

# Online Research @ Cardiff

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository: <https://orca.cardiff.ac.uk/id/eprint/120122/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Mayassi, Toufic, Ladell, Kristin ORCID: <https://orcid.org/0000-0002-9856-2938>, Gudjonson, Herman, McLaren, James E. ORCID: <https://orcid.org/0000-0002-7021-5934>, Shaw, Dustin G., Tran, Mai T., Rokicka, Jagoda J., Lawrence, Ian, Grenier, Jean-Christophe, van Unen, Vincent, Ciszewski, Cezary, Dimaano, Matthew, Sayegh, Hoda E., Kumar, Vinod, Wijmenga, Cisca, Green, Peter H.R., Gokhale, Ranjana, Jericho, Hilary, Semrad, Carol E., Guandalini, Stefano, Dinner, Aaron R., Kupfer, Sonia S., Reid, Hugh H., Barreiro, Luis B., Rossjohn, Jamie ORCID: <https://orcid.org/0000-0002-2020-7522>, Price, David A. ORCID: <https://orcid.org/0000-0001-9416-2737> and Jabri, Bana 2019. Chronic inflammation permanently reshapes tissue-resident immunity in celiac disease. *Cell* 176 (5), 967-981.e19. 10.1016/j.cell.2018.12.039 file

Publishers page: <http://dx.doi.org/10.1016/j.cell.2018.12.039>  
<<http://dx.doi.org/10.1016/j.cell.2018.12.039>>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies.  
See

<http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



**Title**

Chronic inflammation permanently reshapes tissue-resident immunity in celiac disease

**Authors**

Toufic Mayassi<sup>1,2</sup>, Kristin Ladell<sup>3</sup>, Herman Gudjonson<sup>4,5</sup>, James E. McLaren<sup>3</sup>, Dustin G. Shaw<sup>1,2</sup>, Mai T. Tran<sup>6</sup>, Jagoda J. Rokicka<sup>2</sup>, Ian Lawrence<sup>2</sup>, Jean-Christophe Grenier<sup>7</sup>, Vincent van Unen<sup>8</sup>, Cezary Ciszewski<sup>2</sup>, Matthew Dimaano<sup>2</sup>, Hoda E. Sayegh<sup>2</sup>, Vinod Kumar<sup>9</sup>, Cisca Wijmenga<sup>9</sup>, Peter H.R. Green<sup>10</sup>, Ranjana Gokhale<sup>11,12</sup>, Hilary Jericho<sup>11,12</sup>, Carol E. Semrad<sup>2,12</sup>, Stefano Guandalini<sup>11,12</sup>, Aaron R. Dinner<sup>4,5,13</sup>, Sonia S. Kupfer<sup>2,11</sup>, Hugh H. Reid<sup>6,14</sup>, Luis B. Barreiro<sup>7</sup>, Jamie Rossjohn<sup>3,6,14</sup>, David A. Price<sup>3,\*</sup>, Bana Jabri<sup>1,2,15,\*,#</sup>

**Affiliations**

<sup>1</sup>Committee on Immunology, University of Chicago, Chicago, IL, USA

<sup>2</sup>Department of Medicine, University of Chicago, Chicago, IL, USA

<sup>3</sup>Division of Infection and Immunity, Cardiff University School of Medicine, Cardiff, UK

<sup>4</sup>Institute for Biophysical Dynamics, University of Chicago, Chicago, IL, USA

<sup>5</sup>Department of Chemistry, University of Chicago, Chicago, IL, USA

<sup>6</sup>Infection and Immunity Program and Department of Biochemistry and Molecular Biology, Biomedicine Discovery Institute, Monash University, Clayton, Victoria, Australia

<sup>7</sup>Department of Genetics, CHU Sainte-Justine Research Center, Montreal, Quebec, Canada

<sup>8</sup>Department of Immunohematology and Blood Transfusion, Leiden University Medical Center, Leiden, Netherlands

<sup>9</sup>Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, Netherlands

<sup>10</sup>Celiac Disease Center, College of Physicians and Surgeons, Columbia University Medical Center, New York, NY, USA

<sup>11</sup>Section of Gastroenterology, Hepatology, and Nutrition, Department of Pediatrics, University of Chicago, Chicago, IL, USA

<sup>12</sup>University of Chicago Celiac Disease Center, University of Chicago, Chicago, IL, USA

<sup>13</sup>James Franck Institute, University of Chicago, Chicago, IL, USA

<sup>14</sup>Australian Research Council Centre of Excellence in Advanced Molecular Imaging, Monash University, Clayton, Victoria, Australia

<sup>15</sup>Department of Pathology, University of Chicago, Chicago, IL, USA

\*Senior author

#Lead contact

### **Correspondence**

\*David A Price, Division of Infection and Immunity, Cardiff University School of Medicine, Heath Park, University Hospital, Cardiff CF14 4XN, Wales, UK. Tel: (44) 29 2068 7002.

E-mail: [priced6@cardiff.ac.uk](mailto:priced6@cardiff.ac.uk)

\*Bana Jabri, University of Chicago, Departments of Medicine, Pathology, and Pediatrics, 900 East 57<sup>th</sup> Street KCBD, Chicago, IL 60637, USA. Tel: (773) 834-8670. E-mail:

[bjabri@bsd.uchicago.edu](mailto:bjabri@bsd.uchicago.edu)

### **Running title**

Tissue-resident immunity in celiac disease

### **Summary**

Tissue-resident lymphocytes play a key role in immune surveillance, but it remains unclear how these inherently stable cell populations respond to chronic inflammation. In the setting of celiac disease (CeD), where exposure to dietary antigen can be controlled, gluten-induced inflammation triggered a profound depletion of naturally-occurring  $V\gamma 4^+/V\delta 1^+$  intraepithelial lymphocytes (IELs) with innate cytolytic properties and specificity for the butyrophilin-like (BTNL) molecules BTNL3/BTNL8. Creation of a new niche with reduced expression of BTNL8 and loss of  $V\gamma 4^+/V\delta 1^+$  IELs was accompanied by the expansion of gluten-sensitive, interferon- $\gamma$ -producing  $V\delta 1^+$  IELs bearing T cell receptors (TCRs) with a shared non-germline-encoded motif that failed to recognize BTNL3/BTNL8. Exclusion of dietary gluten restored BTNL8 expression but was insufficient to reconstitute the physiological  $V\gamma 4^+/V\delta 1^+$  subset among  $TCR\gamma\delta^+$  IELs. Collectively, these data show that chronic inflammation permanently reconfigures the tissue-resident  $TCR\gamma\delta^+$  IEL compartment in CeD.

### **Keywords**

Butyrophilin-like molecules, celiac disease,  $\gamma\delta$  T cells, intraepithelial lymphocytes, tissue-resident lymphocytes.

### **Introduction**

Tissue-resident lymphocytes have been investigated extensively under steady state conditions and during the induction of local memory populations in response to acute infections (Mueller and Mackay, 2016). In mice, the tissue-resident  $TCR\alpha\beta^+ CD8\alpha\beta^+$  pool is highly stable and responds to secondary antigenic challenge via local proliferation of pre-existing memory cells (Beura et al., 2018; Park et al., 2018), which endure over time despite the accumulation of new tissue-resident populations driven by subsequent

infections (Park et al., 2018). However, it remains unclear if chronic inflammation can permanently reconfigure the tissue-resident T cell compartment.

Intraepithelial lymphocytes (IELs) expressing  $\gamma\delta$  T cell receptors (TCRs) are tissue-resident T cells that play a key role in immune surveillance via dynamic scanning of the intestinal epithelium (Hoytema van Konijnenburg et al., 2017). Murine TCR $\gamma\delta^+$  cells seed the intestine early in life, irrespective of microbial colonization or exposure to dietary antigen (Di Marco Barros et al., 2016), and persist *in situ* as naturally-occurring IELs (Cheroutre et al., 2011). Moreover, the peripheral and intraepithelial TCR $\gamma\delta^+$  compartments are largely non-overlapping as a consequence of distinct migratory characteristics, especially a lack of recirculating IELs (Chennupati et al., 2010; Sugahara et al., 1999).

Celiac disease (CeD) is a gastrointestinal inflammatory disorder triggered and maintained by dietary exposure to gluten (Jabri and Sollid, 2009). Antigen exposure can therefore be controlled *in vivo*, providing a unique opportunity to study the dynamics of human tissue-resident T cells in the setting of chronic inflammation. Active disease is characterized histologically by villous atrophy and immunologically by expanded populations of IELs (Jabri and Sollid, 2009). Adherence to a gluten-free diet (GFD) leads to resolution of the villous abnormalities, together with a decrease in the frequencies of gluten-specific TCR $\alpha\beta^+$  CD4 $^+$  T cells in the lamina propria and cytolytic TCR $\alpha\beta^+$  CD8 $^+$  IELs, which are consequently implicated in the pathogenesis of CeD (Jabri and Sollid, 2009). In contrast, TCR $\gamma\delta^+$  IEL expansions generally persist *in situ*, despite a lack of exposure to gluten (Kutlu et al., 1993). For this reason, tissue-resident TCR $\gamma\delta^+$  IELs are thought to regulate disease activity (Hayday, 2000), potentially by suppressing the influx of circulating T cells and/or by maintaining tissue integrity (Hayday et al., 2001), rather than participate actively in the pathogenesis of CeD (Kutlu et al., 1993). However, these

propositions remain speculative, pending a detailed functional evaluation of TCR $\gamma\delta^+$  IELs in patients with CeD. We set out to address this knowledge gap and more fundamentally to determine the effects of chronic inflammation on human tissue-resident TCR $\gamma\delta^+$  IELs.

## Results

### **V $\delta$ 1<sup>+</sup> T cells displaying hallmarks of tissue-resident lymphocytes are permanently expanded in CeD**

In line with previous studies (Halstensen et al., 1989; Kutlu et al., 1993), we found higher frequencies (Figure 1A) and absolute numbers (Figure 1B) of V $\delta$ 1<sup>+</sup> IELs in patients with CeD relative to healthy controls (Table S1), irrespective of adherence to a GFD (Figure 1C). Moreover, V $\delta$ 1<sup>+</sup> T cells constituted a significantly higher fraction of all TCR $\gamma\delta^+$  IELs in patients with active CeD (Figure 1D). A majority of V $\delta$ 1<sup>+</sup> IELs in healthy controls and patients with CeD expressed markers of tissue residency, namely CD69 and CD103 (Mueller and Mackay, 2016) (Figure 1E). V $\delta$ 1<sup>+</sup> IELs also expressed low levels of CD45RA and CCR7, indicative of an effector memory (T<sub>EM</sub>) phenotype (Sallusto et al., 1999), whereas the corresponding V $\delta$ 1<sup>+</sup> peripheral blood lymphocytes (PBLs) comprised a mixture of naïve and terminally differentiated effector (T<sub>EMRA</sub>) cells (Figure 1F). Expanded populations of V $\delta$ 1<sup>+</sup> IELs were therefore integrated as *bone fide* tissue-resident lymphocytes in patients with CeD.

### **Innate-like V $\delta$ 1<sup>+</sup> IELs are lost in CeD**

TCR $\alpha\beta^+$  CD8 $\alpha\beta^+$  IELs in patients with CeD typically express increased levels of NKG2D and activating CD94/NKG2A<sup>-</sup> NK receptors (Jabri and Sollid, 2009). We found no evidence of a similar phenotype among V $\delta$ 1<sup>+</sup> IELs from patients with active CeD (Figure

S1A and S1B). However, a vast majority of control V $\delta$ 1<sup>+</sup> IELs expressed the activating natural cytotoxicity receptors (NCRs) NKp46 and/or NKp44 (Figure 2A), irrespective of age (Figure 2B) and *in situ* expansion (Figure S1C). These observations suggested a definitive tissue-resident phenotype, reinforced by a lack of NCR expression on the surface of V $\delta$ 1<sup>+</sup> PBLs (Figure S1D). In contrast, V $\delta$ 1<sup>+</sup> IELs from patients with active or GFD-treated CeD rarely expressed NKp46 and almost exclusively lacked NKp44 (Fig. 2A). This disease-associated loss of NCR<sup>+</sup> V $\delta$ 1<sup>+</sup> IELs was refractory to long-term treatment with a GFD, unlike the concomitant loss of NCR<sup>+</sup> CD3<sup>-</sup> IELs, indicating a cell type-restricted effect distinct from the generic microenvironmental perturbations induced by CeD (Figure 2A and Figure S1E). Moreover, V $\delta$ 1<sup>+</sup> IELs expressing NKp46 and NKp44 were absent from the duodenum, the site of tissue destruction, but present in the colon of patients with GFD-treated CeD (Figure 2C). These results suggested that chronic inflammation precipitated an irretrievable and active site-specific loss of NK-like V $\delta$ 1<sup>+</sup> IELs in patients with CeD.

To determine the functional relevance of these NCRs, we measured cellular degranulation *ex vivo* and in response to stimulation with IL-15, which is upregulated under inflammatory conditions and is known to promote NK-receptor-mediated cytolytic activity (Jabri and Abadie, 2015). Co-engagement of NKp46 and NKp44 in conjunction with TCR ligation significantly increased granule exocytosis among control V $\delta$ 1<sup>+</sup> IELs relative to TCR ligation alone (Figure 2D), but only after pre-stimulation with IL-15 (Figure 2D and Figure S1F). V $\delta$ 1<sup>+</sup> IELs from healthy controls also expressed high levels of the cytolytic molecule granzyme B (Figure S1G) and degranulated at significantly higher frequencies than V $\delta$ 1<sup>+</sup> IELs from patients with CeD (Figure 2E). These findings revealed that the healthy small intestine harbored a unique set of innate-like and potentially cytolytic V $\delta$ 1<sup>+</sup> IELs that were displaced in the setting of CeD.

### **Dietary gluten drives the emergence of interferon- $\gamma$ -producing V $\delta$ 1<sup>+</sup> IELs in CeD**

To examine the functional properties of V $\delta$ 1<sup>+</sup> IELs in more depth, we extended our analysis to cytokine production directly *ex vivo*. Control V $\delta$ 1<sup>+</sup> IELs produced very little interferon (IFN)- $\gamma$  or tumor necrosis factor (TNF)- $\alpha$  on a percent cell basis (Figure 3A). In contrast, approximately 50% of V $\delta$ 1<sup>+</sup> IELs from patients with active CeD produced IFN- $\gamma$ , but not TNF- $\alpha$  (Figure 3A). Significantly lower frequencies of IFN- $\gamma$ <sup>+</sup> V $\delta$ 1<sup>+</sup> IELs were detected in patients with GFD-treated CeD (Figure 3A). Irrespective of disease status, V $\delta$ 1<sup>+</sup> PBLs produced both IFN- $\gamma$  and TNF- $\alpha$ , and neither TCR $\alpha\beta$ <sup>+</sup> CD8 $\alpha\beta$ <sup>+</sup> nor TCR $\alpha\beta$ <sup>+</sup> CD4<sup>+</sup> IELs displayed enhanced production of IFN- $\gamma$  in patients with active CeD (Figure 3A). A significant increase in IFN- $\gamma$  production was also detected among V $\delta$ 1<sup>+</sup> IELs isolated from patients with GFD-treated CeD after gluten challenge relative to donor-matched V $\delta$ 1<sup>+</sup> IELs isolated prior to gluten challenge, whereas no such chronological differences were observed in the corresponding TCR $\alpha\beta$ <sup>+</sup> CD8 $\alpha\beta$ <sup>+</sup> or TCR $\alpha\beta$ <sup>+</sup> CD4<sup>+</sup> IEL compartments (Figure 3B). These data suggested a gluten-dependent and cell type-restricted gain of IFN- $\gamma$ -producing function among V $\delta$ 1<sup>+</sup> IELs in patients with active CeD.

### **The transcriptional program of V $\delta$ 1<sup>+</sup> IELs is permanently altered in CeD**

In further analyses, we performed *ex vivo* RNA sequencing (RNA-seq) to determine if naturally-occurring V $\delta$ 1<sup>+</sup> IELs, represented by the prevalent NKp46<sup>+</sup> (NCR<sup>+</sup>) subset in healthy controls, were fundamentally distinct from disease-associated V $\delta$ 1<sup>+</sup> IELs, represented by the prevalent NKp46<sup>-</sup> (NCR<sup>-</sup>) subset in patients with CeD. A total of 645 genes exhibited differential expression between control NCR<sup>+</sup> V $\delta$ 1<sup>+</sup> IELs and NCR<sup>-</sup> V $\delta$ 1<sup>+</sup> IELs from patients with active CeD (Figure S2A). Similar differences were



observed in comparisons between control NCR<sup>+</sup> V $\delta$ 1<sup>+</sup> IELs and NCR<sup>-</sup> V $\delta$ 1<sup>+</sup> IELs from patients with GFD-treated CeD (Figure S2B). Moreover, this overarching dichotomy was confirmed using minimum spanning tree (MST) analysis (Xu et al., 2002), which showed that control NCR<sup>+</sup> V $\delta$ 1<sup>+</sup> IELs formed distinct clusters relative to NCR<sup>-</sup> V $\delta$ 1<sup>+</sup> IELs from patients with active or GFD-treated CeD (Figure 4A).

Targeted analysis of genes encoding archetypal NK receptors and cytolytic effector molecules (Table S2) (Meresse et al., 2006) confirmed the innate-like nature and cytolytic potential of control V $\delta$ 1<sup>+</sup> IELs (Figure 4B). In particular, *GZMK*, *FCGR3A*, and *TYROBP* were significantly overexpressed among control NCR<sup>+</sup> V $\delta$ 1<sup>+</sup> IELs (Figure 4B). A similar analysis of genes encoding various cytokines, chemokines, and growth factors (Table S2) revealed that *IFNG*, *CCL4*, and *IL10* were significantly overexpressed among NCR<sup>-</sup> V $\delta$ 1<sup>+</sup> IELs from patients with CeD (Figure 4C). Transcripts encoding IL-4, IL-9, IL-13, and IL-17 were not detected. Moreover, control NCR<sup>+</sup> V $\delta$ 1<sup>+</sup> IELs expressed significantly more *AREG*, which encodes a growth factor implicated in tissue repair (Zaiss et al., 2015), relative to NCR<sup>-</sup> V $\delta$ 1<sup>+</sup> IELs from patients with active CeD (Figure 4C). Analogous patterns were observed for other genes associated with tissue repair (Linehan et al., 2018) (Figure 4D and Table S2). Clear differences also emerged with respect to the expression of transcription factors associated with immune function (Table S2). For example, control NCR<sup>+</sup> V $\delta$ 1<sup>+</sup> IELs specifically expressed *GATA3* and *IRF8*, which have been implicated in the regulation of innate immunity (Adams et al., 2018; Zhu, 2017), whereas *IRF1* (Kano et al., 2008) and *RUNX1* (Wang et al., 2014) were selectively overexpressed among NCR<sup>-</sup> V $\delta$ 1<sup>+</sup> IELs from patients with active CeD (Figure 4E), consistent with ability of these cells to produce IFN- $\gamma$  (Figure 3A).

To assess the relative impact of origin versus phenotype, we also compared NCR<sup>+</sup> V $\delta$ 1<sup>+</sup> IELs from healthy controls with NCR<sup>+</sup> V $\delta$ 1<sup>+</sup> IELs from patients with active CeD. The

latter were occasionally found in younger patients (Figure S2C), but lacked expression of NKp44, in contrast to control NCR<sup>+</sup> V $\delta$ 1<sup>+</sup> IELs. Unbiased multidimensional scaling (MDS) analysis revealed that control NCR<sup>+</sup> V $\delta$ 1<sup>+</sup> IELs formed a distinct subset, whereas NCR<sup>+</sup> and NCR<sup>-</sup> V $\delta$ 1<sup>+</sup> IELs from patients with active CeD clustered together (Figure S2D). Accordingly, differences observed between control NCR<sup>+</sup> V $\delta$ 1<sup>+</sup> IELs and NCR<sup>-</sup> V $\delta$ 1<sup>+</sup> IELs from patients with active CeD (Figure S2B) were mimicked between control NCR<sup>+</sup> V $\delta$ 1<sup>+</sup> IELs and NCR<sup>+</sup> V $\delta$ 1<sup>+</sup> IELs from patients with active CeD (Figure S2E). Transcriptional differences among V $\delta$ 1<sup>+</sup> IELs were therefore driven by disease state rather than NCR expression, irrespective of adherence to a GFD.

In line with the functional and phenotypic data, these results demonstrated that V $\delta$ 1<sup>+</sup> IELs formed distinct subsets characterized by divergent gene expression programs in healthy controls and patients with CeD.

### **The V $\delta$ 1<sup>+</sup> IEL TCR repertoire is permanently reshaped in CeD**

To determine if the switch from innate-like NCR<sup>+</sup> V $\delta$ 1<sup>+</sup> IELs in the healthy state to IFN- $\gamma$ -producing NCR<sup>-</sup> V $\delta$ 1<sup>+</sup> IELs in CeD was associated with cellular turnover, we used an unbiased molecular approach to characterize all expressed *TRG* and *TRD* gene rearrangements in flow-sorted V $\delta$ 1<sup>+</sup> T cell populations isolated directly *ex vivo* from donor-matched PBLs and IELs (Figure S3A–C, Table S3, and Table S4) (Davey et al., 2017; Quigley et al., 2011). Control V $\delta$ 1<sup>+</sup> IELs almost exclusively used the *TRGV4* gene, unlike donor-matched V $\delta$ 1<sup>+</sup> PBLs (Figure 5A, Figure 5B, and Table S5A). No such preference was observed in patients with active or GFD-treated CeD (Figure 5A, Figure 5B, and Table S5A). Accordingly, *TRGV4* gene transcripts were significantly enriched among control V $\delta$ 1<sup>+</sup> IELs relative to V $\delta$ 1<sup>+</sup> IELs from patients with active or GFD-treated CeD (Figure 5A, Figure 5B, and Table S5A). These results were corroborated in a separate cohort via RNA-

seq analysis, which also showed that *TRGV* gene use was not related to NCR expression (Figure S3D). Moreover, control  $V\delta 1^-$  IELs displayed a similarly extreme preference for the *TRGV4* gene and expressed NCRs at frequencies equivalent to those observed among control  $V\delta 1^+$  IELs (Figure 5C and Figure S3E). In contrast,  $V\delta 1^-$  IELs from patients with CeD displayed no obvious preference for a particular *TRGV* gene and lacked expression of NCRs, akin to  $V\delta 1^+$  IELs from patients with CeD (Figure 5C and Figure S3E). This disease-associated loss of *TRGV4* gene transcripts occurred independently of human leukocyte antigen (*HLA*) class II alleles linked with CeD (Table S3). Of note no *TRGJ* gene bias was observed among groups or tissues, reflecting widespread use of the *TRGJ1* gene (Figure S3F and Table S5B). Collectively, these data suggested that *TRGV4* gene rearrangements facilitated the selection of naturally-occurring  $TCR\gamma\delta^+$  IELs, which were supplanted in patients with CeD.

### **A molecular signature defines $V\delta 1^+$ IEL expansions in active CeD**

Next, we sought evidence of antigen-driven clonal expansions within the remodeled  $V\delta 1^+$  IEL TCR repertoires in patients with CeD. A subset of patients with active CeD harbored low-diversity repertoires, which were not apparent in patients with GFD-treated CeD (Figure S4A). These results suggested that gluten consumption stimulated the expansion of particular clonotypes in the  $V\delta 1^+$  IEL pool, at least in some patients with active CeD, whereas gluten withdrawal allowed diversification, potentially via a loss of antigenic drive and/or *de novo* recruitment of non-expanded  $V\delta 1^+$  IELs.

To address this possibility, we tested for amino acid (AA) preferences among unique CDR3 $\delta$  sequences, with the aim of identifying  $V\delta 1^+$  TCR motifs associated with active CeD. No group-specific AA enrichments were detected to indicate a disease-associated molecular signature (Figure S4B and Table S5C). Of note, *TRDJ1* gene

transcripts predominated across study groups (Figure S4C). At the genetic level, preferential use of *TRDD3* and, to a lesser extent, *TRDD2* in the forward frame was observed across the CDR3 $\delta$  data set as a whole (Figure S4D), but there was no evidence for preferential use of a specific *TRDD* gene among V $\delta$ 1<sup>+</sup> IELs from patients with active CeD (Figure S4E). These data suggested that gluten-induced reshaping of the V $\delta$ 1<sup>+</sup> IEL repertoire was not associated with a clear CDR3 $\delta$  motif. In line with this interpretation, two previously described CDR3 $\delta$  motifs associated with gluten challenge, CxxxxxPxLGD (PxLGD) and CxxxxxxxxYWGI (YWGI) (Han et al., 2013), were distributed at low frequencies in the corresponding V $\delta$ 1<sup>+</sup> repertoires obtained from patients with active or GFD-treated CeD (Figure S4E). However, the fraction of V $\delta$ 1<sup>+</sup> IEL-derived CDR3 $\delta$  sequences that incorporated no *TRDD* gene-encoded AAs was significantly higher in healthy controls relative to patients with CeD (Figure S4F and Table S5D), and the corresponding CDR3 $\delta$  sequences were significantly shorter in healthy controls relative to patients with CeD (Figure S4G).

In contrast, a similar analysis of CDR3 $\gamma$  sequences revealed that histidine (H) was significantly enriched in the V $\delta$ 1<sup>+</sup> IEL repertoires obtained from patients with active CeD relative to the corresponding repertoires obtained from healthy controls and patients with GFD-treated CeD (Figure 6A and Table S5E). This observation was supported by iceLogo motif analysis (Colaert et al., 2009), which revealed a significant enrichment for H adjacent to the J $\gamma$  segment among CDR3 $\gamma$  sequences obtained from patients with active CeD (Figure S4H). In addition, H was found adjacent to the *TRGJ1*-encoded J $\gamma$  segment (H-J1 motif) in four public CDR3 $\gamma$  sequences, collectively shared across six patients with active CeD (Figure 6B and Table S6). It is notable that these H-J1 CDR3 $\gamma$  sequences associated with various *TRGV* gene segments (Figure 6B and Table S6). This pattern of mosaicism hinted at a unique selection pressure focused on the somatically rearranged CDR3 $\gamma$  loop

in patients with CeD. A significantly higher fraction of unique CDR3 $\gamma$  sequences carried the H-J1 motif in the V $\delta$ 1<sup>+</sup> IEL repertoires obtained from patients with active CeD relative to the corresponding repertoires obtained from healthy controls and patients with GFD-treated CeD (Figure 6C). More specifically, H-J1<sup>+</sup> CDR3 $\gamma$  sequences were present in seven patients with active CeD versus only two patients with GFD-treated CeD, irrespective of *HLA-DQ2/DQ8* genotype (Figure S4I and Table S3). The most common V $\delta$ 1<sup>+</sup> IEL-derived CDR3 $\gamma$  sequence detected in four patients with active CeD incorporated the H-J1 motif (Figure 6D and Figure 6E). In contrast, the associated V $\delta$ 1<sup>-</sup> IEL repertoires in these patients did not exhibit H-J1<sup>+</sup> expansions, highlighting the lineage specificity of this molecular signature (Figure 6F). The most common V $\delta$ 1<sup>+</sup> PBL-derived CDR3 $\gamma$  sequence detected in three patients with active CeD also incorporated the H-J1 motif (Figure 6D and Figure 6E). Accordingly, V $\delta$ 1<sup>+</sup> TCR repertoires incorporating a dominant H-J1<sup>+</sup> CDR3 $\gamma$  sequence were significantly enriched among patients with active CeD relative to healthy controls and patients with GFD-treated CeD, compared across PBLs and IELs (Figure 6E).

Collectively, these observations indicated that gluten-induced reshaping of the V $\delta$ 1<sup>+</sup> IEL compartment in patients with active CeD was associated with preferential recruitment and/or expansion of clonotypes bearing the H-J1<sup>+</sup> CDR3 $\gamma$  motif, which in turn suggested that the underlying chronic inflammatory process incorporated a degree of TCR-mediated specificity for a putative ligand associated with CeD.

### **V $\delta$ 1<sup>+</sup> IELs display hallmarks of TCR-mediated activation in patients with CeD**

The combined observations that V $\delta$ 1<sup>+</sup> IELs acquired the ability to produce IFN- $\gamma$  in response to gluten challenge (Figure 3B) and preferentially incorporated clonotypes bearing a shared H-J1<sup>+</sup> CDR3 $\gamma$  motif in patients with active but not GFD-treated CeD

(Figure 6D and Figure 6E) suggested the possibility of *in vivo* activation via the TCR. Analysis of a set of genes associated with T cell activation and TCR signaling (Table S2) (Fabregat et al., 2018) showed that NCR<sup>-</sup> V $\delta$ 1<sup>+</sup> IELs from patients with active CeD significantly overexpressed transcripts associated with antigenic stimulation relative to control NCR<sup>+</sup> V $\delta$ 1<sup>+</sup> IELs (Figure 6G). More specifically, *MKI67*, *CTLA4*, *PDCD1*, and a subset of genes encoding HLA class II molecules were significantly overexpressed among NCR<sup>-</sup> V $\delta$ 1<sup>+</sup> IELs from patients with active CeD relative to NCR<sup>+</sup> V $\delta$ 1<sup>+</sup> IELs from healthy controls and NCR<sup>-</sup> V $\delta$ 1<sup>+</sup> IELs from patients with GFD-treated CeD (Figure 6G). Gluten withdrawal was therefore associated with curtailed activation and a corresponding lack of TCR-driven selection among NCR<sup>-</sup> V $\delta$ 1<sup>+</sup> IELs (Figure 6C and Figure S4A). Direct *ex vivo* analysis of the transcription factor Nur77, which indicates signaling via the TCR (Ashouri and Weiss, 2017), further supported a direct link between gluten exposure and the activation status of NCR<sup>-</sup> V $\delta$ 1<sup>+</sup> IELs (Figure 6H). Accordingly, these data provided evidence for gluten-driven immune responses mediated via the engagement of specific V $\delta$ 1<sup>+</sup> IEL TCRs in patients with active CeD.

### **BTNL3/8-reactive V $\delta$ 1<sup>+</sup> IELs are permanently lost in CeD**

Recent studies have shown that butyrophilin (BTN) and butyrophilin-like (BTNL) molecules play a key role in  $\gamma\delta$  T cell biology (Di Marco Barros et al., 2016; Melandri et al., 2018; Vantourout et al., 2018). In line with the notion that human colonic V $\gamma$ 4<sup>+</sup> IELs may be selected under physiological conditions by BTNL molecules (Di Marco Barros et al., 2016), the loss of *TRGV4* gene transcripts among V $\delta$ 1<sup>+</sup> IELs from patients with CeD (Figure 5A) was associated with corresponding decreases in *BTNL8* gene transcript (Figure 7A) and protein expression levels (Figure 7B). Moreover, gluten withdrawal failed to restore *TRGV4* gene use among V $\delta$ 1<sup>+</sup> IELs, despite normalization of *BTNL8* expression

(Figure 7A, Figure 7B, and Figure S5A). In addition,  $V\delta 1^+$  IELs from patients with active or GFD-treated CeD were not activated in the presence of BTNL3/8<sup>+</sup> HEK293T cells (Figure 7C and Figure S5B), but responded efficiently to generic stimulation via cross-linking of TCRs (Figure S5C). Similar results were obtained with  $V\delta 1^-$  IELs from patients with CeD, which lacked expression of *TRGV4*, but not with control  $V\delta 1^+$  and  $V\delta 1^-$  IELs, which displayed an extreme preference for *TRGV4* and responded in the presence of BTNL3/8<sup>+</sup> HEK293T cells (Figure 5C and Figure 7D). Importantly, BTNL3/8-reactive  $V\delta 1^+$  IELs were lost from the duodenum, but not the colon, in a patient with GFD-treated CeD (Figure S5D). This finding concurred with the site-specific depletion of NCR<sup>+</sup>  $V\delta 1^+$  IELs (Figure 2C). Collectively, these results suggested that the recovery of BTNL molecules after exclusion of dietary gluten was insufficient to reconstitute the niche favored by naturally-occurring  $V\gamma 4^+/V\delta 1^+$  IELs, potentially reflecting an absolute homeostatic requirement for sustained expression of BTNL3/8.

In some cases, patients with active CeD harbored populations of  $V\gamma 4^+/V\delta 1^+$  IELs, which tended to express TCRs that were enriched for the H-J1 motif (Figure 7E). As  $V\gamma 4^+$  T cells from healthy human colon have been shown to react with BTNL3/8 (Di Marco Barros et al., 2016; Melandri et al., 2018), we formally tested the ability of  $V\gamma 4^+/V\delta 1^+$  IEL H-J1<sup>+</sup> TCRs from patients with active CeD to recognize BTNL3/8. Two  $V\gamma 4^+/V\delta 1^+$  IEL TCRs from healthy controls triggered dose-dependent responses to BTNL3/8, as expected, whereas a  $V\gamma 3^+/V\delta 1^+$  IEL TCR from a patient with active CeD failed to recognize BTNL3/8 (Figure 7F, Figure S5E, and Figure S5F). Strikingly, a  $V\gamma 4^+/V\delta 1^+$  IEL H-J1<sup>+</sup> TCR from a patient with active CeD also failed to recognize BTNL3/8 (Figure 7F, Figure S5E, and Figure S5F). This particular TCR was characterized by the presence of a long CDR3 $\delta$  loop (22 AAs), whereas control  $V\gamma 4^+/V\delta 1^+$  IEL TCRs typically incorporated shorter CDR3 $\delta$  loops (Figure S4F). Moreover, the control  $V\gamma 4^+/V\delta 1^+$  IEL TCR with the shortest CDR3 $\delta$

loop (12 AAs, TCR 95) displayed the strongest reactivity against BTNL3/8 (Figure 7F). These observations suggested that the inability of the  $V\gamma 4^+/V\delta 1^+$  IEL H-J1<sup>+</sup> TCR from the patient with active CeD to recognize BTNL3/8 was related to the presence of a long CDR3 $\delta$  loop, consistent with a recent study in which a  $V\gamma 4^+/V\delta 1^+$  IEL H-J1<sup>+</sup> TCR with a short CDR3 $\delta$  loop recognized BTNL3/8 (Melandri et al., 2018). It remains to be determined if the HJ-1 motif can impede  $V\gamma 4^+/V\delta 1^+$  TCR interactions with BTNL3/8. Of note, all expressed TCRs transduced a functional signal in the presence of  $\alpha$ CD3/ $\alpha$ CD28 beads (Figure S5G and S5H). Collectively, these results provided direct evidence to support the contention that gluten-induced inflammatory remodeling of the TCR $\gamma\delta^+$  IEL compartment in patients with CeD was associated with a loss of productive TCR-mediated interactions with BTNL3/8.

### **Dynamic remodeling of the $V\delta 1^+$ IEL compartment precedes tissue damage in CeD**

The  $V\delta 1^+$  IEL compartment underwent dynamic changes both during the chronic inflammatory process associated with active CeD and during the resolution phase associated with strict adherence to a GFD (Figure S6). However, it was still unclear whether these alterations were causally or reactively linked with tissue damage. To address this issue, we analyzed patients with potential CeD, defined as a state of CD4<sup>+</sup> T cell-mediated intolerance to dietary gluten without histological evidence of villous atrophy (Husby et al., 2012). This heterogeneous group encompassed patients with various levels of  $V\delta 1^+$  IEL infiltration (Figure S7A), loss of *BTNL8* gene transcripts (Figure S7B), loss of NCR<sup>+</sup>  $V\delta 1^+$  IELs (Figure S7C), and loss of *TRGV4* gene transcripts (Figure S7D). This heterogeneity was further exemplified by two patients with potential CeD, one of which exhibited an expansion of NCR<sup>+</sup>  $V\delta 1^+$  IELs with conserved *ex vivo* reactivity against BTNL3/8, and the other of which harbored physiological numbers of  $V\delta 1^+$  IELs that lacked



NCR expression and *ex vivo* reactivity against BTNL3/8 (Figure 7G). In conjunction with the presence of occasional V $\delta$ 1<sup>+</sup> IEL TCRs bearing the H-J1<sup>+</sup> CDR3 $\gamma$  motif in patients with potential CeD (Figure S7E), these observations suggested that the loss of naturally-occurring innate-like V $\delta$ 1<sup>+</sup> IELs may precede tissue destruction and the emergence of adaptive features within the disease-associated V $\delta$ 1<sup>+</sup> IEL repertoire.

## Discussion

Current paradigms stipulate that tissue-resident immunity is established and maintained by long-lived, anatomically compartmentalized populations of lymphocytes, which adapt via local homeostasis and proliferation to novel antigenic challenges without displacing pre-existing memory specificities (Beura et al., 2018; Mueller and Mackay, 2016; Park et al., 2018). However, the impact of chronic inflammation on the tissue-resident lymphocyte pool has not been defined in previous studies, in part due to a lack of suitable models. We addressed this knowledge gap by taking advantage of the fact that antigen exposure can be controlled in CeD, a complex T cell-mediated inflammatory disorder with an autoimmune component (Jabri and Sollid, 2009).

In healthy controls, we found that the tissue-resident TCR $\gamma\delta$ <sup>+</sup> IEL compartment was dominated by a unique subset of innate-like, semi-invariant V $\gamma$ 4<sup>+</sup>/V $\delta$ 1<sup>+</sup> and V $\gamma$ 4<sup>+</sup>/V $\delta$ 1<sup>-</sup> IELs that recognized BTNL3/BTNL8, constitutively expressed NCRs, and persisted throughout life, exemplifying the longevity and stability required to maintain durable immunity at barrier sites under physiological conditions. Importantly, these V $\gamma$ 4<sup>+</sup>/V $\delta$ 1<sup>+</sup> IELs were ideally poised to maintain homeostasis in the local microenvironment, either by eliminating virus-infected or malignant cells in response to innate signals, such as NCR ligands and IL-15, or by promoting tissue healing via the production of growth factors, such as amphiregulin (Zaiss et al., 2015).

In patients with CeD, the tissue-resident  $V\delta 1^+$  IEL compartment was profoundly altered, even after exclusion of the inciting antigen and resolution of the associated inflammation. The key changes included an irretrievable loss of innate-like  $V\gamma 4^+/V\delta 1^+$  IELs and the emergence of gluten-sensitive IFN- $\gamma$ -producing  $V\delta 1^+$  IELs characterized by clonal expansions incorporating H-J1<sup>+</sup> TCRs that lacked reactivity against BTN3A1/BTN3B. These observations challenge the assumption that pre-existing tissue-resident  $V\delta 1^+$  IELs expand in CeD (Hayday et al., 2001). We therefore propose a multistep model, whereby naturally-occurring  $V\gamma 4^+/V\delta 1^+$  IELs expand in response to inflammation, triggered by the loss of tolerance to dietary gluten, allowing initial preservation of NCR expression and BTN3A1/BTN3B reactivity, as observed in some patients with potential CeD. The ongoing disease process then reaches a tipping point, as observed in a subset of patients with potential or active CeD, where the chronic loss of BTN3A1 expression eventually leads to the loss of  $V\gamma 4^+/V\delta 1^+$  IELs, which likely depend on this ligand for survival. Circulating TCR $\gamma\delta^+$  cells are then recruited into the vacant immunological space and subsequently acquire a tissue-resident phenotype in response to local signals. In line with this proposition,  $V\delta 1^+$  PBLs from several patients with active CeD expressed TCRs incorporating the H-J1<sup>+</sup> CDR3 $\gamma$  motif, indicative of priming by a locally induced ligand in gut-associated lymphoid tissue (Guy-Grand et al., 2013). Moreover, gluten withdrawal precipitated a contraction of H-J1<sup>+</sup> TCRs among  $V\delta 1^+$  IELs, consistent with the *in situ* expression of an antigen-dependent ligand in patients with active CeD. It is important to note that NCR<sup>+</sup>  $V\delta 1^+$  IELs were occasionally found in patients with active CeD. However, these cells aligned transcriptionally with disease-associated NCR<sup>-</sup>  $V\delta 1^+$  IELs rather than control NCR<sup>+</sup>  $V\delta 1^+$  IELs, lacked expression of NKp44 and the *TRGV4* gene, and were only found in children with CeD. Accordingly, NCR expression is likely regulated by site-specific

signals, which are retained to some extent in children with active CeD, selectively facilitating the expression of NKp46 on V $\delta$ 1<sup>+</sup> IELs.

A previous study in mice showed that V $\gamma$ 7<sup>+</sup> IELs (the mouse homolog of human V $\gamma$ 4<sup>+</sup> IELs) could not be rescued if the expression of BTNL molecules was delayed beyond the neonatal period, during which these cells typically expand *in situ* (Di Marco Barros et al., 2016). In conjunction with our finding that innate-like V $\gamma$ 4<sup>+</sup>/V $\delta$ 1<sup>+</sup> IELs failed to recover in patients with GFD-treated CeD, despite restoration of the mucosal architecture and BTNL8 expression, this observation suggests that tissue-resident TCR $\gamma\delta$ <sup>+</sup> IELs are exquisitely sensitive to the presence of BTNL molecules in the intestinal epithelium. It therefore seems likely that the profound changes we observed in the tissue-resident compartment of patients with CeD were underpinned by a loss of BTNL molecules in the local microenvironment, although it remains to be determined how long this state of depletion needs to persist to trigger the irretrievable destruction of a fully established niche constituted by mature populations of V $\gamma$ 4<sup>+</sup>/V $\delta$ 1<sup>+</sup> IELs.

It has been proposed that TCR $\gamma\delta$ <sup>+</sup> IELs act to limit the infiltration of systemic T cells in CeD (Hayday et al., 2001), without playing a direct role in disease pathogenesis (Hayday, 2000; Kutlu et al., 1993). Our data challenge this assumption. In particular, we found that V $\delta$ 1<sup>+</sup> IELs in patients with CeD expressed high levels of the chemoattractant *CCL4* and adopted a Th1-like phenotype, characterized by expression of *IRF1* and *RUNX1* and the production of IFN- $\gamma$ . Moreover, only V $\delta$ 1<sup>+</sup> IELs showed enhanced IFN- $\gamma$  in response to gluten challenge, which is highly relevant in the context of a Th1-mediated disorder like CeD. Previous studies have shown that IFN- $\gamma$  induces the upregulation of HLA-E on intestinal epithelial cells (Meresse et al., 2006) and further upregulates major histocompatibility complex (MHC) class I molecules (Früh and Yang, 1999), which in turn can contribute to the activation of TCR $\alpha\beta$ <sup>+</sup> CD8 $\alpha\beta$ <sup>+</sup> IELs, the primary mediators of

intestinal epithelial cell destruction in CeD (Jabri and Sollid, 2009). On the flip side of this argument,  $V\delta 1^+$  IELs also expressed genes associated with the regulation of chronic inflammatory responses, such as *IL-10*, *CTLA4*, *PDCD1*, and *ZNF683*, in patients with CeD. However, the corresponding molecules can equally act as proxies of chronic inflammation, in line with a pathogenic role for IFN- $\gamma$ -producing  $V\delta 1^+$  IELs. For example, Th1 cells self-regulate via the production of IL-10 (O'Garra and Vieira, 2007; Saraiva et al., 2009), and gluten-specific CD4<sup>+</sup> T cells can express both IFN- $\gamma$  and IL-10 (Nilsen et al., 1995).

In conclusion, we have shown that chronic site-specific inflammation permanently reconfigures the tissue-resident TCR $\gamma\delta^+$  IEL compartment in patients with CeD. A similar process of 'immunological scarring' may contribute to the pathogenesis of other intestinal immune disorders, such as ulcerative colitis, which is also characterized by decreased levels of *BTNL8 in situ*. Further studies are therefore required to establish the general applicability of our findings across disease states and to determine the impact of chronic inflammatory processes on the stability of adaptive immune cell populations. In anticipation of these complementary data, we speculate that the irretrievable loss of tissue-resident subsets with unique innate-like specificities and cytolytic properties may have long-term implications for the health of patients with CeD.

### **Author Contributions**

T.M. and B.J. conceived the study; T.M., B.J., and D.A.P. designed experiments; T.M., K.L., J.E.M., D.G.S., M.T., J.J.R., C.C., H.E.S., and H.H.R. performed experiments; T.M., K.L., H.G., J.E.M., and V.vU. analyzed and interpreted data; K.L. and J.E.M. performed primary analysis of TCR repertoires; T.M. and H.G. performed computational analysis of TCR repertoires; T.M. and J.C.G. performed RNA-seq analysis; D.A.P. supervised experimental acquisition and primary analysis of TCR repertoires; A.R.D. supervised

computational analysis of TCR repertoires; L.B.B. supervised analysis of RNA-seq data; V.K. and C.W. supervised HLA genotyping experiments; I.L., M.D., P.H.R.G, R.G., H.J., C.E.S., S.G., and S.S.K. contributed samples; T.M. and B.J. drafted the manuscript; D.A.P. critically edited the manuscript; K.L., H.G., J.E.M., D.G.S., A.R.D., and J.R. critically read the manuscript; J.R., D.A.P., and B.J. acquired funds to support the work; B.J. directed the study.

### **Acknowledgements**

We thank the patients and their families for making this study possible, Robert Kavitt, Edwin McDonald, Atsushi Sakuraba, Joel Pekow, Neil Sengupta, and Sushila Dalal for patient care, Kathryn Lesko, Diane McKiernan, Sarbani Adhikari, and Fengshi Dong for patient recruitment, Mathieu Platteel for assistance with HLA genotyping, Steven Erickson and Delphine Guy-Grand for editorial input and thoughtful discussions, and staff at the University of Chicago Flow Cytometry Core Facility. This work was supported by the National Institutes of Health via T32 AI07090 to T.M., T32 GM007281 to D.G.S, R01 DK067180 to B.J., and the Digestive Diseases Research Core Center P30 DK42086 at the University of Chicago to B.J. J.R. is an Australian Research Council Laureate Fellow. D.A.P. is a Wellcome Trust Senior Investigator (100326Z/12/Z).

### **Declaration of Interests**

The authors declare no competing interests.

## References

Adams, N.M., Lau, C.M., Fan, X., Rapp, M., Geary, C.D., Weizman, O.-E., Diaz-Salazar, C., and Sun, J.C. (2018). Transcription factor IRF8 orchestrates the adaptive natural killer cell response. *Immunity* *48*, 1172–1182.e1176.

Ashouri, J.F., and Weiss, A. (2017). Endogenous Nur77 is a specific indicator of antigen receptor signaling in human T and B Cells. *J Immunol* *198*, 657–668.

Beura, L.K., Mitchell, J.S., Thompson, E.A., Schenkel, J.M., Mohammed, J., Wijeyesinghe, S., Fonseca, R., Burbach, B.J., Hickman, H.D., Vezys, V., et al. (2018). Intravital mucosal imaging of CD8<sup>+</sup> resident memory T cells shows tissue-autonomous recall responses that amplify secondary memory. *Nat Immunol* *19*, 173–182.

Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* *34*, 525–527.

Brochet, X., Lefranc, M.P., and Giudicelli, V. (2008). IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. *Nucleic Acids Res* *36*, W503–W508.

Chennupati, V., Worbs, T., Liu, X., Malinarich, F.H., Schmitz, S., Haas, J.D., Malissen, B., Förster, R., and Prinz, I. (2010). Intra- and intercompartmental movement of  $\gamma\delta$  T cells: intestinal intraepithelial and peripheral  $\gamma\delta$  T cells represent exclusive nonoverlapping populations with distinct migration characteristics. *J Immunol* *185*, 5160–5168.

Cheroutre, H., Lambolez, F., and Mucida, D. (2011). The light and dark sides of intestinal intraepithelial lymphocytes. *Nat Rev Immunol* *11*, 445–456.

Colaert, N., Helsen, K., Martens, L., Vandekerckhove, J., and Gevaert, K. (2009). Improved visualization of protein consensus sequences by iceLogo. *Nat Meth* *6*, 786–787.

Davey, M.S., Willcox, C.R., Joyce, S.P., Ladell, K., Kasatskaya, S.A., McLaren, J.E., Hunter, S., Salim, M., Mohammed, F., Price, D.A., et al. (2017). Clonal selection in the human V $\delta$ 1 T cell repertoire indicates  $\gamma\delta$  TCR-dependent adaptive immune surveillance. *Nat Commun* *8*, 14760.

Di Marco Barros, R., Roberts, N.A., Dart, R.J., Vantourout, P., Jandke, A., Nussbaumer, O., Deban, L., Cipolat, S., Hart, R., Iannitto, M.L., et al. (2016). Epithelia use butyrophilin-like molecules to shape organ-specific  $\gamma\delta$  T cell compartments. *Cell* *167*, 203–218.e217.

Fabregat, A., Jupe, S., Matthews, L., Sidiropoulos, K., Gillespie, M., Garapati, P., Haw, R., Jassal, B., Korninger, F., May, B., et al. (2018). The Reactome Pathway Knowledgebase. *Nucleic Acids Res* *46*, D649–D655.

Ferrari, S., and Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. *J Appl Stat* *31*, 799–815.

Früh, K., and Yang, Y. (1999). Antigen presentation by MHC class I and its regulation by interferon gamma. *Curr Opin Immunol* *11*, 76–81.

Gras, S., Chen, Z., Miles, J.J., Liu, Y.C., Bell, M.J., Sullivan, L.C., Kjer-Nielsen, L., Brennan, R.M., Burrows, J.M., Neller, M.A., et al. (2010). Allelic polymorphism in the T cell receptor and its impact on immune responses. *J Exp Med* *207*, 1555–1567.

Guy-Grand, D., Vassalli, P., Eberl, G., Pereira, P., Buren-Defranoux, O., Lemaitre, F., Di Santo, J.P., Freitas, A.A., Cumano, A., and Bandeira, A. (2013). Origin, trafficking, and intraepithelial fate of gut-tropic T cells. *J Exp Med* *210*, 1839–1854.

Halstensen, T.S., Scott, H., and Brandtzaeg, P. (1989). Intraepithelial T cells of the TcR  $\gamma/\delta^+$  CD8<sup>-</sup> and V $\delta$ 1/J $\delta$ 1<sup>+</sup> phenotypes are increased in coeliac disease. *Scand J Immunol* *30*, 665–672.

Han, A., Newell, E.W., Glanville, J., Fernandez-Becker, N., Khosla, C., Chien, Y.-H., and Davis, M.M. (2013). Dietary gluten triggers concomitant activation of CD4<sup>+</sup> and CD8<sup>+</sup>  $\alpha\beta$  T cells and  $\gamma\delta$  T cells in celiac disease. *Proc Natl Acad Sci U S A* *110*, 13073–13078.

Hayday, A.C. (2000).  $\gamma\delta$  cells: a right time and a right place for a conserved third way of protection. *Annu Rev Immunol* *18*, 975–1026.

Hayday, A., Theodoridis, E., Ramsburg, E., and Shires, J. (2001). Intraepithelial lymphocytes: exploring the third way in immunology. *Nat Immunol* *2*, 997–1003.

Holst, J., Szymczak-Workman, A.L., Vignali, K.M., Burton, A.R., Workman, C.J., and Vignali, D.A.A. (2006). Generation of T-cell receptor retrogenic mice. *Nature Protoc* *1*, 406–417.

Hoytema van Konijnenburg, D.P., Reis, B.S., Pedicord, V.A., Farache, J., Victora, G.D., and Mucida, D. (2017). Intestinal epithelial and intraepithelial T cell crosstalk mediates a dynamic response to infection. *Cell* *171*, 783–794.e13.

Husby, S., Koletzko, S., Korponay-Szabó, I.R., Mearin, M.L., Phillips, A., Shamir, R., Troncone, R., Giersiepen, K., Branski, D., Catassi, C., et al. (2012). European society for pediatric gastroenterology, hepatology, and nutrition guidelines for the diagnosis of coeliac disease. *J Pediatr Gastroenterol Nutr* *54*, 136–160.

Jabri, B., and Abadie, V. (2015). IL-15 functions as a danger signal to regulate tissue-resident T cells and tissue destruction. *Nat Rev Immunol* *15*, 771–783.

Jabri, B., and Sollid, L.M. (2009). Tissue-mediated control of immunopathology in coeliac disease. *Nat Rev Immunol* *9*, 858–870.

Kano, S.-I., Sato, K., Morishita, Y., Vollstedt, S., Kim, S., Bishop, K., Honda, K., Kubo, M., and Taniguchi, T. (2008). The contribution of transcription factor IRF1 to the interferon-gamma-interleukin 12 signaling axis and T<sub>H</sub>1 versus T<sub>H</sub>-17 differentiation of CD4<sup>+</sup> T cells. *Nat Immunol* *9*, 34–41.

Kutlu, T., Brousse, N., Rambaud, C., Le Deist, F., Schmitz, J., and Cerf-Bensussan, N. (1993). Numbers of T cell receptor (TCR)  $\alpha\beta^+$  but not of TcR  $\gamma\delta^+$  intraepithelial

lymphocytes correlate with the grade of villous atrophy in coeliac patients on a long term normal diet. *Gut* *34*, 208–214.

Lebrero-Fernández, C., Wenzel, U.A., Akeus, P., Wang, Y., Strid, H., Simrén, M., Gustavsson, B., Börjesson, L.G., Cardell, S.L., Öhman, L., et al. (2016). Altered expression of butyrophilin (BTN) and BTN-like (BTNL) genes in intestinal inflammation and colon cancer. *Immun Inflamm Dis* *4*, 191–200.

Lefranc, M.P. (2003). IMGT, the international ImmunoGeneTics database(R). *Nucleic Acids Res* *31*, 307–310.

Linehan, J.L., Harrison, O.J., Han, S.-J., Byrd, A.L., Vujkovic-Cvijin, I., Villarino, A.V., Sen, S.K., Shaik, J., Smelkinson, M., Tamoutounour, S., et al. (2018). Non-classical immunity controls microbiota impact on skin immunity and tissue repair. *Cell* *172*, 784–796.e18.

Maddelain, D., Colaert, N., Buchanan, I., Hulstaert, N., Gevaert, K., and Martens, L. (2015). The iceLogo web server and SOAP service for determining protein consensus sequences. *Nucleic Acids Res* *43*, W543–W546.

Melandri, D., Zlatareva, I., Chaleil, R.A.G., Dart, R.J., Chancellor, A., Nussbaumer, O., Polyakova, O., Roberts, N.A., Wesch, D., Kabelitz, D., et al. (2018). The  $\gamma\delta$ TCR combines innate immunity with adaptive immunity by utilizing spatially distinct regions for agonist selection and antigen responsiveness. *Nat Immunol* *19*, 1352–1365.

Meresse, B., Curran, S.A., Ciszewski, C., Orbelyan, G., Setty, M., Bhagat, G., Lee, L., Tretiakova, M., Semrad, C., Kistner, E., et al. (2006). Reprogramming of CTLs into natural killer-like cells in celiac disease. *J Exp Med* *203*, 1343–1355.

Monsuur, A.J., de Bakker, P.I.W., Zhernakova, A., Pinto, D., Verduijn, W., Romanos, J., Auricchio, R., Lopez, A., van Heel, D.A., Crusius, J.B.A., et al. (2008). Effective detection of human leukocyte antigen risk alleles in celiac disease using tag single nucleotide polymorphisms. *PLoS ONE* *3*, e2270.

Mueller, S.N., and Mackay, L.K. (2016). Tissue-resident memory T cells: local specialists in immune defence. *Nat Rev Immunol* *16*, 79–89.

Nilsen, E.M., Lundin, K.E., Krajci, P., Scott, H., Sollid, L.M., and Brandtzaeg, P. (1995). Gluten specific, HLA-DQ restricted T cells from coeliac mucosa produce cytokines with Th1 or Th0 profile dominated by interferon gamma. *Gut* *37*, 766–776.

O'Garra, A., and Vieira, P. (2007).  $T_H1$  cells control themselves by producing interleukin-10. *Nat Rev Immunol* *7*, 425–428.

Park, S.L., Zaid, A., Hor, J.L., Christo, S.N., Prier, J.E., Davies, B., Alexandre, Y.O., Gregory, J.L., Russell, T.A., Gebhardt, T., et al. (2018). Local proliferation maintains a stable pool of tissue-resident memory T cells after antiviral recall responses. *Nat Immunol* *19*, 183–191.

Picelli, S., Faridani, O.R., Björklund, A.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc* *9*, 171–181.



Price, D.A., Brenchley, J.M., Ruff, L.E., Betts, M.R., Hill, B.J., Roederer, M., Koup, R.A., Migueles, S.A., Gostick, E., Wooldridge, L., et al. (2005). Avidity for antigen shapes clonal dominance in CD8<sup>+</sup> T cell populations specific for persistent DNA viruses. *J Exp Med* *202*, 1349–1361.

Quigley, M.F., Almeida, J.R., Price, D.A., and Douek, D.C. (2011). Unbiased molecular analysis of T cell receptor expression using template-switch anchored RT-PCR. *Curr Protoc Immunol Chapter 10*, Unit10.33.

Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* *43*, e47.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* *26*, 139–140.

Sallusto, F., Lenig, D., Förster, R., Lipp, M., and Lanzavecchia, A. (1999). Two subsets of memory T lymphocytes with distinct homing potentials and effector functions. *Nature* *401*, 708–712.

Saraiva, M., Christensen, J.R., Veldhoen, M., Murphy, T.L., Murphy, K.M., and O'Garra, A. (2009). Interleukin-10 production by Th1 cells requires interleukin-12-induced STAT4 transcription factor and ERK MAP kinase activation by high antigen dose. *Immunity* *31*, 209–219.

Smithson, M., and Verkuilen, J. (2006). A better lemon squeezer? Maximum-likelihood regression with beta-distributed dependent variables. *Psychol Methods* *11*, 54–71.

Sugahara, S., Shimizu, T., Yoshida, Y., Aiba, T., Yamagiwa, S., Asakura, H., and Abo, T. (1999). Extrathymic derivation of gut lymphocytes in parabiotic mice. *Immunology* *96*, 57–65.

Trynka, G., Hunt, K.A., Bockett, N.A., Romanos, J., Mistry, V., Szperl, A., Bakker, S.F., Bardella, M.T., Bhaw-Rosun, L., Castillejo, G., et al. (2011). Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nat Genet* *43*, 1193–1201.

Vantourout, P., Laing, A., Woodward, M.J., Zlatareva, I., Apolonia, L., Jones, A.W., Snijders, A.P., Malim, M.H., and Hayday, A.C. (2018). Heteromeric interactions regulate butyrophilin (BTN) and BTN-like molecules governing  $\gamma\delta$  T cell biology. *Proc Natl Acad Sci U S A* *115*, 1039–1044.

Wang, Y., Godec, J., Ben-Aissa, K., Cui, K., Zhao, K., Pucsek, A.B., Lee, Y.K., Weaver, C.T., Yagi, R., and Lazarevic, V. (2014). The transcription factors T-bet and Runx are required for the ontogeny of pathogenic interferon- $\gamma$ -producing T helper 17 cells. *Immunity* *40*, 355–366.

Xu, Y., Olman, V., and Xu, D. (2002). Clustering gene expression data using a graph-theoretic approach: an application of minimum spanning trees. *Bioinformatics* *18*, 536–545.

Zaiss, D.M.W., Gause, W.C., Osborne, L.C., and Artis, D. (2015). Emerging functions of amphiregulin in orchestrating immunity, inflammation, and tissue repair. *Immunity* 42, 216–226.

Zhu, J. (2017). GATA3 Regulates the Development and Functions of Innate Lymphoid Cell Subsets at Multiple Stages. *Front Immunol* 8, 1571.

## FIGURE LEGENDS

### Figure 1. $V\delta 1^+$ IELs with hallmarks of tissue residency are permanently expanded in CeD

(A) Frequency of  $V\delta 1^+$  cells among  $CD3^+$  lymphocytes. Right: box plots display first and third quartiles.  $***p < 0.001$ . One-way ANOVA with Tukey's test for multiple comparisons. (B) Absolute numbers of  $V\delta 1^+$  IELs from 3–5 biopsies per donor. Box plot displays first and third quartiles  $***p < 0.001$ . One-way ANOVA with Tukey's test for multiple comparisons. (C) Frequency of  $V\delta 1^+$  IELs among  $CD3^+$  lymphocytes versus the duration of treatment with a GFD. Linear regression. (D) Frequency of  $V\delta 1^+$  IELs among total  $TCR\gamma\delta^+$  T cells. Bottom: cumulative distribution. Healthy controls:  $n = 99$ . Patients with active CeD:  $n = 62$ . Patients with GFD-treated CeD:  $n = 57$ . Kolmogorov-Smirnov test. (E) Frequency of  $CD69^+/CD103^+$  cells among  $V\delta 1^+$  PBLs and IELs. Bottom: box plot displays first and third quartiles. (F) Fraction of cells defined as naïve, central memory (TCM), effector memory (TEM), or terminal effector (TEMRA) based on expression of CD45RA and CCR7.

### Figure 2. Innate-like $V\delta 1^+$ IELs are lost in CeD

(A) Frequency of IELs expressing NKp46 with or without NKp44. Right: box plots display first and third quartiles.  $***p < 0.001$ . One-way ANOVA with Tukey's test for multiple comparisons. (B) Expression of NKp46 or NKp46/NKp44 on control  $V\delta 1^+$  IELs versus age.

(C) Expression of NKp46 and NKp44 on  $V\delta 1^+$  IELs from donor-matched duodenal and right colonic biopsies. (D) Expression of CD107a on IL-15-treated IELs after stimulation with plate-bound  $\alpha TCR\gamma\delta \pm \alpha NKp46$ . \* $p < 0.05$ . Paired t-test. (E) Expression of CD107a on  $V\delta 1^+$  IELs after stimulation with phorbol myristate acetate and ionomycin. Right: box plot displays first and third quartiles. \*\* $p < 0.01$ , \*\*\* $p < 0.001$ . One-way ANOVA with Tukey's test for multiple comparisons.

**Figure 3. Dietary gluten drives the emergence of IFN- $\gamma$ -producing  $V\delta 1^+$  IELs in CeD**

(A) Expression of IFN- $\gamma$  and TNF- $\alpha$  in cells stimulated *ex vivo* with phorbol myristate acetate and ionomycin. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ . One-way ANOVA with Tukey's test for multiple comparisons. (B) Expression of IFN- $\gamma$  and TNF- $\alpha$  in cells from patients with GFD-treated CeD stimulated as in (A) before and after a six-week gluten challenge. Duration of GFD: 1.5 years, 4 years, 7 years, and 20.5 years. \* $p < 0.05$ . Paired t-test.

**Figure 4. The transcriptional program of  $V\delta 1^+$  IELs is permanently altered in CeD**

(A) Transcriptional profiles of NCR $^+$   $V\delta 1^+$  IELs from healthy controls, NCR $^-$   $V\delta 1^+$  IELs from patients with active CeD, and NCR $^-$   $V\delta 1^+$  IELs from patients with GFD-treated CeD compared using minimum spanning tree analysis. (B) Differentially expressed genes from the NK module passing a false discovery rate  $< 10\%$  from any two-way contrast between the  $V\delta 1^+$  IEL populations in (A). Expression values were standardized (mean centered) on a per gene basis. (C) Genes from the cytokine module passing the criteria in (B). (D) Genes from the tissue healing module passing the criteria in (B). (E) Genes from the transcription factor module passing the criteria in (B).

**Figure 5. The  $V\delta 1^+$  IEL TCR repertoire is permanently reshaped in CeD**

(A) Proportion of unique CDR3 $\gamma$  sequences using a particular *TRGV* gene among V $\delta$ 1<sup>+</sup> PBLs and IELs. White lines demarcate individual contributions. Healthy controls: PBLs, n = 7; IELs, n = 8. Patients with active CeD: PBLs, n = 8; IELs, n = 8. Patients with GFD-treated CeD: PBLs, n = 5; IELs, n = 7. \*p<0.05, \*\*p<0.01, \*\*\*p<0.001. Firth's penalized logistic regression and beta regression. See Table S5A. (B) Data in (A) summarized by individual. (C) Proportion of unique CDR3 $\gamma$  sequences using a particular *TRGV* gene among V $\delta$ 1<sup>-</sup> IELs summarized by individual.

**Figure 6. A molecular signature defines V $\delta$ 1<sup>+</sup> IEL expansions in active CeD**

(A) Proportion of unique CDR3 $\gamma$  sequences using a particular amino acid (AA). White lines demarcate individual contributions. Donor numbers as in Figure 5A. ‡ denotes AAs with significant differences between two groups. Firth's penalized logistic regression and beta regression. See Table S5E. (B) Overlapping CDR3 $\gamma$  sequences incorporating the H-J1 motif. (C) Frequency of unique H-J1<sup>+</sup> CDR3 $\gamma$  sequences. Box plots display first and third quartiles. \*\*p<0.01. Kruskal-Wallis rank sum test with Dunn's test for multiple comparisons. (D) Dominant H-J1<sup>+</sup> CDR3 $\gamma$  sequences among patients with active CeD. (E) Frequency of dominant H-J1<sup>+</sup> CDR3 $\gamma$  sequences. 0 denotes samples in which the dominant CDR3 $\gamma$  sequence lacked H-J1. Kruskal-Wallis rank sum test with Dunn's test for multiple comparisons. (F) Frequency of unique H-J1<sup>+</sup> CDR3 $\gamma$  sequences among V $\delta$ 1<sup>+</sup> and V $\delta$ 1<sup>-</sup> IELs from patients with active CeD. (G) Genes from the TCR activation module passing the criteria described in Figure 4B. (H) Expression of Nur77 in V $\delta$ 1<sup>+</sup> IELs versus CD3. Right: box plot displays first and third quartiles. \*\*p<0.01. One-way ANOVA with Tukey's test for multiple comparisons.

**Figure 7. BTNL3/8-reactive V $\delta$ 1<sup>+</sup> IELs are lost in CeD.**

(A) Expression of *BTNL3* and *BTNL8* relative to *GAPDH* in small intestinal biopsies via qPCR. Box plots display first and third quartiles. \*\* $p < 0.01$ , \*\*\* $p < 0.001$ . Kruskal-Wallis rank sum test with Dunn's test for multiple comparisons. (B) Immunohistochemical analysis of *BTNL8* expression in duodenal sections. (C) Downregulation of CD3 and V $\delta$ 1 on the surface of IELs pre-gated for V $\delta$ 1 expression after overnight incubation with HEK293T-*BTNL8*<sup>+</sup> or HEK293T-*BTNL3/8*<sup>+</sup> cells. Gating was patient-specific based on the HEK293T-*BTNL8*<sup>+</sup> condition. Right: box plot displays first and third quartiles. \* $p < 0.05$ , \*\*\* $p < 0.001$ . Kruskal-Wallis rank sum test with Dunn's test for multiple comparisons. (D) Downregulation of CD3 on the surface of V $\delta$ 1<sup>-</sup> IELs. Details and statistics as in (C). (E) Proportion of unique CDR3 $\gamma$  sequences expressing *TRGV4* gene transcripts. Circles highlighted in black for patients with active CeD denote cases where the dominant *TRGV4* transcript incorporated the H-J1 motif. Box plot displays first and third quartiles. (F) SKW3 cells stably expressing clonal TCRs were cultured for 2 hr with varying numbers of untransduced HEK293T cells (HEK293T-UT) or HEK293T-*BTNL3/8*<sup>+</sup> cells. Activation was assessed via the induction of intracellular Nur77 from 3–5 independent experiments per TCR. Error bars display SD. \* $p < 0.05$ . Paired t-test (HEK293T-UT versus HEK293T-*BTNL3/8*<sup>+</sup>). (G) Frequency of V $\delta$ 1<sup>+</sup> IELs (left), expression of NKp46 and NKp44 on V $\delta$ 1<sup>+</sup> IELs (middle), and expression of CD3 and V $\delta$ 1 on IELs after overnight incubation with HEK293T-*BTNL8*<sup>+</sup> or HEK293T-*BTNL3/8*<sup>+</sup> cells (right) for two patients with potential CeD.

## SUPPLEMENTAL INFORMATION

### Figure S1. Innate-like V $\delta$ 1<sup>+</sup> IELs are lost in CeD, Related to Figure 2

(A) Expression of NKG2D on V $\delta$ 1<sup>+</sup> PBLs and IELs. Box plots display first and third quartiles. (B) Expression of CD94 and NKG2A on V $\delta$ 1<sup>+</sup> PBLs and IELs. Box plots display

first and third quartiles. CD94<sup>+</sup>/NKG2A<sup>-</sup>, activating; CD94<sup>+</sup>/NKG2A<sup>+</sup>, inhibitory. \*p<0.05, \*\*\*p<0.001. One-way ANOVA with Tukey's test for multiple comparisons. (C) Frequency of V $\delta$ 1<sup>+</sup> IELs expressing NKp46 among total CD3<sup>+</sup> lymphocytes. The red box depicts individuals with V $\delta$ 1<sup>+</sup> IEL expansions of similar magnitude to those found in patients with CeD. (D) Expression of NKp46 and NKp44 on PBLs. Bottom: box plots display first and third quartiles. (E) Frequency of V $\delta$ 1<sup>+</sup> IELs expressing NKp46 or NKp46/NKp44 versus the duration of treatment with a GFD. (F) Expression of CD107a on V $\delta$ 1<sup>+</sup> IELs after stimulation with plate-bound  $\alpha$ TCR $\gamma\delta$   $\pm$   $\alpha$ NKp46 and  $\alpha$ NKp44. Bottom: data summary. \*p<0.05. Paired t-test. (G) Expression of granzyme B among subsets of IELs. Bottom: box plot displays first and third quartiles.

**Figure S2. The transcriptional program of V $\delta$ 1<sup>+</sup> IELs is permanently altered in CeD, Related to Figure 4**

(A) Differentially expressed genes (DEGs) between NCR<sup>+</sup> V $\delta$ 1<sup>+</sup> IELs from healthy controls and NCR<sup>-</sup> V $\delta$ 1<sup>+</sup> IELs from patients with active CeD. DEGs highlighted in blue were more highly expressed in NCR<sup>-</sup> V $\delta$ 1<sup>+</sup> IELs from patients with active CeD (FDR <5%), and DEGs highlighted in red were more highly expressed in NCR<sup>+</sup> V $\delta$ 1<sup>+</sup> IELs from healthy controls (FDR <5%). (B) DEGs shown in (A) were used to correlate the magnitude of gene expression differences between NCR<sup>+</sup> V $\delta$ 1<sup>+</sup> IELs from healthy controls and NCR<sup>-</sup> V $\delta$ 1<sup>+</sup> IELs from patients with active CeD (x-axis) versus the magnitude of gene expression differences between NCR<sup>+</sup> V $\delta$ 1<sup>+</sup> IELs from healthy controls and NCR<sup>-</sup> V $\delta$ 1<sup>+</sup> IELs from patients with GFD-treated CeD (y-axis). Genes with log<sub>2</sub>FC values >0 were more highly expressed in NCR<sup>-</sup> V $\delta$ 1<sup>+</sup> IELs from patients with active or GFD-treated CeD relative to NCR<sup>+</sup> V $\delta$ 1<sup>+</sup> IELs from healthy controls, and genes with log<sub>2</sub>FC values <0 were more highly expressed in NCR<sup>+</sup> V $\delta$ 1<sup>+</sup> IELs from healthy controls relative to NCR<sup>-</sup> V $\delta$ 1<sup>+</sup> IELs from

patients with active or GFD-treated CeD. Top left:  $\log_2FC$  distribution for all genes in the dot plot summarized as a histogram for each comparison. Pearson correlation. (C) Frequency of  $V\delta 1^+$  IELs expressing NKp46 or NKp46/NKp44 versus age for patients with active CeD. The red box depicts individuals <16 years old with high frequencies of NKp46<sup>+</sup>  $V\delta 1^+$  IELs. (D) Multidimensional scaling plot showing gene expression profile similarity among  $NCR^+$   $V\delta 1^+$  IELs from healthy controls,  $NCR^+$   $V\delta 1^+$  IELs from patients with active CeD,  $NCR^-$   $V\delta 1^+$  IELs from patients with active CeD,  $NCR^+$   $V\delta 1^+$  IELs from patients with GFD-treated CeD, and  $NCR^-$   $V\delta 1^+$  IELs from patients with GFD-treated CeD. (E) DEGs shown in (A) were used to correlate the magnitude of gene expression differences between  $NCR^+$   $V\delta 1^+$  IELs from healthy controls and  $NCR^-$   $V\delta 1^+$  IELs from patients with active CeD (x-axis) versus the magnitude of gene expression differences between  $NCR^+$   $V\delta 1^+$  IELs from healthy controls and  $NCR^+$   $V\delta 1^+$  IELs from patients with active CeD (y-axis). Genes with  $\log_2FC$  values >0 were more highly expressed in  $NCR^+$  and  $NCR^-$   $V\delta 1^+$  IELs from patients with active CeD relative to  $NCR^+$   $V\delta 1^+$  IELs from healthy controls, and genes with  $\log_2FC$  values <0 were more highly expressed in  $NCR^+$   $V\delta 1^+$  IELs from healthy controls relative to  $NCR^+$  and  $NCR^-$   $V\delta 1^+$  IELs from patients with active CeD. Top left:  $\log_2FC$  distribution for all genes in the dot plot summarized as a histogram for each comparison. Pearson correlation.

**Figure S3. The *TRGV4* gene-associated ‘gut signature’ is lost in CeD, Related to Figure 5**

(A) Gating strategy: live  $CD3^+$   $TCR\gamma\delta^+$   $V\delta 1^+$  lymphocytes were flow-sorted for molecular analysis of expressed TCRs. (B) Number of clones per individual/tissue yielding productive sequences for  $TCR\gamma$  and  $TCR\delta$ . (C) Number of unique CDR3 sequences per group/tissue for  $TCR\gamma$  and  $TCR\delta$ . (D) Expression of *TRGV* genes in  $NCR^+$   $V\delta 1^+$  IELs from

healthy controls (n = 8), NCR<sup>+</sup> V $\delta$ 1<sup>+</sup> IELs from patients with active CeD (n = 9), NCR<sup>-</sup> V $\delta$ 1<sup>+</sup> IELs from patients with active CeD (n = 9), NCR<sup>+</sup> V $\delta$ 1<sup>+</sup> IELs from patients with GFD-treated CeD (n = 3), and NCR<sup>-</sup> V $\delta$ 1<sup>+</sup> IELs from patients with GFD-treated CeD (n = 5). Germline transcripts were extracted from the RNA-seq data set. Expression values were standardized (mean centered) on a per gene basis. (E) Frequency of V $\delta$ 1<sup>-</sup> IELs expressing NKp46 or NKp46/NK/p44. Box plot displays first and third quartiles. \*\*\*p<0.001. One-way ANOVA with Tukey's test for multiple comparisons. (F) Proportion of unique CDR3 $\gamma$  sequences using a particular *TRGJ* gene summarized by individual.

**Figure S4. V $\delta$ 1<sup>+</sup> IELs express TCRs with longer CDR3 $\delta$  loops in CeD, Related to Figure 6**

(A) Shannon diversity indices summarized in violin plots for CDR3 $\gamma$  and CDR3 $\delta$  sequences. (B) Proportion of unique CDR3 $\delta$  sequences using a particular amino acid (AA). White lines demarcate individual contributions. Healthy controls: PBLs, n = 7; IELs, n = 8. Patients with active CeD: PBLs, n = 8; IELs, n = 8. Patients with GFD-treated CeD: PBLs, n = 5; IELs, n = 7. ‡ denotes amino acids with significant differences between two groups. Firth's penalized logistic regression and beta regression. See Table S5C. (C) Proportion of unique CDR3 $\delta$  sequences using a particular *TRDJ* gene summarized by individual. (D) Statistical assignment of *TRDD* gene use for each unique CDR3 $\delta$  sequence. Each candidate forward and reverse *TRDD* gene sequence (rows) was tested for a significant substring match to each unique CDR3 $\delta$  sequence (columns). Significant *TRDD* gene assignments (FDR <0.05) are shown in red; non-significant *TRDD* gene assignments (FDR >0.05) are shown in blue. (E) Frequency of unique CDR3 $\delta$  sequences incorporating a particular CDR3 $\delta$  motif summarized by individual. (F) Proportion of unique CDR3 $\delta$  sequences using a particular feature. Healthy controls: PBLs, n = 7; IELs, n = 8.



Patients with active CeD: PBLs, n = 8; IELs, n = 8. Patients with GFD-treated CeD: PBLs, n = 5; IELs, n = 7. \* $p < 0.05$ , \*\*\* $p < 0.001$ . Firth's penalized logistic regression and beta regression. See Table S5D. (G) Cumulative distribution for CDR3 $\delta$  length across groups. \*\* $p < 0.01$ . Kolmogorov-Smirnov test. (H) Unique CDR3 $\gamma$  sequences among V $\delta$ 1<sup>+</sup> IELs from patients with active CeD visualized using iceLogo for enrichment of non-germline-encoded amino acids relative to unique CDR3 $\gamma$  sequences among V $\delta$ 1<sup>+</sup> IELs from healthy controls and patients with GFD-treated CeD. Position 14 is closest to the *TRGJ* gene-encoded segment. (I) Frequency of unique CDR3 $\gamma$  sequences with an H-J1 motif summarized by individual.

**Figure S5. BTNL3/8-reactive V $\delta$ 1<sup>+</sup> IELs are lost in CeD, Related to Figure 7**

(A) Proportion of unique CDR3 $\gamma$  sequences using the *TRGV4* gene versus relative expression of *BTNL3* and *BTNL8* for patients with GFD-treated CeD. Linear regression. (B) Expression of BTNL3 (myc<sup>+</sup>) and BTNL8 (HA<sup>+</sup>) on untransduced (UT) HEK293T cells (left), HEK293T cells transduced with BTNL8-HA (middle), and HEK293T cells transduced with BTNL3-myc and BTNL8-HA (right). (C) Downregulation of CD3 and V $\delta$ 1 on the surface of IELs pre-gated for V $\delta$ 1 expression after stimulation for 2 hr with 1.5  $\mu$ g/mL of plate-bound purified  $\alpha$ CD3. (D) Downregulation of CD3 and V $\delta$ 1 on the surface of IELs pre-gated for V $\delta$ 1 expression after overnight incubation with HEK293T-BTNL8<sup>+</sup> or HEK293T-BTNL3/8<sup>+</sup> cells. (E) SKW3 cell lines stably expressing clonal TCRs were cultured overnight with HEK293T-UT (black), HEK293T-BTNL8<sup>+</sup> (blue), or HEK293T-BTNL3/8<sup>+</sup> cells (red). Top: representative histogram overlays displaying surface expression of CD3 on SKW3 cells. Bottom: box plots show first and third quartiles (n = 3 independent experiments). \* $p < 0.05$ , \*\* $p < 0.01$ . One-way ANOVA followed by Tukey's test for multiple comparisons. (F) SKW3 cells stably expressing clonal TCRs were cultured for

2 hr with varying numbers of untransduced HEK293T cells (HEK293T-UT) or HEK293T-BTNL3/8<sup>+</sup> cells. Representative contour plots from a single experiment display expression of CD3 and Nur77. (G) Expression of CD3 on unstimulated (black) or  $\alpha$ CD3/ $\alpha$ CD28-stimulated SKW3 cells (red) stably expressing the indicated TCRs. (H) Surface expression of CD3 and intracellular expression of Nur77 for the indicated SKW transductants cultured with HEK293T-UT cells (untreated) or stimulated with  $\alpha$ CD3/ $\alpha$ CD28 beads.

**Figure S6. The tissue-resident V $\delta$ 1<sup>+</sup> IEL compartment is permanently reshaped in CeD, Related to Figures 1–7**

(A) Top: V $\delta$ 1<sup>+</sup> IELs are expanded in patients with CeD and adopt a tissue-resident phenotype characterized by expression of CD69 and CD103. Mean frequency values are summarized by group/tissue. Middle/top: V $\delta$ 1<sup>+</sup> IELs expressing NKp46 are lost in CeD and replaced by IFN- $\gamma$ -producing V $\delta$ 1<sup>+</sup> IELs. Mean frequency values are summarized by group/tissue. Middle/bottom: V $\delta$ 1<sup>+</sup> IELs lose the *TRGV4* gene-associated ‘gut signature’ in patients with CeD (data summarized by group/tissue). This loss is associated with the emergence of CDR3 $\gamma$  sequences incorporating the H-J1 motif among V $\delta$ 1<sup>+</sup> IELs in patients with active CeD. These H-J1<sup>+</sup> CDR3 $\gamma$  sequences become less common after exclusion of dietary gluten. Bottom: *BTNL8* expression is lost in patients with active CeD, and V $\delta$ 1<sup>+</sup> IELs no longer recognize BTNL3/8. Although *BTNL8* expression levels recover on a strict GFD, BTNL3/8 reactivity is permanently lost among V $\delta$ 1<sup>+</sup> IELs. (B) V $\delta$ 1<sup>+</sup> IELs in the healthy state (black) express NKp46 and NKp44, as well as V $\gamma$ 4<sup>+</sup>/V $\delta$ 1<sup>+</sup> TCRs that recognize BTNL3/8. These activating NCRs allow healthy V $\delta$ 1<sup>+</sup> IELs to recognize and eliminate stressed, infected, or malignant IECs. In patients with CeD, decreased expression of *BTNL3* and *BTNL8* is accompanied by a loss of V $\gamma$ 4<sup>+</sup>/V $\delta$ 1<sup>+</sup> IELs, which are replaced by V $\delta$ 1<sup>+</sup> IELs (red) that produce IFN- $\gamma$  in a gluten-dependent manner and express TCR $\gamma$

chains enriched for the H-J1<sup>+</sup> CDR3<sub>γ</sub> motif that fail to recognize BTNL3/8. These H-J1<sup>+</sup> Vδ1<sup>+</sup> IELs contract after withdrawal of dietary gluten, but are not replaced by NCR<sup>+</sup> Vγ4<sup>+</sup>/Vδ1<sup>+</sup> IELs. Instead, Vδ1<sup>+</sup> IELs in patients with GFD-treated CeD are enriched for TCRs that fail to recognize BTNL3/8. Repertoire diversity also increases (purple color gradient), suggesting of a lack of selection pressure in the absence of gluten-induced inflammation. This model is consistent with a fundamental reshaping of the tissue-resident Vδ1<sup>+</sup> IEL compartment after the onset of CeD.

**Figure S7. Alterations to the Vδ1<sup>+</sup> IEL compartment precede tissue damage in CeD, Related to Figure 7**

(A) Frequency of Vδ1<sup>+</sup> cells among total CD3<sup>+</sup> lymphocytes. Box plot displays first and third quartiles. \*\*\*p<0.001. One-way ANOVA with Tukey's test for multiple comparisons. (B) Expression of *BTNL3* and *BTNL8* relative to *GAPDH* in small intestinal biopsies determined via qPCR. Box plots display first and third quartiles. \*p<0.05, \*\*p<0.01, \*\*\*p<0.001. Kruskal-Wallis rank sum test with Dunn's test for multiple comparisons. (C) Frequency of Vδ1<sup>+</sup> IELs expressing NKp46 with or without NKp44. Box plot displays first and third quartiles. \*\*p<0.01. One-way ANOVA with Tukey's test for multiple comparisons. (D) Proportion of unique CDR3<sub>γ</sub> sequences using a particular *TRGV* gene among Vδ1<sup>+</sup> IELs. White lines demarcate individual contributions. (E) Frequency of unique CDR3<sub>γ</sub> sequences incorporating the H-J1 motif among Vδ1<sup>+</sup> IELs. Box plot displays first and third quartiles. \*\*p<0.01. Kruskal-Wallis rank sum test with Dunn's test for multiple comparisons.

**STAR METHODS**

## **CONTACT FOR REAGENT AND RESOURCE SHARING**

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Bana Jabri ([bjabri@bsd.uchicago.edu](mailto:bjabri@bsd.uchicago.edu)).

## **EXPERIMENTAL MODEL AND SUBJECT DETAILS**

Patients were classified into four groups for the purposes of this study. Control – symptoms of upper gastrointestinal tract disease, with no histological evidence of duodenal inflammation, no family history of CeD, and negative serology for TG2. Active – histological evidence of villous atrophy, with positive serology for TG2. GFD – established diagnosis of CeD, with no histological evidence of villous atrophy, and negative serology for TG2. Potential – no histological evidence of villous atrophy, with positive serology for TG2. Exclusion criteria included immunosuppressive medication and coincident diagnoses of Barrett's esophagus, eosinophilic esophagitis, cancer, cirrhosis, or inflammatory bowel disease (IBD). In challenge experiments, patients with GFD-treated CeD consumed 6 g of gluten daily for six weeks. Duodenal biopsies were taken from 3–5 distinct sites, together with 2–8 mL of venous blood, under protocol 12623B approved by the Chicago Biomedicine Institutional Review Board.

## **METHOD DETAILS**

### **Lymphocyte isolation**

PBLs were isolated from whole blood via standard density gradient centrifugation. IELs were isolated from duodenal biopsies via mechanical disruption. Briefly, duodenal tissues were shaken at 250 rpm for 30 min at 37 °C in 7 mL of RPMI 1640 medium supplemented with 1% dialyzed fetal bovine serum (Biowest), 2 mM EDTA (Corning), and 1.5 mM MgCl<sub>2</sub> (Thermo Fisher Scientific). The procedure was repeated once with fresh medium to

enhance cell recovery. Cells were harvested from the biopsy-free media via centrifugation and pooled for subsequent analyses.

### **Flow cytometry**

The following directly conjugated antibodies were used to identify cell surface markers:  $\alpha$ V $\delta$ 1 APC (REA173),  $\alpha$ V $\delta$ 1 FITC (TS8.2),  $\alpha$ V $\delta$ 2 PE (B6),  $\alpha$ V $\delta$ 2 PerCP (B6),  $\alpha$ TCR $\alpha\beta$  BV421 (IP26),  $\alpha$ TCR $\gamma\delta$  PE (5A6.E9),  $\alpha$ TCR $\gamma\delta$  PE-Cy5 (5A6.E9),  $\alpha$ CD3 APC (UCHT1),  $\alpha$ CD3 APC-Cy7 (UCHT1),  $\alpha$ CD3 PE-Cy7 (UCHT1),  $\alpha$ CD3 V450 (UCHT1),  $\alpha$ CD4 APC (RPA-T4),  $\alpha$ CD4 BV786 (SK3),  $\alpha$ CD8a BUV496 (RPA-T8),  $\alpha$ CD8a BV510 (RPA-T8),  $\alpha$ CD8a BV650 (RPA-T8),  $\alpha$ CD45 BV711 (HI30),  $\alpha$ CD45RA BV510 (HI100),  $\alpha$ CD69 PE (FN50),  $\alpha$ CD69 PE-CF594 (FN50),  $\alpha$ CD103 BUV395 (Ber-ACT8),  $\alpha$ CD107a BUV395 (H4A3),  $\alpha$ CCR7 PE-Cy7 (G043H7),  $\alpha$ NKp44 APC (p44-8),  $\alpha$ NKp44 PE (p44-8),  $\alpha$ NKp46 BV605 (9E2),  $\alpha$ NKp46 PE (9E2),  $\alpha$ Nur77 PE (12.14),  $\alpha$ Myc-Tag PE (9B11), and  $\alpha$ HA-Tag Alexa Fluor 647 (6E2). For intracellular cytokine detection, cells were fixed/permeabilized using a BD Cytotfix/Cytoperm Plus Fixation/Permeabilization Solution Kit (BD Biosciences) and stained with the following directly conjugated antibodies:  $\alpha$ IFN- $\gamma$  APC (4S.B3) and  $\alpha$ TNF- $\alpha$  PE-Cy7 (MAb11). Antibodies were purchased from BD Biosciences, BioLegend, eBioscience, or Thermo Fisher Scientific. Dead cells were excluded from the analysis using LIVE/DEAD Fixable Aqua or LIVE/DEAD Fixable Near-IR (Thermo Fisher Scientific). All flow cytometry data were analyzed using FlowJo software version 10.2 (Tree Star).

### **Transcriptome sequencing**

Full-length cDNA and sequencing libraries were generated using a modified version of the single-cell Smart-seq2 protocol (Picelli et al., 2014). Briefly, 50–100 NCR<sup>+</sup> (NKp46<sup>+</sup>) or NCR<sup>-</sup> (NKp46<sup>-</sup>) V $\delta$ 1<sup>+</sup> IELs were sorted into a lysis buffer containing oligo-dT and dNTPs. Reverse transcription was performed after hybridization of oligo-dT to poly-A RNA. The

resulting cDNA was amplified over 18 thermocycles, purified using an AMPure XP Kit (Beckman Coulter), and quality controlled using a Bioanalyzer High Sensitivity DNA Kit (Agilent). A total of 0.25 ng of cDNA from each sample was fragmented, ligated with adapters (Nextera XT), amplified over 12 thermocycles, and quality controlled using a Bioanalyzer High Sensitivity DNA Kit (Agilent). Libraries were then pooled in equimolar amounts and sequenced on two lanes in duplicate using a HiSeq4000 System (Illumina).

### **HLA genotyping**

DNA samples were genotyped using an ImmunoArray BeadChip (Illumina) with single nucleotide polymorphism (SNP) probes as described previously (Trynka et al., 2011). Data analysis was performed using PLINK v1.9. Five of the six HLA-tagging SNPs reported previously (Monsuur et al., 2008) were covered on the ImmunoArray BeadChip. Genotype call rates were higher than 95%. HLA-DQ was inferred from a combination of genotypes at these five SNPs.

### **TCR sequencing**

Lymphocytes were stained for viability and surface expression of CD3, TCR $\alpha\beta$ , TCR $\gamma\delta$ , V $\delta$ 1, and V $\delta$ 2. Live V $\delta$ 1<sup>+</sup> T cells were then sorted directly into 100  $\mu$ L of RNAlater (Thermo Fisher Scientific) using a FACSAria III flow cytometer (BD Biosciences) according to the gating strategy presented in Figure S3A. Cell numbers and patient details are summarized in Table S3. All expressed *TRG* and *TRD* gene transcripts were amplified using an unbiased template-switch anchored RT-PCR (Davey et al., 2017; Quigley et al., 2011). Amplicons were subcloned, sampled, and sequenced as described previously (Price et al., 2005). Gene use was assigned using the ImMunoGeneTics (IMGT) nomenclature (Lefranc, 2003).

### **TCR/NKR stimulation assay**

Polystyrene flat-bottom 96-well plates (Corning) were coated overnight at 4 °C with one of the following antibody cocktails: mix A – 0.5 µg/mL αTCR $\gamma\delta$  (B1, BioLegend), 1 µg/mL mouse IgG2a  $\kappa$  (MOPC-173, BioLegend), and 1 µg/mL mouse IgG2b  $\kappa$  (MPC-11, BioLegend); or mix B – 0.5 µg/mL αTCR $\gamma\delta$  (B1, BioLegend), 1 µg/mL αNKp44 (195314, R&D Systems), and 1 µg/mL αNKp46 (253415, R&D Systems). Lymphocytes were isolated from whole biopsies either pre-treated or not pre-treated with 10 ng/mL recombinant human IL-15 (BioLegend) for 18 hr at 4 °C and plated at a density of  $2 \times 10^5$  cells per well ( $1 \times 10^6$  cells/mL) in RPMI 1640 medium supplemented with 10% human AB serum (Corning), 0.1% GolgiPlug (BD Biosciences), 0.1% GolgiStop (BD Biosciences), and 10 ng/mL recombinant human IL-15 (if pre-treated with recombinant human IL-15). Surface upregulation of CD107a was quantified via flow cytometry after incubation for 3 hr at 37 °C.

#### **Phorbol myristate acetate/ionomycin stimulation assay**

Lymphocytes were suspended in RPMI 1640 medium supplemented with 10% human AB serum (Corning), 10 ng/mL phorbol myristate acetate (Sigma-Aldrich), 150 ng/mL ionomycin calcium salt (Sigma-Aldrich), 100 U/mL recombinant human IL-2 (NIH AIDS Reagent Program), 0.1% GolgiPlug (BD Biosciences), and 0.1% GolgiStop (BD Biosciences) and distributed at  $2 \times 10^5$  cells per well ( $1 \times 10^6$  cells/mL) in polystyrene flat-bottom 96-well plates (Corning). Surface expression of CD107a and intracellular cytokine production were quantified via flow cytometry after incubation for 3 hr at 37 °C.

#### **Quantitative RT-PCR**

Duodenal biopsies were collected in RNAlater RNA Stabilization Reagent (Qiagen). After incubation for 48 hr at 4 °C, excess solution was removed, and the biopsies were stored at –80 °C. On the day of processing, biopsy material was thawed, suspended in 500 µL of Buffer RLT (Qiagen) supplemented with 1% β-mercaptoethanol (Sigma-Aldrich), and

homogenized in a Bullet Blender 24 (Next Advance) using a 1:1 mix of 0.5 mm and 1.0 mm zirconium oxide beads (Next Advance). RNA was then isolated using an AllPrep DNA/RNA/Protein Mini Kit (Qiagen). cDNA was generated from 200 ng of total RNA using a GoScript Reverse Transcriptase Kit (Promega). Expression of the genes of interest was measured via quantitative RT-PCR using SYBR Advantage qPCR Premix (Clontech) on a Light Cycler 480 Instrument II (Roche). The following parameters were used for amplification: denaturation for 10 sec at 95 °C, annealing for 10 sec at 60 °C, and extension for 10 sec at 72 °C. Human primer sequences were as follows: *GAPDH* forward: 5'-ATGGGGAAGGTGAAGGTCG-3'; *GAPDH* reverse: 5'-GGGGTCATTGATGGCAACAATA-3'; *BTNL3* forward: 5'-TCAGTTTCTACGAGCTGGTGTC-3'; *BTNL3* reverse: 5'-CCAAGGCCTGGACAAACTT-3'; *BTNL8* forward: 5'-GCTCTCATGCTCAGTTTGGTT-3'; and *BTNL8* reverse: 5'-GTCTGGCCCAAACACCTG-3'. The *BTNL3* and *BTNL8* primers were described previously (Lebrero-Fernández et al., 2016).

#### **Generation of HEK293T cell lines expressing BTNL3 and BTNL8**

Synthetic DNA (Genscript) encoding full-length human BTNL8 alone or full-length human BTNL3 and BTNL8 separated by a P2A (Teschovirus A) ribosomal skip sequence was subcloned into pMIG (Holst et al., 2006) and used to make retroviral particles (Gras et al., 2010; Holst et al., 2006). Retrovirally transduced (GFP<sup>+</sup>) HEK293T cells were flow-sorted based on expression of the N-terminal myc (EQKLISEEDL(GGS)) tag to identify BTNL3 and the N-terminal HA (YPYDVPDYA(GSG)) tag to identify BTNL8.

#### **Generation of SKW3 cell lines expressing V $\delta$ 1<sup>+</sup> TCRs**

Flow-sorted V $\delta$ 1<sup>+</sup> IELs isolated from healthy controls and patients with GFD-treated CeD were stimulated *in vitro* for two weeks at a starting density of 3 cells per well with irradiated B-lymphoblastoid cell lines and heterologous PBLs in RPMI 1640 medium supplemented



with 300 U/mL recombinant human IL-2 (NIH AIDS Reagent Program), 1 µg/mL phytohemagglutinin (Calbiochem), and 10% human AB serum (Atlanta Biologicals). Paired TCR $\gamma$  and TCR $\delta$  sequences were obtained from clonally expanded V $\delta$ 1<sup>+</sup> IELs. Clonal expansions containing the H-J1 motif were sufficiently large in two patients with active CeD to allow frequency-based pairing of TCR $\gamma$  and TCR $\delta$  sequences obtained directly *ex vivo* from bulk populations of V $\delta$ 1<sup>+</sup> IELs. Representative V $\delta$ 1<sup>+</sup> TCRs were stably expressed via retroviral transduction of the TCR $\alpha\beta$ -deficient T cell leukemia cell line SKW3 (German Collection of Microorganisms and Cell Cultures) (Gras et al., 2010; Holst et al., 2006).

#### **BTNL3/8 reactivity assay *ex vivo***

Round-bottom 96-well plates were seeded overnight with  $2 \times 10^5$  HEK293T-BTNL8<sup>+</sup> or HEK293T-BTNL3/8<sup>+</sup> cells per well to create a monolayer. IELs were isolated from whole biopsies pre-treated with 10 ng/mL recombinant human IL-15 (BioLegend) for 16 hr at 4 °C and overlaid at  $1 \times 10^5$  cells per well on top of the pre-plated HEK cell monolayer. Cells were co-cultured for 24 hr and analyzed via flow cytometry for surface downregulation of CD3 (UCHT1) and V $\delta$ 1 (REA173). As a positive control, IELs were simulated for 2 hr with 1.5 µg/mL of plate-bound purified  $\alpha$ CD3 (UCHT1).

#### **BTNL3/8 reactivity assay *in vitro***

For CD3 downregulation, round-bottom 96-well plates were seeded overnight with  $2 \times 10^5$  HEK293T-UT or HEK293T-BTNL3/8<sup>+</sup> cells per well to create a monolayer, and  $1 \times 10^5$  TCR $\gamma\delta$ -SKW3 cells were then added to each well and incubated for 24 hr at 37 °C. For Nur77 induction, round-bottom 96-well plates were seeded overnight with  $3.125 \times 10^3$ ,  $6.25 \times 10^3$ ,  $1.25 \times 10^4$ , or  $2.5 \times 10^5$  HEK293T-UT or HEK293T-BTNL3/8<sup>+</sup> cells per well to create a monolayer, and  $1 \times 10^5$  TCR $\gamma\delta$ -SKW3 cells were then added to each well and incubated for 2 hr at 37 °C. Flow cytometry was used to measure surface downregulation

of CD3 (UCHT1) or intracellular induction of Nur77 (12.14). As a positive control,  $1 \times 10^5$  TCR $\gamma\delta$ -SKW3 cells were stimulated for 2 hr with 2.5  $\mu$ L of Dynabeads Human T-Activator CD3/CD28 for T Cell Expansion and Activation (Thermo Fisher Scientific).

### **Immunohistochemistry**

Duodenal biopsies were preserved in formalin and embedded in paraffin. Sections were cut to a thickness of 5  $\mu$ m and stained with Bond RX Automatic Stainer (Leica Biosystems). Slides were then dewaxed three times with xylene, ethanol, and water, treated for 20 min with Epitope Retrieval Solution II (Leica Biosystems), incubated for 5 min with 0.5% casein to prevent non-specific binding, and stained for 1 hr with a 1:200 dilution of  $\alpha$ BTNL8 (2187B, R&D Systems). Antigen-antibody binding was revealed using Bond Polymer Refine Detection (Leica Biosystems). Tissue sections were then blocked with peroxidase, stained with DAB, and counter-stained with hematoxylin. Images were acquired using a ScanScope XT microscope (Leica Biosystems).

## **QUANTIFICATION AND STATISTICAL ANALYSIS**

### **Transcriptome analysis**

A total of 34 RNA-seq libraries were generated from flow-sorted V $\delta$ 1<sup>+</sup> IELs, including 8 from healthy controls (8 NCR<sup>+</sup>), 18 from patients with active CeD (9 NCR<sup>+</sup> and 9 NCR<sup>-</sup>), and 8 from patients with GFD-treated CeD (3 NCR<sup>+</sup> and 5 NCR<sup>-</sup>). Adaptor sequences and low-quality score bases (Phred score < 20) were trimmed using Trim Galore (version 0.4.4). The resulting reads were merged per sample and mapped to the human genome reference sequence (Ensembl GRCh38 release 87) using kallisto (version 0.43.0) (Bray et al., 2016). To account for differences in read counts at the tails of the distribution, samples were normalized using the weighted trimmed mean of M-values algorithm (TMM), as implemented in the R package edgeR (Robinson et al., 2010). Data were then log-

transformed using the voom function in the limma package (Ritchie et al., 2015). Non-coding and lowly-expressed genes with an average (across all samples)  $\log_2(\text{CPM})$  lower than 0 were excluded from downstream analyses, leaving a total of 16,066 genes. Linear models that accounted for differences in age and sex were used to identify genes with expression levels that varied between  $\text{NCR}^+$  samples from healthy controls,  $\text{NCR}^+/\text{NCR}^-$  samples from patients with active CeD, and  $\text{NCR}^+/\text{NCR}^-$  samples from patients with GFD-treated CeD, implemented using the lmFit function in the limma package (Ritchie et al., 2015).

### **General procedure for analysis of TCR characteristics**

Sequences were first annotated for *TRGV/J* or *TRDV/J* using the IMGT database and V-QUEST tool (Brochet et al., 2008). Annotated CDR3 sequences were then imported into R for further analysis. All detected sequences were captured in the analysis, including potentially non-functional *TRGV10* gene transcripts. Clonal distribution in each sample was assessed using the Shannon diversity index. Sequence characteristics (*TRV*, *TRD*, and *TRJ* gene usage, and AA composition) were tested for association with patient groups and tissues. For each potentially distinguishing sequence characteristic (*TRV*, *TRD*, and *TRJ* gene usage, and AA composition), enrichment was scored by overall occurrence per group and by trends in occurrence per individual. Results that weighted each unique sequence equally were found to be more robust to potentially non-representative sampling than results that weighted each unique sequence by clonal size. Unique sequences were therefore weighted equally throughout the analysis, and the most strongly enriched features were identified as those passing both group and individual tests. For overall group enrichment testing, each unique TCR sequence labeled with or without the characteristic under question was used as the response variable in Firth's penalized logistic regression performed with the group label as the predictor (logistf R package). This method allowed nearly complete separation of each characteristic by the group label predictor. Linear

hypothesis testing on the resulting group coefficients was then used to estimate the significance of each group comparison (multcomp R package). To test if the occurrence of a characteristic per individual shifted consistently between groups, characteristic proportions were explicitly modeled using beta regression (Smithson and Verkuilen, 2006). The proportion of unique TCR sequences with the characteristic under question was first calculated for each individual. These proportion values were then used as the response variable in beta regression performed with the group label as the predictor (betareg R package). Due to the common presence of 0.0 and 1.0 proportion values (completely absent or present characteristics, respectively), raw proportion values ( $p$ ) were transformed using  $(p \cdot (n - 1) + 0.5)/n$ , where  $n$  is the sample size, to ensure modeled values fell within the open interval (0,1) required for beta regression modeling (Ferrari and Cribari-Neto, 2004). Linear hypothesis testing on beta distribution mean coefficients per group was then used to estimate the significance of each group comparison (car R package).

### **Shannon diversity index**

For each sample, the Shannon diversity index was calculated as  $H = -\sum_n p_i \ln p_i$ , which provides a measure of repertoire diversity. To account for variations in the numbers of recovered CDR3 sequences, individual datasets were subsampled 100 times to 50 CDR3 sequences. The median index was then used to represent the score for each sample (vegan R package).

### **Analysis of *TRGV* and *TRGJ* gene usage**

As described above, logistic regression was used to test for enrichment of *TRGV* and *TRGJ* gene segments in each group and each tissue. Significantly enriched gene segments were then tested for enrichment at the individual level using beta regression.

### **Analysis of CDR3 AAs**

For each full-length CDR3 sequence, the longest substring of AAs from the V and J ends that matched the corresponding *TRV* and *TRJ* germline sequences was identified using reference data provided by the IMGT database (Lefranc, 2003). The remaining AAs were considered non-germline, irrespective of potential contributions from the TRDD segment. The absence or presence of each AA was first determined in each unique CDR3. As described above, logistic regression was then used to test for the enrichment of each AA in each group, and significantly enriched AAs were tested for enrichment at the individual level using beta regression.

### **Analysis of *TRDD* gene usage**

To determine the likely contribution of *TRDD* genes to each TCR $\delta$  sequence, an empirical statistic was generated to distinguish true germline matches from potentially spurious matches arising from random nucleotide additions. For each possible *TRDD* gene-derived nucleotide sequence, the longest substring match was first identified in each candidate CDR3 sequence, and 100,000 random nucleotide strings weighted by the observed nucleotide composition among all non-germline sequences were generated for each possible CDR3 length. By collecting match length statistics between a *TRDD* sequence and these random sequence sets, empirical null distributions of match length were obtained for each *TRDD* gene in both the forward and reverse frames. In each of these distributions, each match length bin contained at least 10 positive occurrences, assuring sufficient sampling coverage. These calculations allowed us to estimate an empirical p-value for an observed *TRDD* gene sequence match, representing the probability of obtaining a match of the observed length or longer by chance. To account for testing a *TRDD* gene sequence against every unique CDR3 sequence, the set of empirical p-values was corrected according to the Benjamini-Hochberg multiple testing procedure, which quantifies the false discovery rate (FDR). This procedure was applied independently to each of three possible *TRDD* genes across both the forward and reverse nucleotide

sequences. Significant *TRDD* gene use was assigned at FDR values <0.05. Sequences were further annotated for each *TRDD* gene to identify the active reading frame. As described above, logistic regression spanning all annotated unique CDR3s was used to test for enrichment of *TRDD* gene segments in each group. Significantly enriched gene segments were then tested for enrichment at the individual level using beta regression. Analysis of *TRDD* gene use, statistical testing, and sequence annotation were performed using custom code written in R.

### **iceLogo motifs**

Non-germline-encoded CDR3 AA sequences were right-justified at the junction with the J segment and analyzed using the iceLogo java application (Maddelein et al., 2015). Unique V $\delta$ 1<sup>+</sup> IEL CDR3 sequences from healthy controls and patients with GFD-treated CeD were pooled to serve as a fixed position background/reference set of sequences for visualization of AA enrichment among V $\delta$ 1<sup>+</sup> IEL CDR3 sequences from patients with active CeD.

### **DATA AND SOFTWARE AVAILABILITY**

RNA-sequencing data were submitted to the Gene Expression Omnibus under accession code [GEO: GSE123649](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE123649).

**Table S1 – Summary of key cohorts, Related to Figures 1–7**

**Table S2 – Gene modules, Related to Figure 4 and Figure 6**

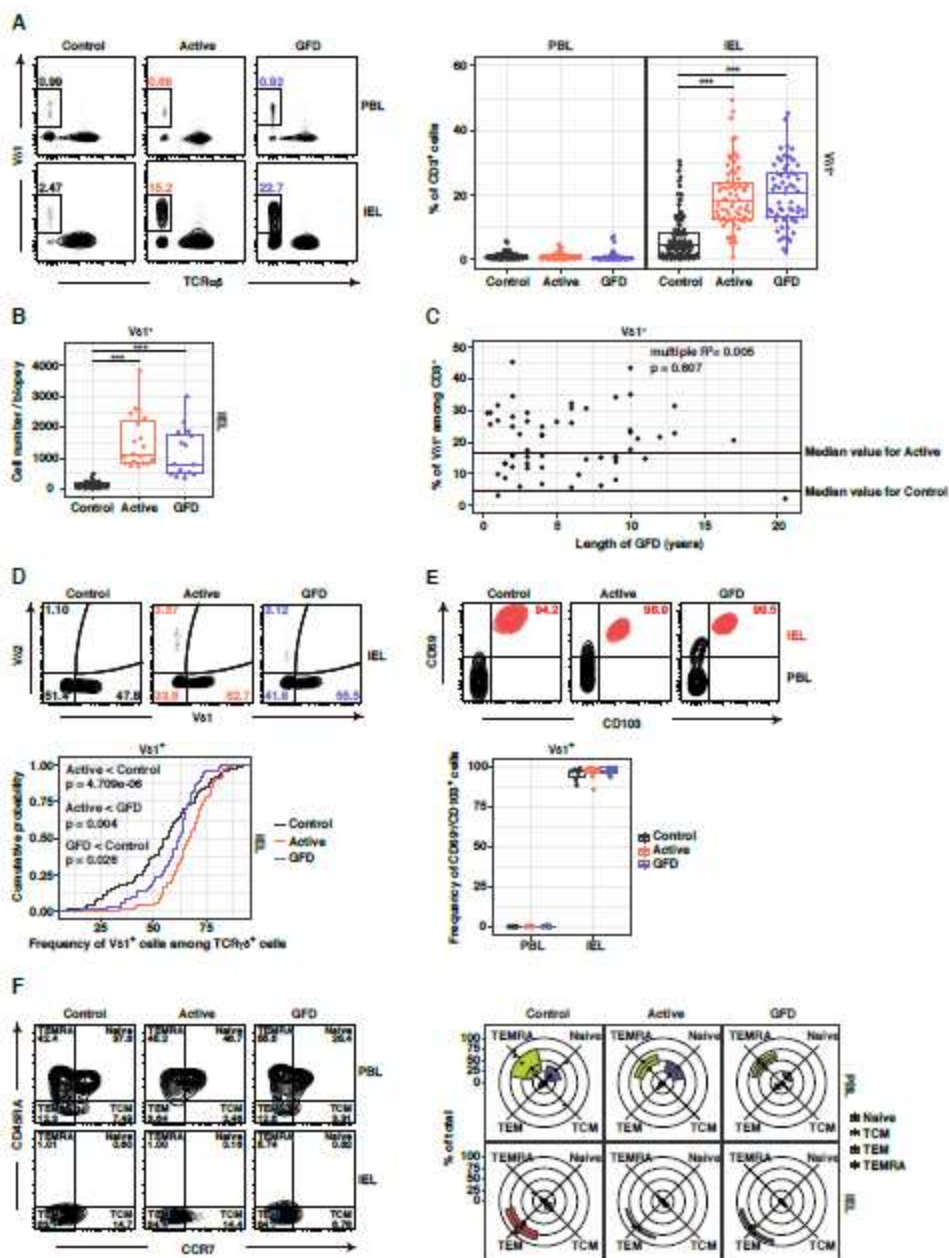
**Table S3 – TCR sequencing cohort, Related to Figure 5 and Figure 6**

**Table S4 – TCR sequencing data set, Related to Figure 5 and Figure 6**

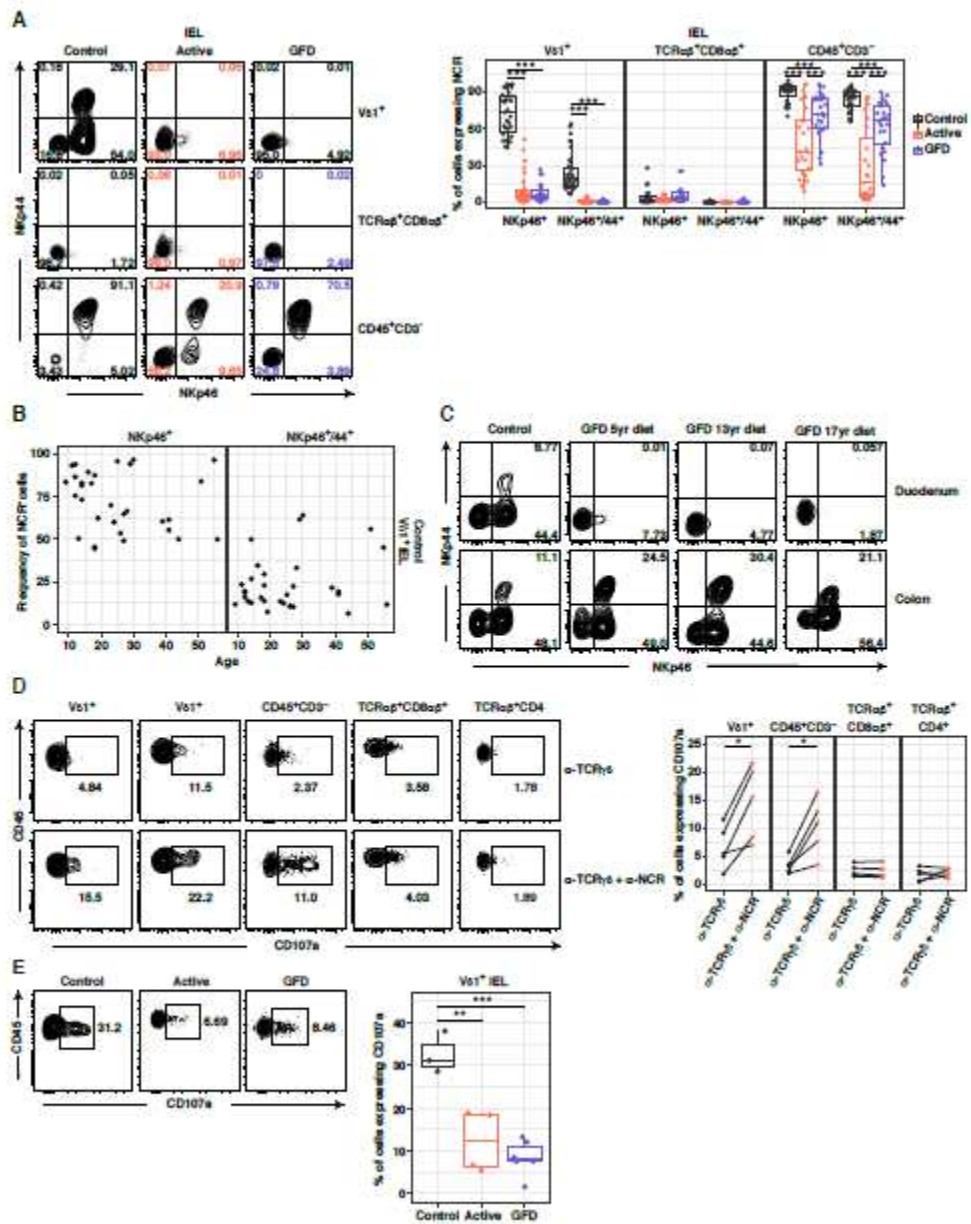
**Table S5 – TCR analysis statistics, Related to Figure 5 and Figure 6**

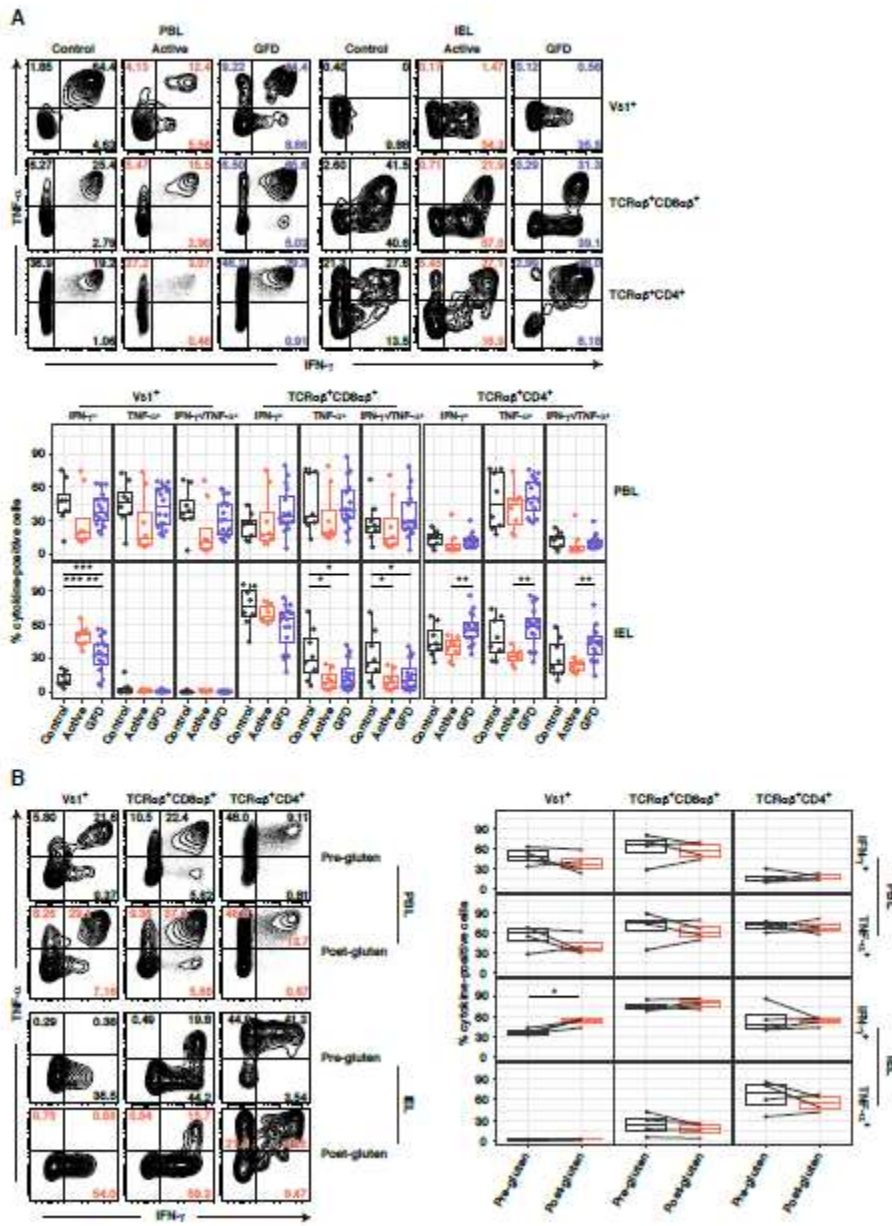
Significant differences that pass both the group and individual level tests are highlighted in red. (A) *TRGV* gene enrichment statistics. (B) *TRGJ* gene enrichment statistics. (C) CDR3 $\delta$  amino acid enrichment statistics. (D) *TRDD* gene enrichment statistics. (E) CDR3 $\gamma$  amino acid enrichment statistics.

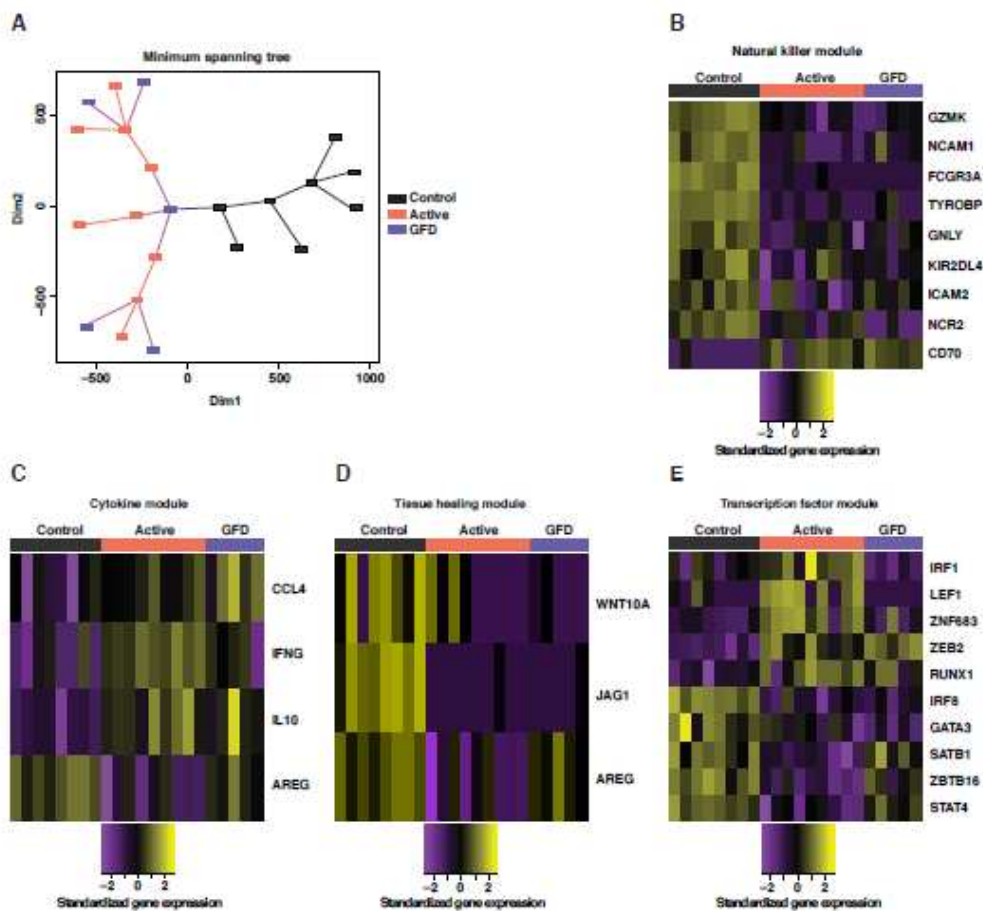
**Table S6 – List of overlapping CDR3 sequences, Related to Figure 6**

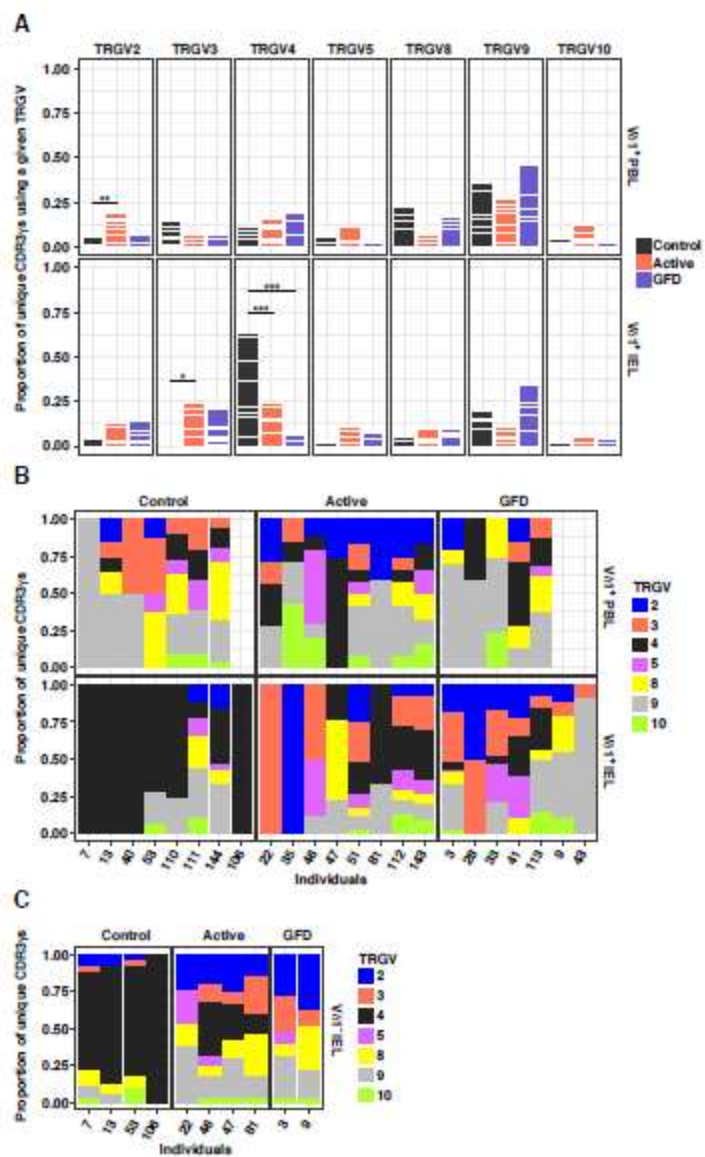


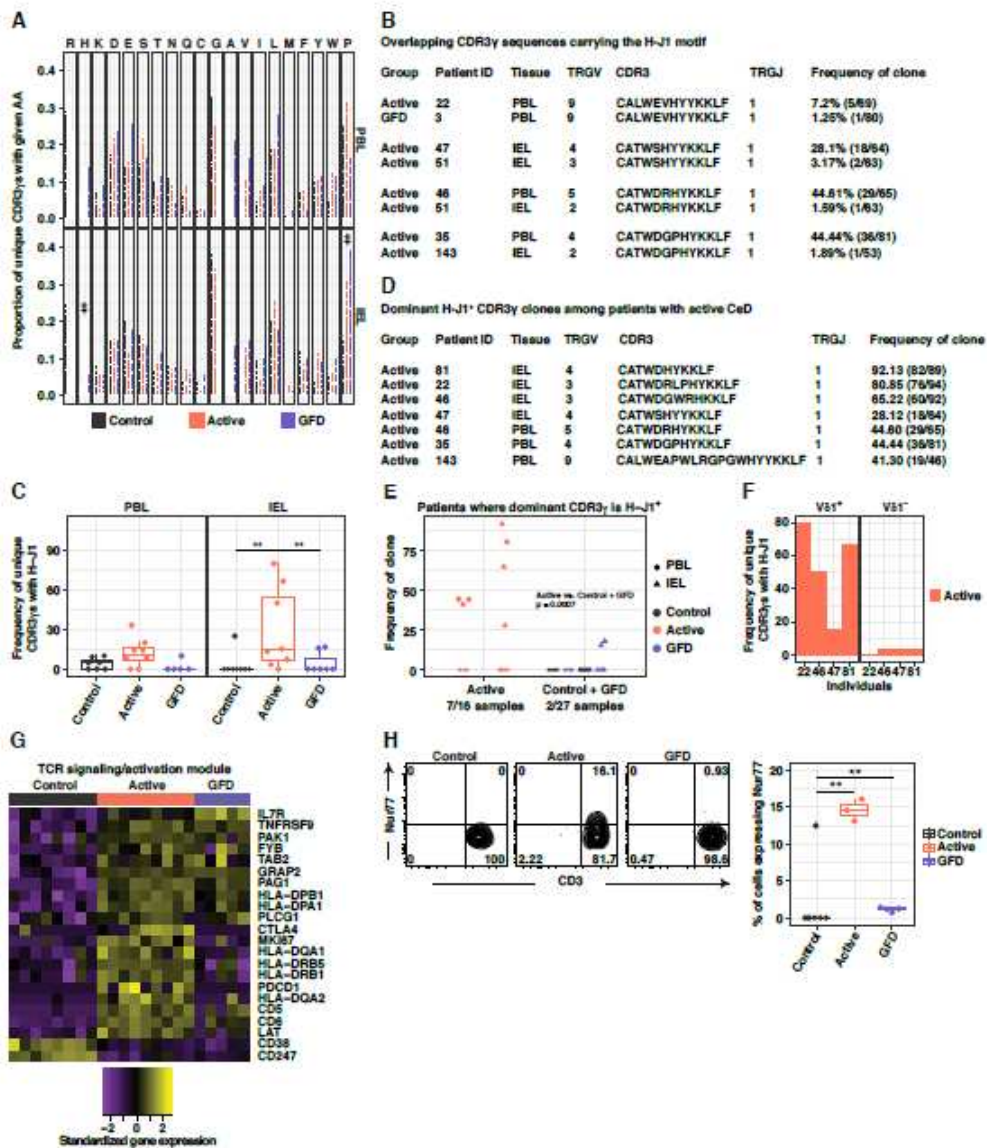




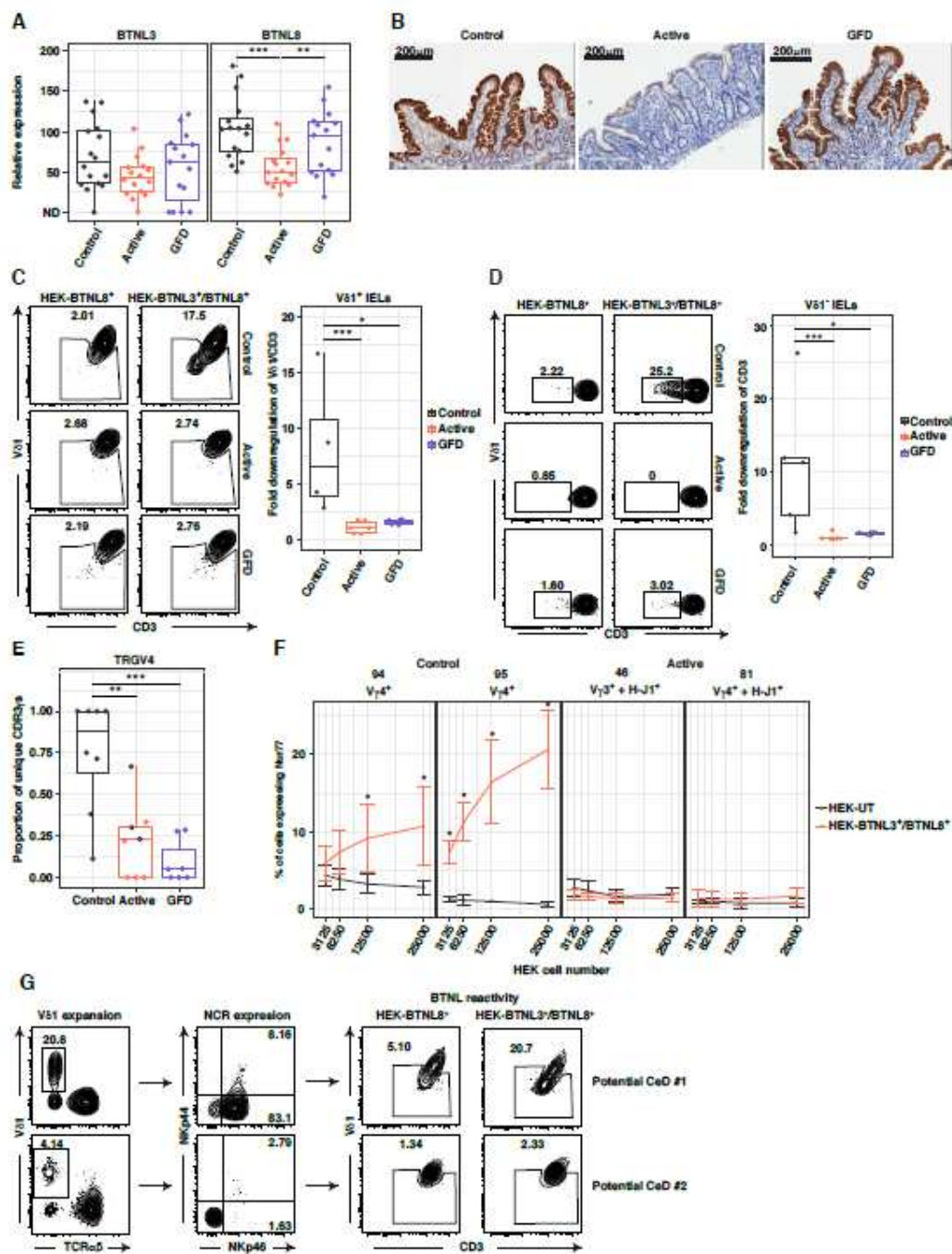


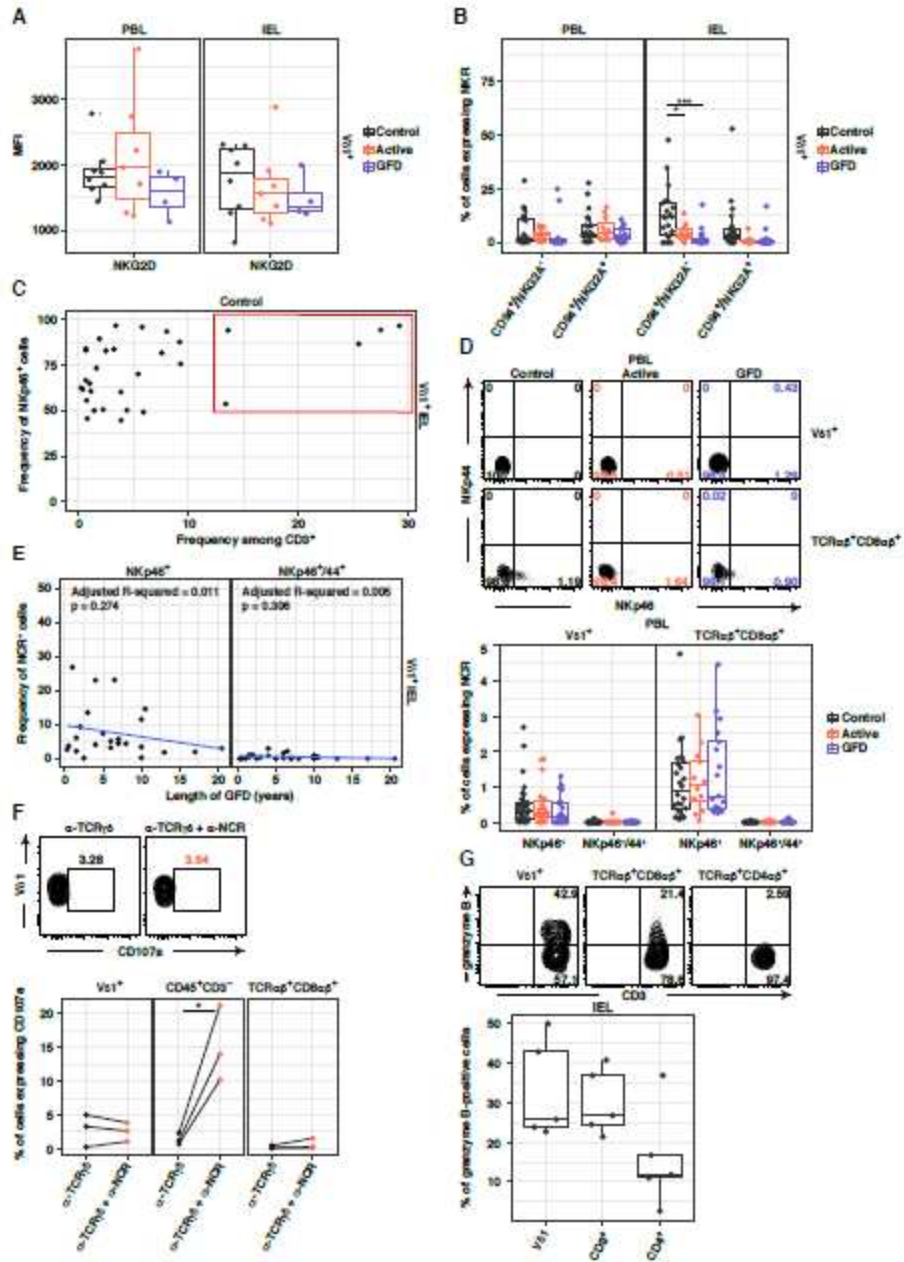


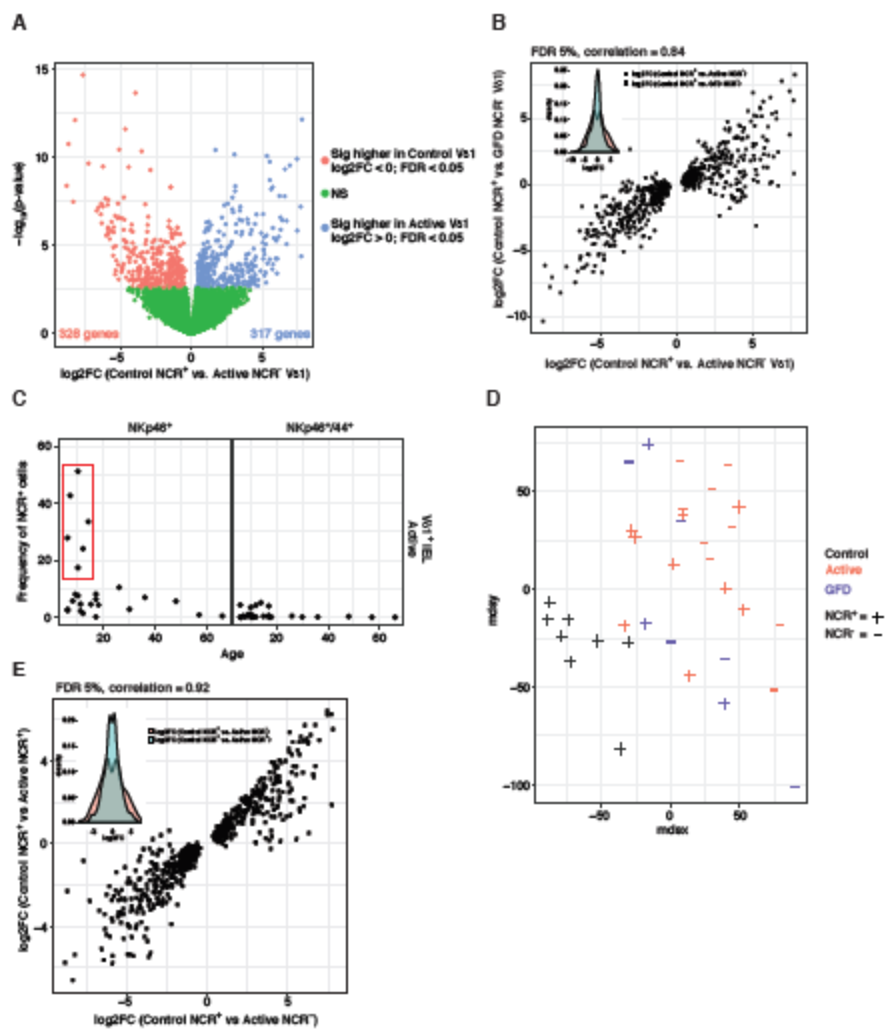




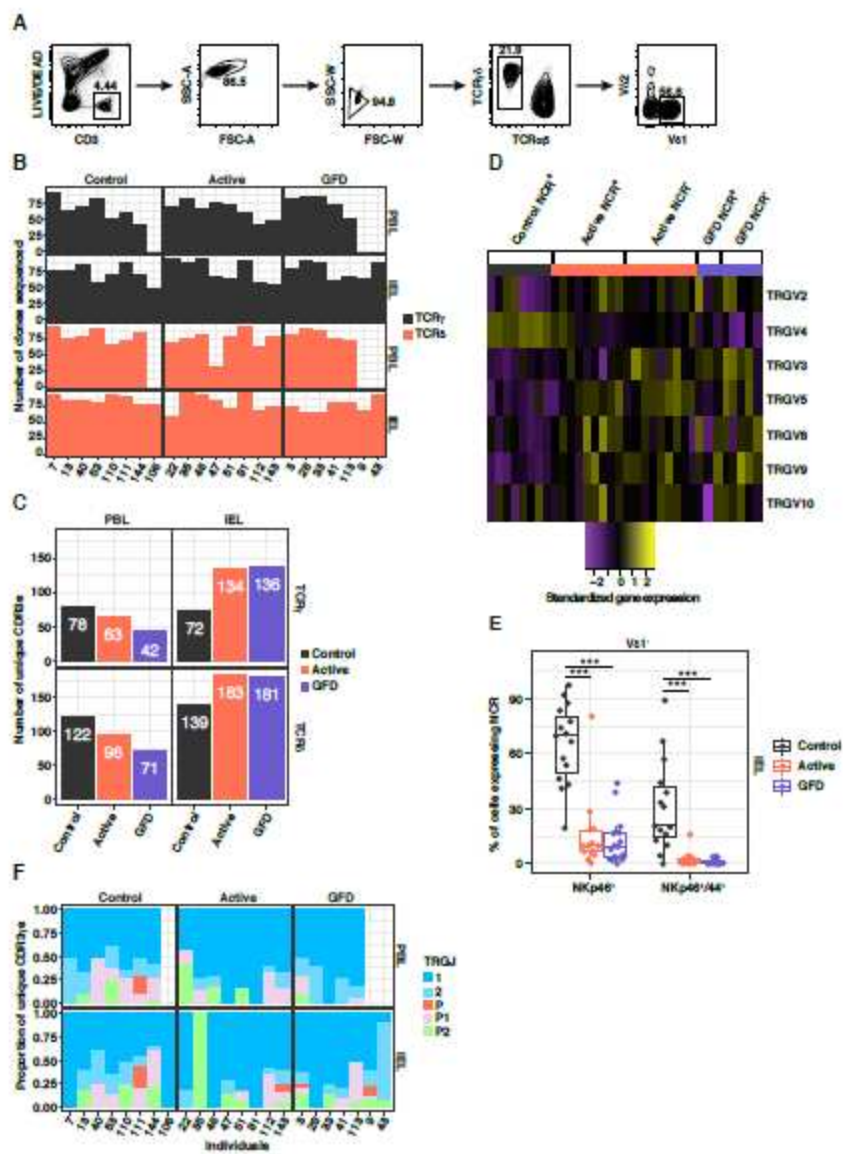


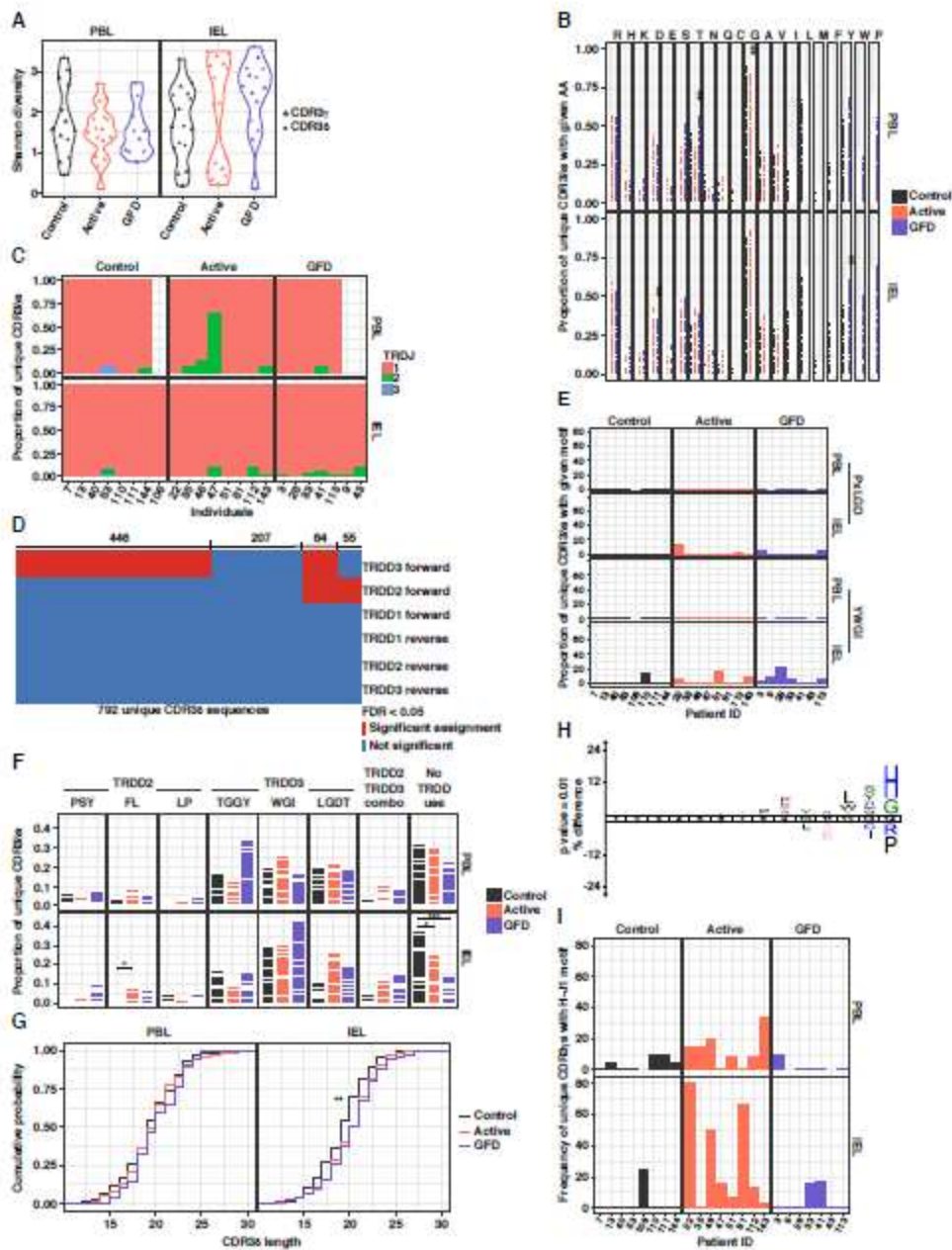


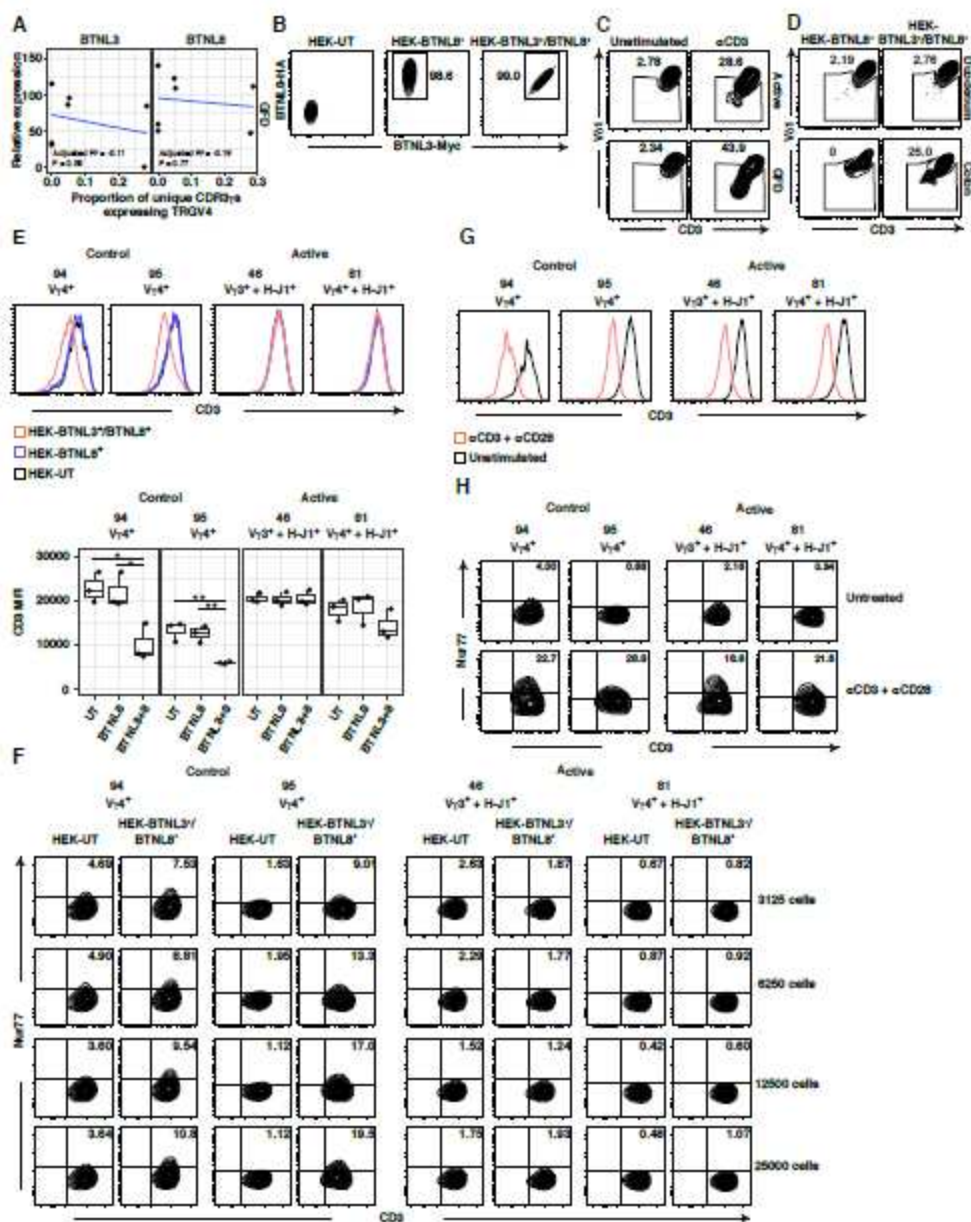




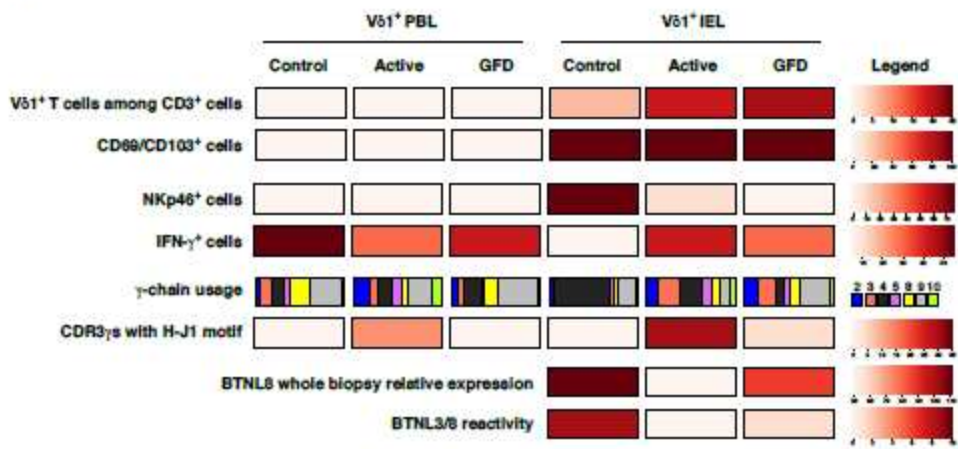








A



B

