
anti-CD45 - 89Y - HI30 - Fluidigm - 304045 - 2:100
anti-CD45 - 108Pd - HI30 - Biolegend - 304045 - 8 µg/ml
anti-CD57 - 115In - HCD57 - Biolegend - 322325 - 1.5 µg/ml
anti-CD28 - 139La - CD28.2 - Biolegend - 302937 - 4 µg/ml
anti-Intebrin-α4/CD49d - 141Pr - 9F10 - Fluidigm - 3141004B - 1:100
anti-KLRG1 - 142Nd - 13F12F2 - Thermo Fisher Scientific - 16-9488-85 - 3 µg/ml
anti-CD278/ICOS - 143Nd - C398.4A - Fluidigm - 3143025B - 0.5:100
anti-CD38 - 144Nd - HIT2 - Fluidigm - 3144014B - 1.5:100
anti-CD4 - 145Nd - RPA-T4 - Fluidigm - 3145001B - 0.5:100
anti-CD8a - 146Nd - RPA-T8 - Fluidigm - 3146001B - 0.6:100
anti-CD137 (41BB) - 147Sm - 4-1BB - R&D systems - AF838 - 12 µg/ml
anti-CD27 - 148Nd - O323 - Biolegend - 302839 - 1 µg/ml
anti-CD56 (NCAM) - 149Sm - NCAM16.2 - Fluidigm - 3149021B - 0.5:100
anti-CD127 - 150Nd - A019D5 - Biolegend - 351337 - 1 µg/ml
anti-CD11c - 151Eu - Bu15 - Biolegend - 337221 - 2 µg/ml
anti-CD19 - 151Eu - HIB19 - Biolegend - 302247 - 1 µg/ml
anti-CD14 - 151Eu - M5E2 - Fluidigm - 3151009B - 1:100
anti-CD244 - 152Sm - 2B4 - R&D systems - AF1039 - 4 µg/ml
anti-CD62L - 153Eu - DREG-56 - Fluidigm - 3153004B - 0.5:100
anti-CD3 - 154Sm - UCHT1 - Fluidigm - 3154003B - 0.8:100
anti-CD279 (PD-1) - 155Gd - EH12.2H7 - Fluidigm - 3155009B - 1.8:100
anti-CD195 (CCR5) - 156Gd - NP-6G4 - Fluidigm - 3156015A - 4:100
anti-CD194 (CCR4) - 158Gd - L291H4 - Fluidigm - 3158032A - 0.5:100
anti-CD161 - 159Tb - HP-3G10 - Fluidigm - 3159004B - 0.5:100
anti-CD39 - 160Gd - A1 - Fluidigm - 3160004B - 1:100
anti-CD152 (CTLA-4) - 161Dy - 14D3 - Fluidigm - 3161004B - 5:100
anti-Integrin-β7 - 162Dy - FIB504 - Fluidigm - 3162026B - 0.5:100
anti-CD183 (CXCR3) - 163Dy - G025H7 - Fluidigm - 3163004B - 0.75:100
anti-OX40 (CD134) - 164Dy - Ber-ACT35 - Biolegend - 350015 - 8 µg/ml
anti-Phycoerythrin - 165Ho - PE001 - Fluidigm - 3165015B - 1.25:100
anti-CD85j/ILT2 - 166Er - GHI/75 - Fluidigm - 3166020B - 1:100
anti-CD197 (CCR7) - 167Er - G043H7 - Fluidigm - 3167009A - 1:100
anti-CD73 - 168Er - AD2 - Fluidigm - 3168015B - 1:100
anti-CD25 (IL-2R) - 169Tm - 2A3 - Fluidigm - 3169003B - 0.6:100
anti-CD45RA - 170Er - HI100 - Fluidigm - 3170010B - 0.1:100
anti-CD185 (CXCR5) - 171Yb - RF8B2 - Fluidigm - 3171014B - 0.75:100
anti-CD69 - 172Yb - FN50 - Biolegend - 310939 - 1 µg/ml
anti-HLA-DR - 173Yb - L243 - Fluidigm - 3173005B - 0.75:100
anti-CD196 (CCR6) - 174Yb - G034E3 - Biolegend - 353427 - 1 µg/ml
anti-CD184 (CXCR4) - 175Lu - 12G5 - Fluidigm - 3175001B - 1.25:100
anti-Allophycocyanin - 176Yb - APC003 - Fluidigm - 3176007B - 1.25:100
anti-CD11b - 209Bi - ICRF44 - Fluidigm - 3209003B - 0.4:100
anti-TIGIT - 147Sm - 372702 - Biolegend - 372702 - 8 µg/ml
anti-TCRg/d - 152Sm - 11F2 - Fluidigm - 3152008B - 1:100
anti-CD29 - 156Gd - TS2/16 - Biolegend - 303021 - 1.6 µg/ml
anti-CD200 - 164Dy - OX-104 - Biolegend - 329219 - 8 µg/ml
anti-CCR2 - 166Er - K036C2 - Biolegend - 357202 - 2 µg/ml
anti-CX3CR1 - 175Lu - 2A9-1 - Biolegend - 341602 - 6 µg/ml
anti-TCRα/β - 176Yb - IP26 - Fluidigm - 3176015B - 1.5:100

anti-FoxP3 - APC - PCH101 - Thermo Fisher Scientific - 17-4776-42 - 5 µg/ml
anti-CD4 - Alexa 700 - A161A1 - Biolegend - 357418 - 3:100
anti-CD25 - APC-Cy7 - BC96 - Biolegend - 302614 - 4:100
anti-CD45RA - PE-Cy5 - HI100 - Biolegend - 304110 - 1:100
anti-CD127 - PE-Cy7 - A019D5 - Biolegend - 351320 - 5:100
anti-Ki-67 - Alexa 488 - Ki-67 - Biolegend - 350532 - 5 µg/ml
anti-CD11c - Pacific Blue - 3.9 - Biolegend - 301626 - 1.5:100

anti-CD56 - Pacific Blue - 5.1H11 - Biolegend - 362520 - 1.5:101
anti-CD14 - Pacific Blue - HCD14 - Biolegend - 325616 - 1.5:102
anti-CD19 - Pacific Blue - 6D5 - Biolegend - 115523 - 1.5:103
anti-CD3 - BV605 - UCHT1 - Biolegend - 300460 - 4:100
anti-Integrin β7 - BV650 - FIB504 - BD Bisciences - 564284 - 5:100
anti-CD278/ICOS - Alexa 488 - C398.4A - Biolegend - 313514 - 1 µg/ml
anti-CD62L - PerCP - DREG-56 - Biolegend - 304824 - 3:100
anti-CD45RA - Pe-Cy7 - HI100 - Biolegend - 304126 - 2.5:100
anti-CXCR4 - APC - 12G5 - Biolegend - 306510 - 1:100
anti-CXCR5 - APC - J252D4 - Biolegend - 356908 - 1:100
anti-CD4 - APC-H7 - SK3 - BD Bisciences - 641398 - 0.36 µg/ml
anti-CD3 - BV605 - OKT3 - Biolegened - 317322 - 3:100
anti-Alexa 488 - CD25 - BC96 - Thermo Fisher Scientific - 53-0259-42 - 4:100

anti-CD3 - BV570 - UCHT1 - Biolegend - 300436 - 4:100
anti-CD11c - Horizon V450 - B-ly6 - BD Bisciences - 560369 - 3:100
anti-CD27 - PE-Cy7 - LG.7F9 - eBioscience - 25-0271-82 - 0.2:100
anti-IgA - FITC - not given - Southern Biotech - 2050-02 - 2 µg/ml

Validation

Each antobody was titrated using PBMCs from a healthy donor prior to establishment of staining protocols, following manufacturer's instructions. PBMCs from one healthy donor was used as an internal control for the antibody staining panel in mass cytometry experiments.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

T-cell clones were established in the Sollid lab from HLA-DQ2.5:gluten tetramer sorted CD4+ blood or gut T cells of celiac disease patients or from single-cell suspensions of duodenal gut biopsies of celiac disease patients as also described in Online Methods.

Authentication

Each T-cell clone used was generated by the first author of this study. All incoming patients are numbered. The numbering of T-cell clones is based on the patient sample from which they were derived. The T-cell clones were generated in the Sollid Lab (Univeristy of Oslo). Their specificity was verified with tetramer staining and in proliferation assays to the corresponding gluten peptide. They have not been authenticated externally, e.g. through STR profiling.

Mycoplasma contamination

All generated T-cell clones tested negative for mycoplasma contamination.

Commonly misidentified lines
(See ICLAC register)

No commonly misidentified cell lines were used.

Human research participants

Policy information about studies involving human research participants

Population characteristics

All participants, including relevant clinical information, are listed in Supplementary Table 3, 6 and additional information is given in the methods section. The study is not designed such that analysis of and thus information on sex, age or ethnicity are highly relevant to the development of the technique for HLA Class II tetramer staining combined with mass cytometry, nor for the detection of gluten-specific CD4+ T cells in blood and gut of celiac disease patients. We do not have information on age, sex or ethnicity for the blood bank donors. Thus, we cannot exclude that there are differences regarding these parameters, , yet all examined subjects were adults.

Recruitment

Buffy coats (containing PBMCs) were obtained from anonymous blood donors at the Stanford blood center or Oslo University Hospital (blood bank). Celiac disease patients including non-celiac patients undergoing gastroduodenoscopy, from whom we received blood and duodenal biopsies, were recruited through the endoscopy unit at Oslo University Hospital. For the gluten challenge study, we recruited participants through announcements on hospital employee web sites and in fora managed by a patient interest organization for celiac disease. Material from gluten-challenged patients was collected at the endoscopy unit at Oslo University Hospital. Blood samples from participants with systemic sclerosis and systemic lupus erythematosus were recruited from the Rheumatology Departments at Oslo University Hospital and the Immunology and Rheumatology Division at the Department of Medicine at Stanford Hospital. We recruited participants with influenza virus infection from the Emergency Department and the Express Outpatient Clinic at Stanford Hospital. As stated and referred to in Online Methods, all included patients needed to fulfill the diagnostic criteria for celiac disease, systemic scloeris or systemic lupus erythematosus, respectively. All participants were adults, but they were not matched for sex, age or ethnicity, which potentially could impact the results.

Ethics oversight

The studies incling material from celiac disease patients was approved by the the Regional Committee for Medical and Health Research Ethics South-East Norway (2010/2720 and 2013/1237). The study influenza-infected individuals was approved by the Stanford University Administrative Panels on Human Subjects in Medical Research, covered by IRB 22442 (Immune Responses to Influenza-like Illness). The recruitment of patients with systemic sclerosis and systemic lupus erythematosus 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).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	Although this study involves clinical data, it it not a clinical trial. However, blood samples from five gluten challenge patients, recruited as part of a clinical trial: Clinicaltrials.gov identifier: NCT02464150, were used for mass cytometry analysis.
Study protocol	The protocols are approved by the regional ethics committees but are not publicly available. They are are available from the authors upon request
Data collection	Participants were recruited and material collected in 2015-2017 in locations as described above (see 'recruitment'). Samples were analyzed in 2016-2018 at The Human Immune Monitoring Center (Stanford University), the Flow Cytometry Core Facility at Oslo University Hospital – Rikshospitalet and Radiumhospitalet, the Flow Core Facility at the University of Bergen, the Institute for Immunity, Transplantation and Infection at Stanford University School of Medicine and at the Department of Immunology (Oslo University Hospital and Univeristy of Oslo).
Outcomes	In the study we used clinical data. However, outcomes were not predefined as in a clinical trial.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	These steps are described in detail in the manuscript's Online Methods. For flow cytometry studies, PBMCs were separated from blood with Lymphoprep according to the manufacturer's protocol and single-cell suspensions were generated from duodenal biopsies using 1% collagenase. Cell samples were frozen in 10% DMSO and later thawed in 20% FCS/RPMI. The samples were then stained with Fc receptor blocker on ice followed by HLA-DQ2.5:gluten tetramers or HLA-DQ8: gluten tetramers for 40 minutes at room temperature. In the case of tetramer-stained blood samples, these were stained with anti-PE antibody-coated magnetic beads for 20 minutes on ice and then enriched on a magnetized LS column (Miltenyi Biotec) to increase the frequency of tetramer-binding cells within the cell sample. The samples were then stained with live/dead staining according to manufacturer's protocol and then on ice for 20 minutes with a cocktail of cell surface antibodies. Staining of intracellular Ki-67 and FoxP3 was done subsequently according to the manufacturer's protocol. See the Online Methods section of this paper for details on mass cytometry staining.
Instrument	We used either a LSR II instrument or a FACS Aria II cell sorter. Mass cytometry data were acquired on a Helios instrument.
Software	Flow cytometry data were analyzed with FlowJo version 10.4. Mass cytometry data were analyzed with FlowJo version 10.4 to exclude cells that were not CD4+ T cells followed by analysis R using different packages specified by name and version in the methods section of this paper and in this Reporting Summary (see "Software and Code").
Cell population abundance	Tetramer-sorted and in vitro cultured T-cell lines and T-cell clones were purity checked thorough re-staining with the relevant tetramers after in vitro culturing. Purity of cells sorted for RNA seq was not checked after sorting due to the low number of obtained cells.
Gating strategy	For mass cytometry, the gating strategy is visualized in Extended Data Fig. 2 and for the calculation of the frequency of cells with the phenotype of gluten-specific CD4+ T cells, in Extended Data Fig. 10a. For flow cytometry, we used the following gating procedure: Forward scatter vs forward scatter width; forward scatter vs side scatter; Dump (CD11c, CD14, CD19, CD56 and live/dead) vs CD3; CD4; CD4 vs tetramer. We included the gating strategy for flow cytometry in Extended Data Fig. 6.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.