OPEN ACCESS

Original research

A novel unconventional T cell population enriched in Crohn's disease

Elisa Rosati ,[1,2] Gabriela Rios Martini,[1,2] Mikhail V Pogorelyy,[3,4] Anastasia A Minervina,[3,4] Frauke Degenhardt,[1] Mareike Wendorff,[1] Soner Sari,[1] Gabriele Mayr,[1] Antonella Fazio,[1] Christel Marie Dowds,[1] Charlotte Hauser,[5] Florian Tran,[1,6] Witigo von Schönfels,[5] Julius Pochhammer,[5] Maria A Salnikova,[3] Charlot Jaeckel,[1] Johannes Boy Gigla,[1] Sanaz Sedghpour Sabet,[1] Matthias Hübenthal ,[1,7] Esther Schiminsky,[2] Stefan Schreiber,[6] Philip C Rosenstiel ,[1] Alexander Scheffold,[2] Paul G Thomas,[4] Wolfgang Lieb,[8] Bernd Bokemeyer,[9] Maria Witte,[10] Konrad Aden,[1,6] Alexander Hendricks,[5,10] Clemens Schafmayer,[5,10] Jan-Hendrick Egberts,[5] Ilgar Z Mamedov,[3,11,12,13] Petra Bacher,[1,2] Andre Franke [1]

► Additional supplemental material is published online only. To view, please visit the journal online (http://dx.doi.org/10.1136/gutjnl-2021-325373).

For numbered affiliations see end of article.

Correspondence to
Dr Elisa Rosati, Institute of Clinical Molecular Biology, Christian-Albrechts-Universitat zu Kiel, Kiel, Schleswig-Holstein, Germany;
e.rosati@ikmb.uni-kiel.de
Professor Andre Franke;
a.franke@mucosa.de

IZM, PB and AF contributed equally.

IZM, PB and AF are joint senior authors.

Received 9 June 2021
Accepted 22 February 2022

To cite: Rosati E, Rios Martini G, Pogorelyy MV, *et al*. *Gut* Epub ahead of print: [*please include* Day Month Year]. doi:10.1136/gutjnl-2021-325373

ABSTRACT

Objective One of the current hypotheses to explain the proinflammatory immune response in IBD is a dysregulated T cell reaction to yet unknown intestinal antigens. As such, it may be possible to identify disease-associated T cell clonotypes by analysing the peripheral and intestinal T-cell receptor (TCR) repertoire of patients with IBD and controls.

Design We performed bulk TCR repertoire profiling of both the TCR alpha and beta chains using high-throughput sequencing in peripheral blood samples of a total of 244 patients with IBD and healthy controls as well as from matched blood and intestinal tissue of 59 patients with IBD and disease controls. We further characterised specific T cell clonotypes via single-cell RNAseq.

Results We identified a group of clonotypes, characterised by semi-invariant TCR alpha chains, to be significantly enriched in the blood of patients with Crohn's disease (CD) and particularly expanded in the $CD8^+$ T cell population. Single-cell RNAseq data showed an innate-like phenotype of these cells, with a comparable gene expression to unconventional T cells such as mucosal associated invariant T and natural killer T (NKT) cells, but with distinct TCRs.

Conclusions We identified and characterised a subpopulation of unconventional Crohn-associated invariant T (CAIT) cells. Multiple evidence suggests these cells to be part of the NKT type II population. The potential implications of this population for CD or a subset thereof remain to be elucidated, and the immunophenotype and antigen reactivity of CAIT cells need further investigations in future studies.

Significance of this study

What is already known on this subject?

- ► T lymphocytes are known to drive Crohn's disease (CD) pathogenesis and progression.
- ► Previous studies identified T-cell receptor (TCR) repertoire traits and specific clonotypes associated with disease status and prognosis.
- ► Interindividual variability challenged the identification and validation of disease-associated TCRs shared among multiple patients.

What are the new findings?

- ► T cells with a semi-invariant TCRα motif were strongly enriched in the blood of patients with CD.
- ► These clonotypes are mostly found among CD8+ T cells and have an unconventional T cell phenotype.
- ► Multiple evidence suggests these cells to be part of the natural killer T (NKT) type II population.

How might it impact on clinical practice in the foreseeable future?

- ► The identified clonotypes draw attention on the possible role of NKT type II cells in CD and deserve further investigation.
- ► These clonotypes may discriminate CD from other forms of IBD, as well as other diseases and provide a new biomarker for CD.

INTRODUCTION

Dysregulated T cell reactions against intestinal antigens are considered to be causal or driving factors for IBD.[1 2] As hypothesised also for other inflammatory and autoimmune diseases, for example, multiple sclerosis[3] or type 1 diabetes,[4] the excessive immune response in IBD may be directed against common, yet still unknown, disease-driving antigens. In genetically predisposed hosts, antigens derived from a dysbiotic microbiome may come into contact with T cells through the impaired

intestinal epithelial barrier, which may lead to uncontrolled and dysregulated immune responses.[2] Accordingly, it may be possible to identify specific disease-associated T cells within the peripheral blood and/or at the intestinal inflammation site of patients with IBD.

Although the two main forms of IBD, namely Crohn's disease (CD) and UC, are often referred to as generally 'IBD', they are clinically two independent diseases with different disease phenotypes but partially overlapping genotypes.[5] CD may localise in any part of the GI tract, mostly in the small intestine, while UC develops only in the colon.[6] Additionally, while both diseases are thought to be mediated by dysregulated T cell responses, different T cell subsets are hypothesised to play a role in disease pathogenesis and progression.[7] Phenotypes and markers distinguishing the two diseases exist but are limited, sometimes resulting in incorrect diagnoses. Further insights differentiating the two forms of IBD could result in a better understanding of the pathophysiology, as well as the development of future strategies for diagnosis and therapy of the diseases.

Most T cells are characterised by a unique T-cell receptor (TCR), formed by an alpha and beta chain. T cells recognise antigenic peptides presented by the major histocompatibility complex (MHC). On recognition of an antigen via the TCR, T cells may expand clonally, resulting in a population of cells with identical TCRs ('clonotypes'). The TCR repertoire, defined as the collection of unique TCRs of an individual, is extremely variable, with up to 10^{11} unique clonotypes present in a single person.[8] Therefore, the composition of the TCR repertoire, that is, the diversity of different TCRs or the expansion of individual clonotypes, could provide important insights into disease-associated alterations of the T cell reaction.

TCR repertoire profiling using high-throughput sequencing has emerged as a method for the analysis of a variety of diseases and conditions.[9] The identification of disease-associated T cell clonotypes in patients with IBD could be an important step towards elucidating how changes in the TCR repertoire impact the diseases, towards the identification of pathogenic T cells and antigenic triggers, and towards the development of new diagnostic and therapeutic strategies. Previous studies identified expanded clonotypes in IBD intestinal samples as compared with blood samples[9–12] and correlated the presence of patient-specific TCRs in the intestinal tissue with disease recurrence after surgery.[12] However, up to now, the identification of disease-associated clonotypes in multiple individuals has proven to be challenging due to the interindividual variability of the diseases as well as of the TCR repertoire. Further, the acquisition of intestinal material is often a limiting factor.[9 10 12] Moreover, most studies look only at the TCR beta repertoire, which is more variable and considered more informative to infer unique clonotypes as well as antigen and HLA specificity. As a result, the TCR alpha repertoire is greatly understudied. Only very few datasets, none of which includes IBD, are available for both alpha and beta chains.

A subset of T cells is formed by the so-called unconventional T cells, which recognise non-classical HLA molecules. The most known unconventional T cells include mucosal associated invariant T (MAIT) cells which bind to MR1, and natural killer T (NKT) cells which bind to the MHC-like molecule CD1d. MAIT cells and NKT type I or invariant NKT (iNKT) cells are characterised by a semi-invariant TCR alpha repertoire formed by the *TRAV1-2*, *TRAJ33/20/12* and *TRAV10*, *TRAJ18* genes, respectively.[13–16] In contrast, NKT type II cells express a polyclonal repertoire. Although there are still many open questions about unconventional T cells, they have been described to recognise ligands of microbial origin and thus to interact with the intestinal microbiota.[15] While MAIT cells are known to be decreased in blood of patients with IBD and to accumulate in the inflamed intestinal mucosa,[17] the role of NKT type II cells in IBD is unclear, with contradictory evidences on the protective or pathogenic role of these cells.[18]

Here, we performed an integrated analysis of multiple independent sample collections for which both alpha and beta TCR repertoire profiling, and HLA data were generated. We detected a selective enrichment of specific T cell clonotypes in the blood of patients with CD, which was further characterised by single-cell RNAseq and TCRseq. The herein identified clonotypes are characterised by a semi-invariant motif of the TCR alpha chain, as well as shared gene expression (GEX) markers with unconventional T cells. Based on our own results and on works by others,[19] we hypothesise these cells to be a new specific subgroup of NKT type II cells, whose enrichment in patients with CD we show for the first time.

A summary of the performed analyses and sample collections is presented in figure 1 and online supplemental table S1.

RESULTS

Identification of TCRs enriched or depleted in the blood of patients with CD

We employed a prospective blood sample collection of 109 CD, 36 patients with UC and 99 healthy controls (online supplemental table S2) with the aim of identifying specific T cell clonotypes associated with CD. Thus, we performed TCR repertoire profiling of the TCR alpha chain and TCR beta chain.[20] Summary of the analysed bulk TCR data can be found in online supplemental table S3.

First, we investigated whether we could identify specific clonotypes that were selectively enriched or depleted in the blood of patients with CD. As a first exploratory analysis, we compared the frequency of TCR alpha and TCR beta clonotypes grouped in TCR families by the variable (V) gene, the joining (J) gene, and CDR3 region length (in amino acids) in patients with CD versus patients with UC and versus healthy controls. Results showed no differences for TCR beta chains (online supplemental figure S1A,B). For TCR alpha chains, however, clonotypes carrying a CDR3 sequence of 12 amino acids and gene combination *TRAV1-2_TRAJ33* were less frequent in patients with CD as compared with healthy individuals and patients with UC (figure 2A and online supplemental figure S1C). This was confirmed via Fisher's exact test of single TCR alpha sequences of this group (p values in online supplemental table S3B). These CDR3 length and gene combination have previously been annotated as MAIT cells.[14] In fact, MAIT cells have been shown to be decreased in the blood of patients with IBD.[21 22]

In patients with CD, a group of TCR alpha chains defined by a 15 amino acids CDR3 region and with the gene combination *TRAV12-1_TRAJ6* were enriched as compared with healthy controls or patients with UC (figure 2A and online supplemental figure S1C). The Fisher's exact test confirmed that >10 clonotypes of the *TRAV12-1_TRAJ6* group were found in a significantly higher number of patients with CD than controls (p values in online supplemental table S3B). Clonotypes of this TCR group, identified as significantly enriched in CD, share the semi-invariant CDR3 motif CVV**A*GGSYIPTF (figure 2B). For simplicity, throughout the manuscript, we refer to these clonotypes as Crohn associated invariant T (CAIT) clonotypes.

While positions 4 and 5 of the identified CDR3 motif are the most variable, positions 6 and 7 are more conserved and

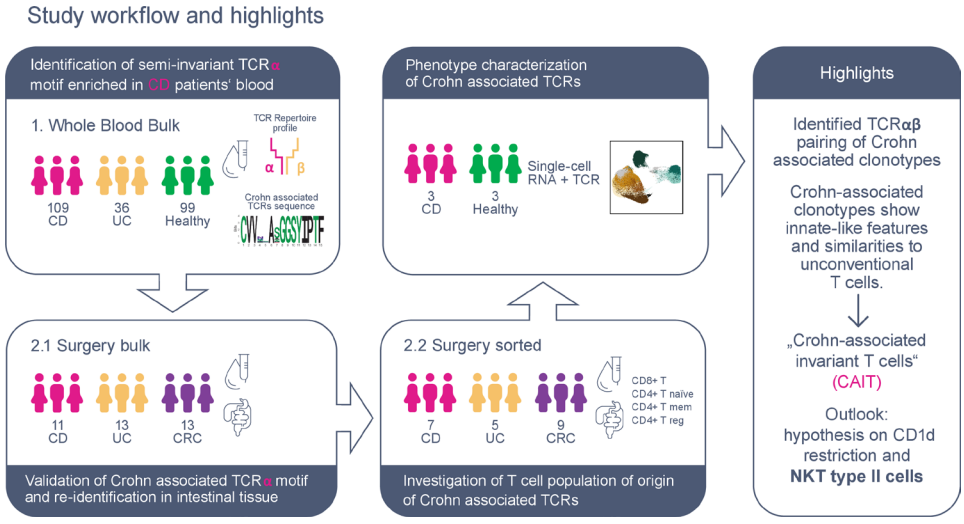


Figure 1 Graphical abstract of study workflow and main results. Two sample collections were used for T-cell receptor (TCR) repertoire profiling of patients with IBD and controls: (1) the whole blood collection included bulk blood samples from IBD and healthy controls; the surgery collection included matched blood, intestinal tissue and intestinal lymph nodes of patients with IBD and disease control (colon cancer, CRC), both from (2.1) bulk tissue and (2.2) sorted T cell populations. Multiple TCRs, sharing a semi-invariant TCR alpha motif, were identified to be enriched in patients with CD and particularly in the $CD8^+$ fraction. Cells carrying this TCR motif were reidentified in single-cell RNA and TCR data of three recontacted patients with CD and three age-matched and sex-matched healthy controls. Single-cell gene expression analysis revealed that CD-associated clonotypes have an innate-like phenotype enriched in $CD8^+$ T cells and comparable to unconventional T cells, that is, mucosal associated invariant T (MAIT) and natural killer T (NKT) cells. Thus, for simplicity, we refer to these clonotypes as Crohn-associated invariant T (CAIT) cells throughout the manuscript. Finally, based on our own result as well as other studies, we hypothesise CAIT clonotypes to be part of NKT type II family and reactive to the CD1d HLA-like molecule.

occupied by A and mostly S amino acids respectively. Moreover, the predicted 3D structure of the TCR alpha CDR3 loop showed that positions 4 and 5 are mostly buried inside the TCR protein structure and may indirectly influence the loop conformation. Positions 6 and 7 are exposed and, together with positions 8 and 11, are predicted to be the most involved into epitope interaction (figure 2C and online supplemental figure S1D). However, CDR3 loops are very flexible and may adapt to different conformations when interacting with different epitopes. Thus, in principle, all CDR3 amino acid positions, excluding the terminal 3–4 residues, may contact the antigen and influence binding affinity.

A higher number of different CAIT clonotypes (different CDR3 regions) were found in patients with CD as compared with healthy controls (figure 2D). In addition, they were also more expanded in the blood of patients with CD (figure 2E) and their cumulative abundance (the sum of their relative abundance in each sample) was accordingly significantly higher in patients with CD as compared with controls (p value=0.008, figure 2F). All in all, at least one CAIT clonotype was present in 49/109 (45%) patients with CD, 7/36 (19%) patients with UC and 37/99 (37%) healthy controls. In 32/109 (29.3%) patients with CD, CAIT TCRs accounted for more than 2.5% of the whole blood repertoire, while this was the case for only 2.7% of patients with UC and 1% of healthy individuals (figure 2F). In contrast, and as previously observed in other studies, clonotypes associated with MAIT cells showed opposite behaviour and were decreased in CD (p value=$3.7*10^{-16}$, online supplemental figure S1E,F).

To further characterise the CAIT clonotypes, we investigated a potential HLA specificity of these TCRs, which would also be an important step towards the identification of antigens against which these TCRs may be reactive to. Furthermore, certain HLA alleles are known to be associated with CD.[23] Accordingly, it is of interest to investigate whether TCRs associated with CD may bind one of them.[24] We genotyped our patients and imputed HLA alleles using previously published reference datasets and methods.[25 26] No significant association was observed between the presence or abundance of CAIT clonotypes and specific HLA alleles of class I (-A, -B, -C) or class II (-DP, -DQ, -DR) (online supplemental table S4A,B). We also investigated whether CAIT TCRs were associated with phenotypic or clinical traits. No significant association was detected between the cumulative abundance or the number of CAIT TCRs, and clinical parameters such as sex, age, smoking behaviour, disease location, disease severity or previous treatment for CD (online supplemental figure S2A-N and online supplemental table S4C). Samples from patients with CD were collected prior to start of treatment with anti-TNF biologics. A non-significant trend was observed, showing increased numbers of CAIT clonotypes in patients non-responding to anti-TNF therapy (p value=0.089, online supplemental figure S2N and online supplemental table S4D). Although not reaching statistical significance, in patients affected by ileal or ileocolonic CD, CAIT TCRs were more abundant as compared with colonic CD (p value=0.07, online supplemental figure S2A and online supplemental table S4C), which may be a potential explanation for the identification of CAIT cells in only a subset of patients.

CAIT clonotypes are present in both blood and intestinal tissue of patients with CD

To verify the presence and abundance of CAIT TCRs in CD, we analysed an independent sample collection (surgery bulk in figure 1) including matched blood and intestinal tissue of 37 individuals undergoing bowel resection surgery, which was composed of 11 patients with CD, 13 patients with UC and 13 patients with colon cancer (CRC) as disease controls (online supplemental table S5). In line with the Whole Blood Bulk dataset, CAIT TCRs were confirmed to be increased in number

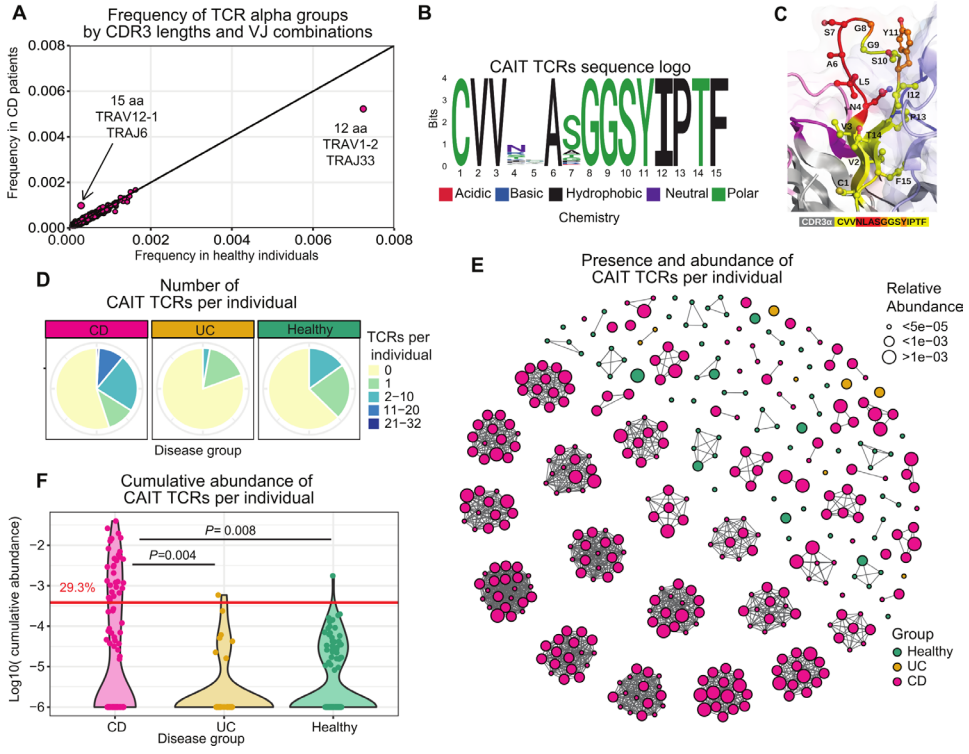


Figure 2 Identification of Crohn-associated invariant T (CAIT) alpha chains in blood. (A) Scatterplot of the frequency of all CDR3 lengths and VJ gene combinations in patients with Crohn's disease (CD) and healthy controls. Enrichment of 15 aa long TRAV12-1_TRAJ6 clonotypes and depletion of 12 aa long TRAV1-2_TRAJ33 clonotypes was observed in patients with CD. (B) Logo plot of the TRAV12-1_TRAJ6 CDR3 amino acid motif enriched in CD. T-cell receptors (TCRs) with this motif are hereafter referred to as 'Crohn-associated TCRs' and define CAIT cells. (C) Zoom-in of CDR3α positions 1–15 (CVVNLASGGSYIPTF) depicting amino acid positions as ball-and-stick and coloured as indicated in the sequence panel below. Amino acid positions 4–7 (NLAS) are coloured in red. In orange, amino acid positions G8 and Y11, that are predicted to directly interact with the epitope together with 6A and 7S due to the orientation of their side chains towards the surface. (D) Pie chart showing the proportion of individuals carrying different numbers of CD-associated clonotypes. (E) Network plot showing enrichment of the specific sequences in CD. Only samples where CAIT clonotypes were present are shown. Each separated group of nodes/cluster represents one individual. Each node represents one CAIT TCR of the (motif group of figure 2B). The size of the node reflects the abundance of the clonotype in a specific sample/individual. Colour of the nodes represents the disease group. It should be noted that the same TCR may be present multiple times in the plot if it is found in multiple individuals. (F) Log-transformed cumulative abundance of CD-associated clonotypes per individual. An abundance of 0 is represented as –6 on the log scale. In patients above the horizontal red line, CAIT clonotypes account for more than 2.5% of the whole blood repertoire, in detail 32/109 tested patients with CD (29.3%) but in only 1/36 (2.7%) patients with UC and 1/99 (1%) healthy individuals. Differences between disease groups have been assessed using Mann-Whitney U test, followed by false discovery rate (FDR) multiple comparison correction.

(figure 3A) and cumulative abundance in the blood of patients with CD (p value=0.002, figure 3B–D). CAIT clonotypes were observed in the blood of 9/11 (82%) patients with CD, 4/13 (31%) patients with UC and 3/13 (23%) patients with CRC. In 7/11 (63.6%) patients with CD, CAIT TCRs composed more than 2.5% of the whole blood repertoire, while they composed maximum 0.3% and 0.6% of the blood repertoire in patients with UC and patients with CRC, respectively. The same non-significant trend was observed in intestinal tissue (p value=0.07 after multiple testing correction). For a subset of patients (4 CD and 5 UC), intestinal mesenteric lymph nodes resected together with the inflamed intestinal tissue were also analysed. CAIT clonotypes were significantly more abundant in the intestinal lymph nodes of patients with CD as compared with patients with UC (p value=0.02, figure 3A–D). While most CAIT TCRs were observed in the blood, few CAIT clonotypes were also observed to be present in multiple analysed tissues of the same individual at the same time (figure 3C).

To answer the question whether CAIT cells occur within a particular T cell compartment, blood and intestinal T cells were sorted in $CD4^+$ naïve, $CD4^+$ conventional memory, $CD4^+$ regulatory and $CD8^+$ T cells for additional 7 patients with CD, 5 patients with UC and 9 patients with CRC (online supplemental tables S1 and S6). Surprisingly, CAIT TCRs were found in all analysed fractions, but at different frequencies. CAIT cells were most abundant in the $CD8^+$ fraction, particularly in patients with CD, with up to 22 CAIT TCRs detected per individual and higher cumulative abundance as compared with the other cell compartments (figure 4A,B). In contrast, CAIT sequences were almost absent in the blood of patients with UC and higher abundance of CAIT sequences in patients with CRC was mostly found in T_{reg} and $T_{Naïve}$ fractions.

Beta chains paired with CAIT alpha chains are private of the single individual and show preferential V gene usage

For three patients with CD with a high proportion of CAIT cells, we next FACS-purified the $CD4^+$ and $CD8^+$ memory fractions and performed single-cell RNA GEX and TCR sequencing using 10× Genomics' Chromium technology. The same analysis was performed for three sex-matched and age-matched healthy controls.

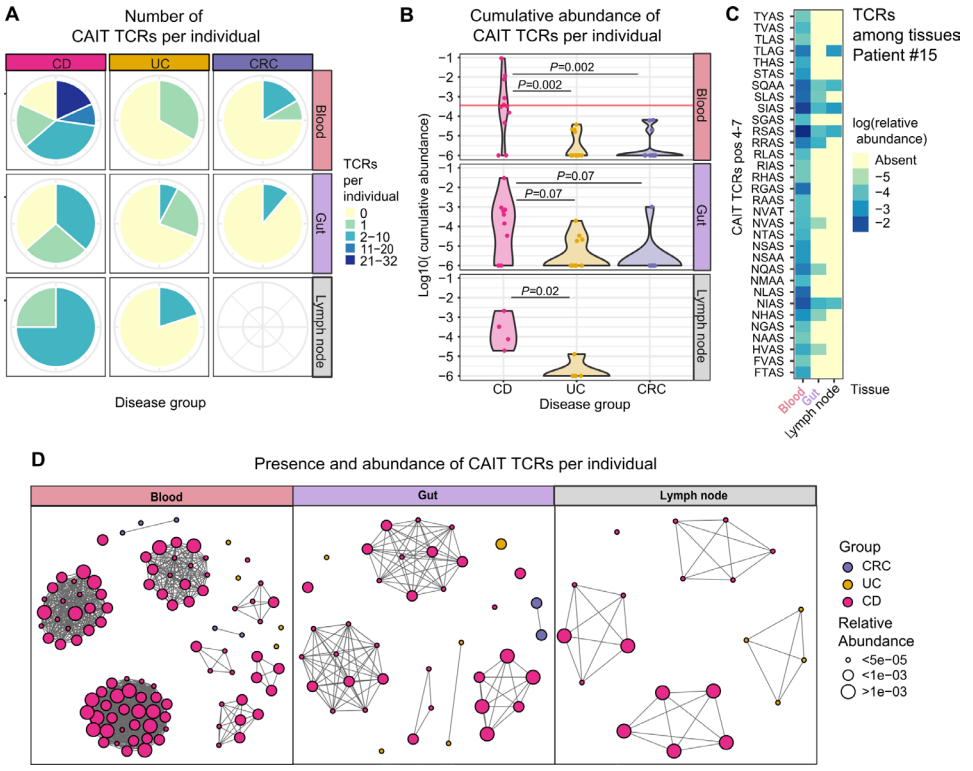


Figure 3 CAIT T-cell receptor (TCR) alpha chains in blood, intestinal tissue and intestinal lymph node. (A) Pie chart showing the proportion of individuals (11 Crohn's disease (CD), 13 UC, 13 colon cancer (CRC)) carrying different numbers of Crohn-associated invariant T (CAIT) clonotypes in blood, gut and, for UC and CD only, intestinal lymph node. (B) Log-transformed cumulative abundance of Crohn-associated clonotypes per individual and tissue. In patients above the horizontal red line, CAIT clonotypes account for more than 2.5% of the whole blood repertoire (7/11 (63.6%) patients with CD, 0% patients with UC and patients with CRC). Differences between disease groups have been assessed using Mann-Whitney U test, followed by FDR multiple comparison correction. (C) Heatmap of the log-transformed relative abundance of CAIT TCRs among the tissues (x-axis) of one representative patient with CD (#15). Labels on the y-axis indicate the variable positions[4–7] of the CDR3 sequence of CAIT TCRs. (D) Network plot showing enrichment of CAIT sequences in CD among the three analysed tissue samples. Each separated cluster thus represents one individual/sample. Each node is one TCR of the motif group of figure 1B. The size of the node expresses the abundance of the clonotype in a specific sample/individual. Colour of the vertices represents the disease group. Of note, the same TCR may be present multiple times in the plot if it is found in multiple individuals.

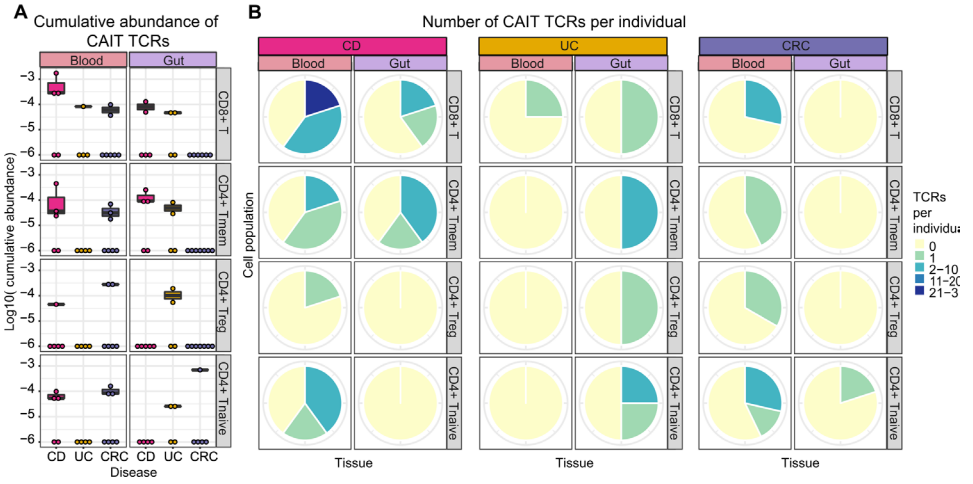


Figure 4 Crohn-associated invariant T (CAIT) T-cell receptor (TCR) alpha chains in sorted T cell populations. (A) Presence and log10-transformed relative abundance of CAIT TCRs (y-axis) in each individual among sorted T cell populations and analysed tissues. Particularly in $CD8^+$ (first panel from top) and $CD4^+$ Tmem (second panel from top), more CAIT TCRs are found in patients with Crohn's disease (CD). (B) Pie chart showing the proportion of individuals, separated by disease group (CD: left, UC: middle, colon cancer (CRC): right) carrying different numbers of Crohn-associated clonotypes in both tissues and four cell populations. Seven patients with CD, five patients with UC and nine patients with CRC were analysed.

In accordance with our initial findings, CAIT cells were identified at high abundance (22–238 CAIT cells per patient) in the three patients with CD through their TCRs (figure 5A,B and online supplemental table S7A), while they were rare in healthy controls (0–2 CAIT cells per healthy individual, figure 5A). Analysis of TCR alpha/beta pairs showed diverse TCR beta chains associated with the semi-invariant CAIT alpha sequences (figure 5B and online supplemental table S7B). Analysis of the TCR alpha/beta chain pairing in CAIT clonotypes, showed a preferential pairing (38%–55% of CAIT clonotypes) with beta chains carrying the *TRBV7-9* gene in all three patients (figure 5C). None of the identified paired TCR beta chains were found in any other individual of the previously analysed bulk TCR datasets (244 whole blood and 59 surgery samples). Thus, these TCR beta sequences seem to be private to single individuals or very rare in the examined population (online supplemental table S7C).

CAIT clonotypes show an innate-like GEX profile

Based on Seurat clustering of GEX analysis,[27] we defined 12 T cell populations of the sorted $CD4^+$ and $CD8^+$ cells, which are shown in figure 5D. Marker genes for each cluster are listed in online supplemental table S7D and summarised in figure 5E. CAIT cells were found mostly in the $CD8^+$ fraction, more precisely in effector memory clusters characterised by high expression of *KLRD1* (CD94) and *KLRB1* (CD161), respectively, which are known markers for innate-like T and natural-killer cells (figure 5A–E and online supplemental figure S3A). Most CAIT cells were found in the $CD8^+$ effector memory $CD161^+$ cluster, which includes known unconventional T cell subtypes such as MAIT and NKT cells. The cluster was also characterised by upregulation of genes such as *SLC4A10*, *CEBDP* and *IL18R* that are typical GEX markers of MAIT cells[28 29] (figure 5E and online supplemental figure S3A,B). In fact, MAIT and iNKT cells were identified in this cluster through their TCRs[30 31] (figure 5A). A smaller subset of CAIT, MAIT and iNKT cells was also found in the CD8 effector memory cluster mostly characterised by expression of *KLRD1*, encoding for CD94, which is another marker of innate immunity and present in both NK cells, NKT cells and subsets of MAIT cells.[32 33] Thus, CAIT cells showed an innate-like GEX profile comparable to known unconventional T cells. Moreover, their abundance in these patients was comparable to MAIT cells, while iNKT cells were very rare. To further investigate phenotypic differences between CAIT and MAIT cells, we run a differential expression analysis specifically on these two groups. It resulted that although the GEX profile of CAIT and MAIT cells is very similar, CAIT cells express higher levels of natural-killer associated genes, such as *KLRD1, GZMH, FGFBP2* and *CX3CR1*, while they downregulate *IL7R* (figure 5F and online supplemental table S7E).[34 35] Additionally, CAIT cells expressed *CD69*, *IFNg* and *TNF* too (online supplemental table S7E).

Individual CAIT and MAIT cells were additionally found in the $CD4^+$ memory clusters, confirming our observations at bulk TCR level, that CAIT clonotypes occur in both fractions with an enrichment in $CD8^+$ cells (figure 4A,B)

Variable frequency of CAIT clonotypes among TRAV12-1+ cells

Next, we wanted to evaluate whether it is possible to efficiently capture and characterise CAIT clonotypes by flow cytometry. We analysed the peripheral blood of 39 patients with CD, 20 patients with UC and 18 healthy controls, as well as gut biopsies from 14 patients with CD, 7 patients with UC and 19 patients with CRC (online supplemental tables S8 and S9). By gating on $CD3^+TRAV12\text{-}1^+CD161^+IL18R^+$ cells (gating strategy in online supplemental figure 4A), we found a significant enrichment of this population in the blood of CD, and surprisingly also in patients with UC, as compared with healthy controls (online supplemental figure 4B). Among these cells, the $CD8^+$ fraction ($CD8^+TRAV12\text{-}1^+CD161^+IL18R^+$) was clearly enriched in patients with CD as compared with healthy donors (online supplemental figure 4C). Contrarily, as shown before by other studies,[17] MAIT cells ($CD3^+TRAV1\text{-}2^+CD161^+IL18R^+$) were decreased in the blood of patients with IBD both in the total $CD3^+$ and the $CD3^+CD8^+$ compartment (online supplemental figure 4D,E). However, analyses of gut biopsies did not fully recapitulate these findings. Although the 'CAIT-like' population ($CD3^+TRAV12\text{-}1^+CD161^+IL18R^+$) had in general a higher frequency in the intestinal tissue (0%–4% of $CD3^+$cells) as compared with blood (0%–1.8% of $CD3^+$cells), we observed only a trend, but no significant enrichment of this population in CD (online supplemental figure 4F-H), contrarily to MAIT cells (online supplemental figure 4I,J).

Specific clonotypes cannot be captured by flow cytometry if the HLA epitope is unknown and thus the accuracy of their detection is limited to the TCR V region, for which specific antibodies are commercially available. Therefore, we used our single-cell data to assess the frequency of CAIT clonotypes among clonotypes expressing a TCR using the *TRAV12-1* gene. We also compared it with the frequency of MAIT clonotypes among $TRAV1\text{-}2^+$ cells. MAIT TCRs accounted for 33%–80% of the analysed $TRAV1\text{-}2^+$ cells in patients with CD and 58%–95% in healthy controls. In contrast, the percentage of CAIT clonotypes among $TRAV12\text{-}1^+$ cells was highly variable, ranging from 1.9% to 36% in patients with CD, and 0% to 0.3% in healthy controls (figure 6A). The frequency of CAIT cells was increased when focussing on cells expressing *KLRB1* (figure 6B), encoding for the CD161 protein. In this subset, CAIT clonotypes were 6.0%–70% of $TRAV12\text{-}1^+$ cells in patients with CD and 0%–2% in healthy controls, while the proportion of MAIT clonotypes among $TRAV1\text{-}2^+$ was 64%–96% in patients with CD and 82%–99% in controls (figure 6C). Maximum frequencies were reached in the $TRAV12\text{-}1^+CD161^+CD8^+$ cells where CAIT cells ranged from 18% to 77% in patients with CD and 0% to 5.8% in healthy controls (figure 6D).

These GEX data suggest that while MAIT clonotypes are strongly represented in a flow cytometry analysis of $TRAV1\text{-}2^+CD161^+$ T cells, CAIT clonotypes have high frequency in some individuals but low frequency in others. Therefore, although GEX levels do not directly translate into protein expression levels, it is reasonable to hypothesise that $TRAV12\text{-}1^+CD161^+$ cells analysed via flow cytometry include CAIT clonotypes but that the high variability between individuals limits the use of flow cytometry for their detection.

DISCUSSION

By means of high-throughput TCR profiling, we here identified a specific group of T cells, defined by a semi-invariant TCR alpha motif present in both blood and intestinal tissue. CAIT clonotypes enrichment in CD was significant in blood and mesenteric lymph nodes while only a trend was observed in intestinal tissue. Potentially, these cells may be activated into the gut and then enter the peripheral circulation. Their accumulation in peripheral blood may also be a sign of systemic disease. We did not

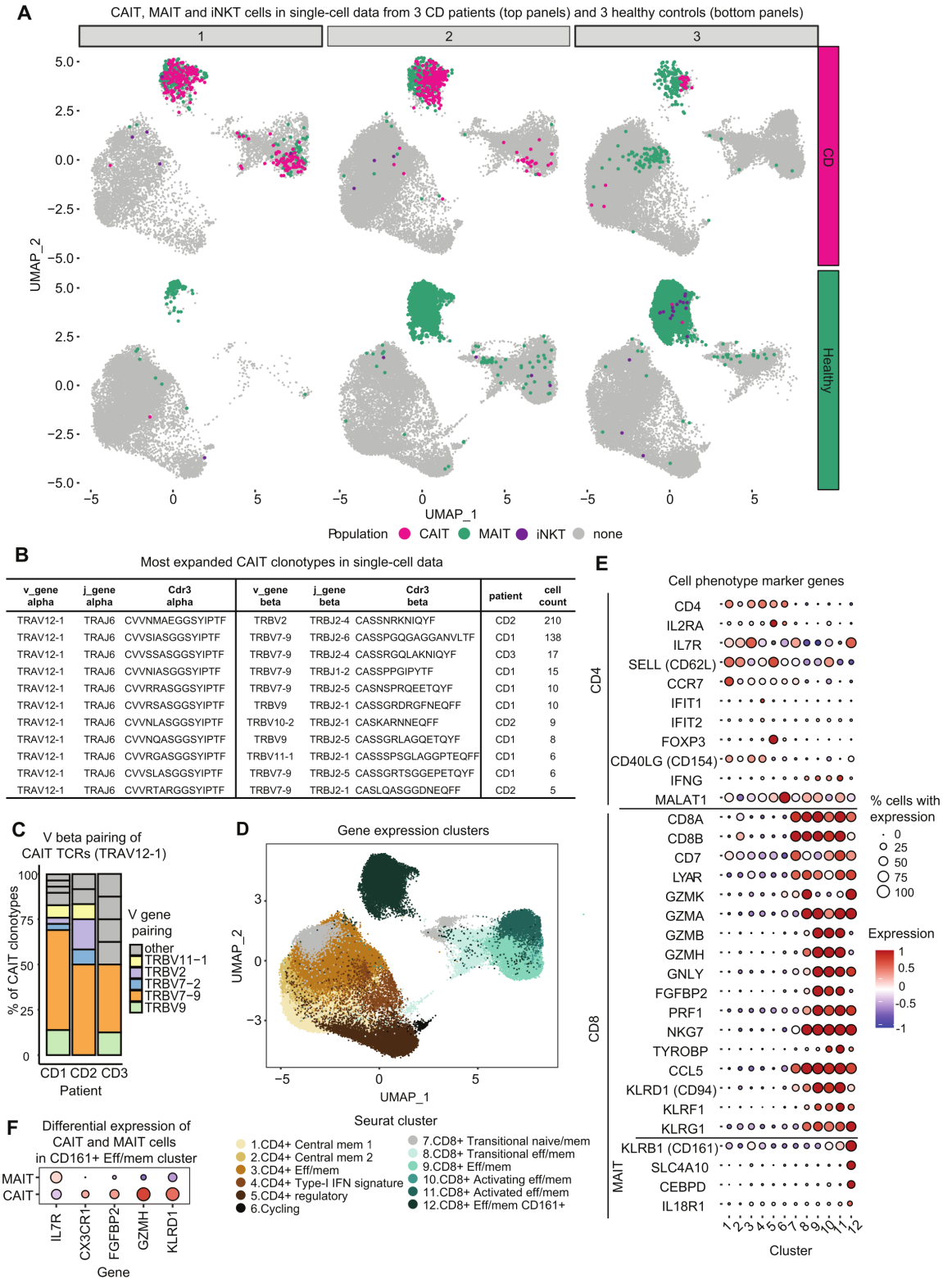


Figure 5 Single-cell analysis of sorted memory CD4 and CD8 cells from three patients with Crohn's disease (CD) and three matched healthy controls. (A) Uniform Manifold Approximation and Projection (UMAP) (co-)localisation of mucosal associated invariant T (MAIT) (green), invariant NKT (iNKT) (violet) and Crohn-associated invariant T (CAIT) (pink) cells. MAIT cells are abundant in both patients with CD (top 3 panels) and healthy controls (bottom 3 panels), while CAIT cells are abundant in patients with CD but very sparse in healthy controls, only 1 (bottom panel left) and 2 (bottom panel right) CAIT cells were found in healthy controls respectively. (B) Paired T-cell receptor (TCR) alpha and beta chains of the most expanded CAIT clonotypes observed in three patients with CD through single-cell TCR analysis. (C) V genes of TCR beta chain pairing for CAIT in the three patients with CD analysed. Plotted as the proportion of unique clonotypes with a certain V gene beta pairing. (D) UMAP single-cell visualisation of Seurat functional clusters. (E) Bubble plot for Seurat cluster marker genes defining the $CD4^+$ and $CD8^+$ different populations. (F) Differentially expressed genes between MAIT and CAIT cells of the same function cluster (cluster 12, effector memory $CD8^+$ $CD161^+$ T cells).

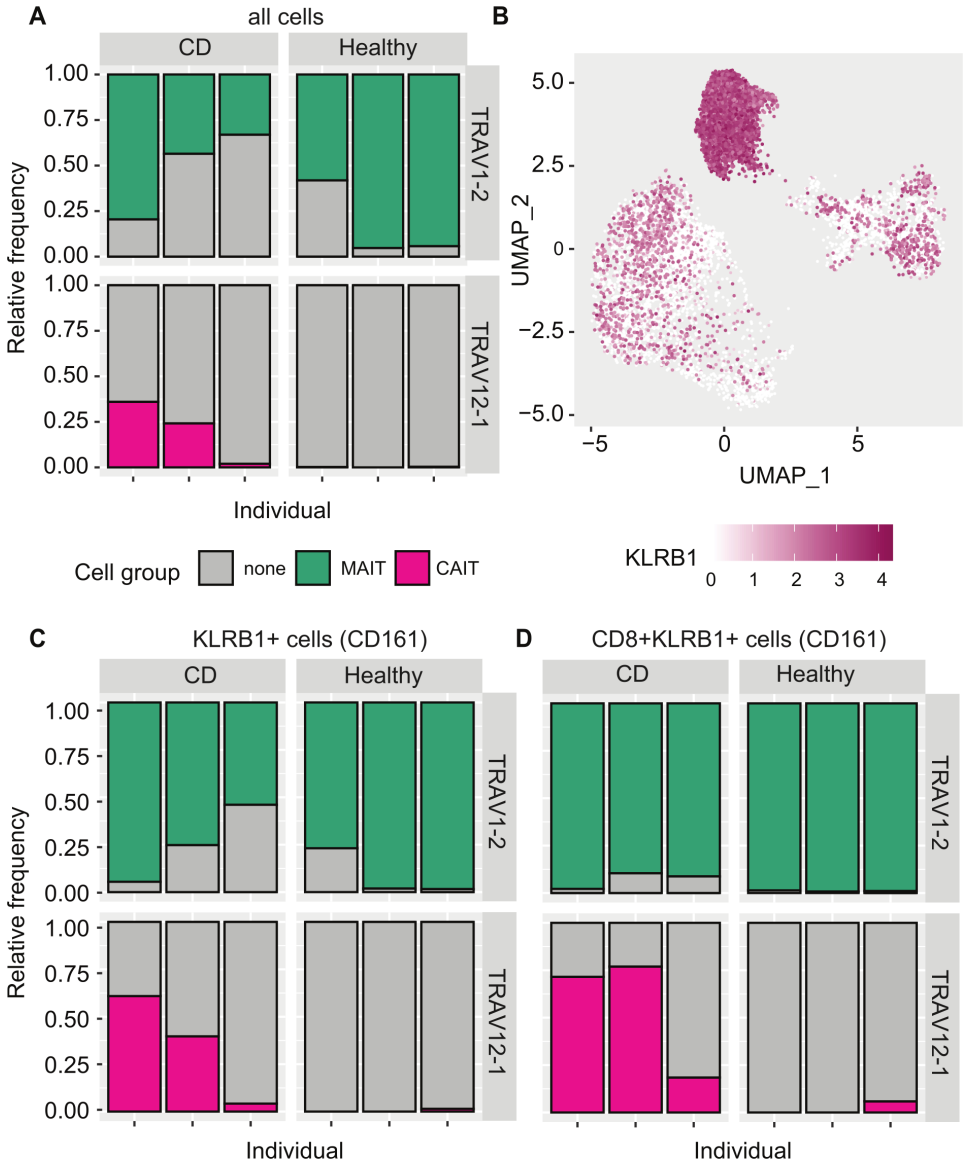


Figure 6 Mucosal associated invariant T (MAIT) and Crohn-associated invariant T (CAIT) cell distribution in TRAV1–2$^+$ and TRAV12–1$^+$ cells the single-cell dataset. TRAV1-2$^+$ and TRAV12-1$^+$ cells were selected from the single-cell dataset in figure 5. (A) Distribution of MAIT and CAIT cells among the three patients with CD and three healthy controls samples for TRAV1-2$^+$ and TRAV12-1$^+$ cells. (B) Distribution of the expression levels of KLRB1. The expression is higher in the same cell groups where MAIT and CAIT cells mostly localise. (C) Distribution of MAIT and CAIT cells in the cell subset expressing KLRB1, encoding for the CD161 protein. (D) Distribution of MAIT and CAIT cells in the CD8$^+$ cell subset expressing KLRB1.

observe correlations between occurrence of the CAIT TCR alpha motif and clinical parameters or other confounding factors.

We observed CAIT cells to be present in both CD4$^+$ and CD8$^+$ fractions, but further characterisation through single-cell GEX of presorted blood memory CD4$^+$ and CD8$^+$ T cell fractions showed higher enrichment in the CD8$^+$ fraction. Single-cell data also revealed the alpha/beta TCR pairs of CAIT cells. Pairing with diverse beta chains was observed, with a preferential usage of the *TRBV7-9* gene.

CAIT cells were mostly found in CD8$^+$ effector memory clusters, with a compatible phenotype to unconventional T cells as MAIT and NKT cells. Innate-like cell markers such as *KLRB1* (CD161), *SLC4A10, CEBPD, IL18R* and *KLRD1* (CD94) were indeed highly expressed by CAIT cells. In contrast to MAIT cells, that are known to be decreased in the blood of patient with CD, CAIT clonotypes were significantly increased in our patients' blood, suggesting a different role of CAIT cells in CD. Despite the high abundance of CAIT clonotypes in certain subjects, the proportion of these cells among individuals varies greatly. Therefore, T cell populations capturable by flow cytometry, based on our current analysis, that is, TRAV12-1$^+$CD161$^+$, are not specific enough to further dissect the immunophenotype and functional profile of CAIT clonotypes. The observed differences of flow-cytometric detection in blood versus intestinal tissue may also relate to a different cell composition of the TRAV12-1$^+$CD161$^+$ compartment, which needs to be investigated in future studies. Also, disease activity and disease location may have an impact on the observed results, as mostly colonic biopsies and patients in disease remission could be obtained for flow cytometric analysis. Thus, although TCR analyses highlighted the presence of CAIT clonotypes in the gut, further analyses stratifying patients not only by disease condition but also by frequency of CAIT cells among CD3$^+$, will be necessary to clarify the role of these clonotypes in the intestinal tissue and in IBD.

Recently, Almeida *et al* published a study characterising a subset of CD1d-reactive NKT type II cells.[19] Apart for MAIT cells, so far only the CD1d-reactive iNKT cells are known to have a restricted TCR alpha repertoire.[15] NKT type II cells are described to have an oligoclonal repertoire.[15] The cells described by Almeida showed indeed a polyclonal TCR repertoire, but with preferential usage of the *TRAV12-1*, *TRAJ6* gene combination, the same observed in CAIT clonotypes. The main GEX marker of NKT type II cells in humans is CD161, thus fitting the phenotype of CAIT cells.[15 18] Moreover, two of the TCR alpha clonotypes therein described fit the CAIT CDR3 motif (figure 1B).[19 36] Given these results, we hypothesise that CAIT cells may be reactive to CD1d and thus be a semi-invariant subgroup of the NKT type II family. Indeed, identification of HLA and antigen specificity is an essential step in T cell characterisation. We did not observe a correlation with a specific classical MHC allele in the patients with high abundance of CAIT cells, suggesting that classical MHC-restriction is indeed not driving the expansion of these cells. This hypothesis was further supported by the fact that CAIT cells were found in both the $CD4^+$ and the $CD8^+$ compartments.

The cells identified by Almeida *et al* are reactive to small lipids produced by the microbiome as well as to PentamethylBenzofuranSulfonates (PBFs). Therefore, CAIT cells may be potentially reactive towards microbiome metabolites or small lipids, which were recently shown to modulate NKT cell responses in the gut.[37]

The study herewith described also has limitations that we want to highlight here. In our study, we were able to control for only some confounding factors such as age, sex, disease activity, smoking behaviour and anti-TNF treatment, while other factors such as medication and disease history may impact on the observed cell population. Moreover, CD is a very heterogeneous disease with high interindividual variability, demanding larger sample sizes than examined here to define larger subgroups for stratified analyses. Additionally, control intestinal samples were available only for non-inflammatory disease controls (CRC) but not healthy individuals, which is another limitation of our study. Finally, technical limitations and interindividual variability do not allow for efficiently capture CAIT cells via flow cytometry. Thus, further studies, including longitudinal sampling strategies, will be necessary to better assess the effect that medications, as well as the transition from active disease to disease remission, may have on CAIT cells.

In summary, we describe a new semi-invariant subgroup of unconventional T cells to be specifically enriched in the blood of patients with CD. Evidences suggest these cells to be NKT type II cells. The consequences of this enrichment remain to be clarified, and CAIT cells immunophenotype and antigen reactivity need to be further elucidated in future studies. Finally, given the semi-invariant nature of CAIT cells, similarly to MAIT and iNKT cells, they are likely to be public in humans at population level, thus constituting an interesting target of study in both health and disease.

MATERIALS AND METHODS

Study design: summary

Because dysregulated T cell reactions against intestinal antigens are considered causal or driving factors for CD, we aimed at identifying potentially existing disease-associated specific T cell clonotypes through the analysis of the TCR repertoire. Identification of disease-associated TCRs may shed light on specific dysregulated T cell responses and may take us a step closer to the identification of potential antigenic triggers of the disease. In our initial analysis, blood samples of patients with CD were selected from the available baseline samples of a prospective registry of patients with CD undergoing treatment with biologics which was established by the German 'Competence Network IBD', Kiel, Germany in 2008. Additional goals included investigation of correlation between the identified disease-associated TCRs[20 38–42] and genotypes[43] and clinical phenotypes of the patients. Therefore, patients included in our initial analysis were characterised by active disease status[44–47] and were known to be responding or non-responding to anti-TNF therapeutic treatment after 6 months from baseline sampling. Age-matched and sex-matched healthy controls were selected from a local registry of blood donors. UC samples were selected from a patient cohort collected at the local hospital to verify the disease specificity of the identified CD-associated TCRs. Finally, 109 patients with CD, 36 patients with UC and 99 healthy controls were included in the analysis. Samples excluded because of failure in library preparation, sequencing or because not passing quality filtering were not included in the manuscript. A second sample collection was analysed to validate the findings in the first collection and to investigate presence and abundance of the identified Crohn-associated clonotypes in the intestinal tissue. Surgical specimens were collected from patients undergoing bowel resection (N=37). Patients with CRC were included as disease controls. It is important to notice that the used intestinal tissue from patients with CRC was not tumour tissue, but adjacent, macroscopically healthy, tissue. T cell subpopulations were sorted from blood and gut of an additional subgroup of patients (N=21) to investigate the cellular population of origin of the identified Crohn associate clonotypes. To investigate in detail the phenotype of the disease-associated cells as well as the TCR alpha/beta pairing, the T cell populations most enriched for these cells, namely $CD4^+$ and $CD8^+$ memory cells, were sorted from freshly collected blood of three recontacted patients with CD of the surgery sample collection and underwent single-cell TCR-seq and RNA-seq.[27 48 49]

Detailed methodologies are described in the online supplemental materials and methods.

Author affiliations
[1]Institute of Clinical Molecular Biology, Christian-Albrechts University of Kiel, Kiel, Schleswig-Holstein, Germany
[2]Institute of Immunology, Christian-Albrechts University of Kiel, Kiel, Schleswig-Holstein, Germany
[3]Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, Russian Federation
[4]Department of Immunology, St Jude Children's Research Hospital, Memphis, Tennessee, USA
[5]Department of Visceral and Thoracic Surgery, Universitatsklinikum Schleswig-Holstein, Kiel, Schleswig-Holstein, Germany
[6]Department of Internal Medicine I, Universitätsklinikum Schleswig-Holstein, Kiel, Schleswig-Holstein, Germany
[7]Department of Dermatology, University Hospital Schleswig Holstein, Kiel, Schleswig-Holstein, Germany
[8]Institute of Epidemiology and Biobank POPGEN, Christian-Albrechts-University of Kiel, Kiel, Germany
[9]Interdisciplinary Crohn Colitis Centre Minden, Minden, Germany
[10]Department of General Surgery, Rostock University Medical Center, Rostock, Mecklenburg-Vorpommern, Germany
[11]CEITEC, Masaryk University, Brno, Czech Republic
[12]Dmitry Rogachev National Research Center of Pediatric Hematology, Moscow, Russian Federation
[13]Center for Precision Genome Editing and Genetic Technologies for Biomedicine, Moscow, Russian Federation

Twitter Elisa Rosati @elisarosix, Gabriela Rios Martini @gabyriosmar, Mikhail V Pogorelyy @pogorely, Anastasia A Minervina @Minervina_Asya, Stefan Schreiber @

medinflame, Philip C Rosenstiel @PhilipCRosensti, Konrad Aden @KonradAden and Andre Franke @an_franke

Acknowledgements We thank Julia Wilking and Christian Röder for their support during the collection of the surgery samples and Prof. Susanne Nicolaus and Dr. Julia Kümpers for their help in recontacting the patients for the single-cell experiments. We thank Nicole Bekel, Berith Messner, Tanja Wesse, Sandra Ussat, Wolfgang Albrecht, Nicole Braun, Maria Eloina Figuera Basso, Yewgenia Dolshawskaya, Melanie Vollstedt, Catharina Von der Lancken, Melanie Schlap-kohl, Janina Fuß and Sören Franzenburg for the technical support during the sample preparation and sequencing procedures.

Contributors AFranke, PB and IZM conceived and coordinated the study. AF is the study guarantor. CJ, KA, FT, WL, SSchreiber and BB helped in collecting the whole blood samples and patient metadata. MH helped in selecting the whole blood samples. AH, CS, CH and J-HE collected the surgery samples and provided patient metadata. MD, ER and PB collected and processed the surgery samples. ER, AFazio, SSS, MAS and AAM isolated RNA and prepared the TCR libraries. FD and MWendorff performed the HLA imputation. ES, GRM and AS performed flow cytometry and sorting for the single-cell experiments. GRM, SSari, PGT, MW, WS and JP collected samples and supported the analyses during peer revision process. ER performed the data analysis with the help of MVP, AAM, JBG, IZM and PB. GB performed the structural data analysis. ER wrote the first draft of the manuscript with support from PB, GRM, AAM, MVP, MH, PCR, IZM and AFranke. All authors read and approved the final version of the manuscript.

Funding This work was supported by the EU's Horizon 2020 SYSCID program under the grant agree-ment No 733100, the DFG Excellence Cluster EXC306 'Inflammation at Interfaces' and Pre-cision Medicine in Chronic Inflammation (EXC2167-390884018), the DFG Research Training Group 1743 [RTG1743], RFBR 19-54-12011 grant, grant from the Ministry of Science and Higher Education of Russian Federation (075-15-2019-1789), the DFG grant n. 4096610003, the RU5042-miTarget and the European Crohn's and Colitis Organisation (ECCO).

Competing interests ER, AF, PB and IZM hold a pending patent application based on the results described in this manuscript (EP21217058.3).

Patient consent for publication Not applicable.

Ethics approval The study has been approved by the local ethics committees (University of Kiel). For the whole blood collection, ethics votes: A156/03 v. 29.7.2010, A156/03 - 3/15 v. 21.1.16, A156/03-2/13, A161/08. For the surgery collection: samples were collected in collaboration with the Clinic for General and Thoracic Surgery of the UKSH. Ethics vote: D553/16. All study participants signed a written informed consent. Participants gave informed consent to participate in the study before taking part.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data are available upon reasonable request. Data may be obtained from a third party and are not publicly available. Raw sequencing data of bulk TCR repertoire profiling are available on the ENA database with study accession number PRJEB50045. Single-cell gene expression and TCR processed data are available on the FastGenomics database (Seurat_objects_Rosati_TCR_IBD). Raw single cell data (fastq files) are available from the corresponding author upon request. HLA genetic data are available from the corresponding author upon reasonable request. Metadata from the whole blood TCR collection are available through the Popgen biobank.

ORCID iDs
Elisa Rosati http://orcid.org/0000-0002-2635-6422
Matthias Hübenthal http://orcid.org/0000-0002-5956-3006
Philip C Rosenstiel http://orcid.org/0000-0002-9692-8828
Andre Franke http://orcid.org/0000-0003-1530-5811

REFERENCES

1 Kostic AD, Xavier RJ, Gevers D. The microbiome in inflammatory bowel disease: current status and the future ahead. *Gastroenterology* 2014;146:1489–99.
2 Klag T, Stange EF, Wehkamp J. Defective antibacterial barrier in inflammatory bowel disease. *Dig Dis* 2013;31:310–6.
3 Ciccarelli O, Barkhof F, Bodini B, *et al*. Pathogenesis of multiple sclerosis: insights from molecular and metabolic imaging. *Lancet Neurol* 2014;13:807–22.
4 Ilonen J, Lempainen J, Veijola R. The heterogeneous pathogenesis of type 1 diabetes mellitus. *Nat Rev Endocrinol* 2019;15:635–50.
5 Ray K. Ibd: genotypes and phenotypes of IBD. *Nat Rev Gastroenterol Hepatol* 2015;12:672.
6 Kaser A, Zeissig S, Blumberg RS. Inflammatory bowel disease. *Annu Rev Immunol* 2010;28:573–621.
7 Imam T, Park S, Kaplan MH, *et al*. Effector T helper cell subsets in inflammatory bowel diseases. *Front Immunol* 2018;9:1212.
8 Rosati E, Dowds CM, Liaskou E, *et al*. Overview of methodologies for T-cell receptor repertoire analysis. *BMC Biotechnol* 2017;17:61.
9 Doorenspleet ME, Westera L, Peters CP, *et al*. Profoundly Expanded T-cell Clones in the Inflamed and Uninflamed Intestine of Patients With Crohn's Disease. *Journal of Crohn's and Colitis* 2017;11:831–9.
10 Wu J, Pendegraft AH, Byrne-Steele M, *et al*. Expanded TCRβ CDR3 clonotypes distinguish Crohn's disease and ulcerative colitis patients. *Mucosal Immunol* 2018;11:1487–95.
11 Chapman CG, Yamaguchi R, Tamura K, *et al*. Characterization of T-cell receptor repertoire in inflamed tissues of patients with Crohn's disease through deep sequencing. *Inflamm Bowel Dis* 2016;22:1275–85.
12 Allez M, Auzolle C, Ngollo M, *et al*. T cell clonal expansions in ileal Crohn's disease are associated with smoking behaviour and postoperative recurrence. *Gut* 2019;68:1961–70.
13 Treiner E, Duban L, Bahram S, *et al*. Selection of evolutionarily conserved mucosal-associated invariant T cells by MR1. *Nature* 2003;422:164–9.
14 Greenaway HY, Ng B, Price DA, *et al*. NKT and MAIT invariant TCRα sequences can be produced efficiently by VJ gene recombination. *Immunobiology* 2013;218:213–24.
15 Godfrey DI, Uldrich AP, McCluskey J, *et al*. The burgeoning family of unconventional T cells. *Nat Immunol* 2015;16:1114–23.
16 Kaminski H, Couzi L, Eberl M. Unconventional T cells and kidney disease. *Nat Rev Nephrol* 2021;17:795–813.
17 Serriari N-E, Eoche M, Lamotte L, *et al*. Innate mucosal-associated invariant T (MAIT) cells are activated in inflammatory bowel diseases. *Clin Exp Immunol* 2014;176:266–74.
18 Liao C-M, Zimmer MI, Wang C-R. The functions of type I and type II natural killer T cells in inflammatory bowel diseases. *Inflamm Bowel Dis* 2013;19:1330–8.
19 Almeida CF, Smith DGM, Cheng T-Y, *et al*. Benzofuran sulfonates and small self-lipid antigens activate type II NKT cells via CD1d. *Proc Natl Acad Sci U S A* 2021;118:e2104420118.
20 Pogorelyy MV, Elhanati Y, Marcou Q, *et al*. Persisting fetal clonotypes influence the structure and overlap of adult human T cell receptor repertoires. *PLoS Comput Biol* 2017;13:e1005572.
21 Hegazy AN, West NR, Stubbington MJT, *et al*. Circulating and Tissue-Resident $CD4^+$ T Cells With Reactivity to Intestinal Microbiota Are Abundant in Healthy Individuals and Function Is Altered During Inflammation. *Gastroenterology* 2017;153:1320–37.
22 Rosati E, Pogorelyy MV, Dowds CM, *et al*. Identification of disease-associated traits and Clonotypes in the T cell receptor repertoire of monozygotic twins affected by inflammatory bowel diseases. *J Crohns Colitis* 2020;14:778–90.
23 Goyette P, Boucher G, Mallon D, *et al*. High-Density mapping of the MHC identifies a shared role for HLA-DRB1*01:03 in inflammatory bowel diseases and heterozygous advantage in ulcerative colitis. *Nat Genet* 2015;47:172–9.
24 Goyette P, Boucher G, Mallon D, *et al*. High-Density mapping of the MHC identifies a shared role for HLA-DRB1*01:03 in inflammatory bowel diseases and heterozygous advantage in ulcerative colitis. *Nat Genet* 2015;47:172–9.
25 Degenhardt F, Wendorff M, Wittig M, *et al*. Construction and benchmarking of a multi-ethnic reference panel for the imputation of HLA class I and II alleles. *Hum Mol Genet* 2019;28:2078–92.
26 Degenhardt F, Mayr G, Wendorff M, *et al*. Transethnic analysis of the human leukocyte antigen region for ulcerative colitis reveals not only shared but also ethnicity-specific disease associations. *Hum Mol Genet* 2021;30:356–69.
27 Butler A, Hoffman P, Smibert P, *et al*. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat Biotechnol* 2018;36:411–20.
28 Godfrey DI, Koay H-F, McCluskey J, *et al*. The biology and functional importance of MAIT cells. *Nat Immunol* 2019;20:1110–28.
29 De Biasi S, Gibellini L, Lo Tartaro D, *et al*. Circulating mucosal-associated invariant T cells identify patients responding to anti-PD-1 therapy. *Nat Commun* 2021;12:1669.
30 Zhao J, Zhang S, Liu Y, *et al*. Single-Cell RNA sequencing reveals the heterogeneity of liver-resident immune cells in human. *Cell Discov* 2020;6:22.
31 Park D, Kim HG, Kim M, *et al*. Differences in the molecular signatures of mucosal-associated invariant T cells and conventional T cells. *Sci Rep* 2019;9:7094.

32 Lepore M, Kalinichenko A, Kalinicenko A, *et al*. Parallel T-cell cloning and deep sequencing of human MAIT cells reveal stable oligoclonal TCRβ repertoire. *Nat Commun* 2014;5:3866.
33 Lepore M, Lewinsohn DA, Lewinsohn DM. T cell receptor diversity, specificity and promiscuity of functionally heterogeneous human MR1-restricted T cells. *Mol Immunol* 2021;130:64–8.
34 Park H-J, Shin MS, Kim M, *et al*. Transcriptomic analysis of human IL-7 receptor alpha ^{low} and ^{high} effector memory CD8^{+} T cells reveals an age-associated signature linked to influenza vaccine response in older adults. *Aging Cell* 2019;18:e12960.
35 Pereira BI, De Maeyer RPH, Covre LP, *et al*. Sestrins induce natural killer function in senescent-like CD8^{+} T cells. *Nat Immunol* 2020;21:684–94.
36 Van Rhijn I, Young DC, Im JS, *et al*. Cd1D-Restricted T cell activation by nonlipidic small molecules. *Proc Natl Acad Sci U S A* 2004;101:13578–83.
37 Oh SF, Praveena T, Song H, *et al*. Host immunomodulatory lipids created by symbionts from dietary amino acids. *Nature* 2021;600:302–7.
38 Shugay M, Britanova OV, Merzlyak EM, *et al*. Towards error-free profiling of immune repertoires. *Nat Methods* 2014;11:653–5.
39 Bolotin DA, Poslavsky S, Mitrophanov I, *et al*. MiXCR: software for comprehensive adaptive immunity profiling. *Nat Methods* 2015;12:380–1.
40 Benjamini Y, Hochberg Y. *Royal statistical Society*. . Wiley, 1995: 57. 289–300.
41 Csardi Gabor NT. The igraph software package for complex network research. *InterJournal, Complex Systems* 2006;5:1–9.
42 Gowthaman R, Pierce BG. TCRmodel: high resolution modeling of T cell receptors from sequence. *Nucleic Acids Res* 2018;46:W396–401.
43 Zheng X, Shen J, Cox C, *et al*. HIBAG--HLA genotype imputation with attribute bagging. *Pharmacogenomics J* 2014;14:192–200.
44 Harvey RF, Bradshaw JM. A simple index of crohn's-disease activity. *The Lancet* 1980;315:514.
45 Waniek S, di Giuseppe R, Plachta-Danielzik S, *et al*. Association of vitamin E levels with metabolic syndrome, and MRI-Derived body fat volumes and liver fat content. *Nutrients* 2017;9:1143.
46 Koch M, Borggrefe J, Barbaresko J, *et al*. Dietary patterns associated with magnetic resonance imaging-determined liver fat content in a general population study. *Am J Clin Nutr* 2014;99:369–77.
47 Nöthlings U, Krawczak M. [PopGen. A population-based biobank with prospective follow-up of a control group]. *Bundesgesundheitsblatt Gesundheitsforschung Gesundheitsschutz* 2012;55:831–5.
48 Korsunsky I, Millard N, Fan J, *et al*. Fast, sensitive and accurate integration of single-cell data with harmony. *Nat Methods* 2019;16:1289–96.
49 Finak G, McDavid A, Yajima M, *et al*. Mast: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biol* 2015;16:278.

SUPPLEMENTARY MATERIALS

Supplementary Material and Methods

Supplementary Table S1: summary of performed analyses and used sample collections

Dataset	Sample type	Analysis	Disease group	Number of individuals
1 whole blood	PAXgene tubes	Bulk TCR repertoire + HLA imputation	CD	109
			UC	36
			Healthy	99
2.1 surgery	PBMCs + Gut biopses + Mesenteric lymph nodes	Bulk TCR repertoire	CD	11
			UC	13
			CRC	13
2.2 surgery sorted	Sorted T cells from PBMCs + Sorted T cells from gut	Bulk TCR repertoire	CD	7
			UC	5
			CRC	9
3 single-cell	Sorted T cells from PBMCs	Single-cell RNA and TCR	CD	3
			Healthy	3

Whole blood sample collection

Phenotypic details for this collection are described in **Supplementary Table S2**.

CD patient collection: BioCrohn is a non-interventional prospective registry of Crohn's disease patients undergoing biologics treatment which was established by the German "Competence Network IBD", Kiel, Germany in 2008. Samples included in this study were collected prior to the start of anti-TNF therapy (Infliximab).

To evaluate disease activity, we use the Harvey-Bradshaw Index (HBI). We employ this measure to distinguish between inactive ($0 \leq HBI \leq 4$: remission) and active state of the disease ($5 \leq HBI \leq 7$: mild activity, $8 \leq HBI \leq 16$: moderate activity, $17 \leq HBI \leq \infty$: severe activity)(*1*), as described from the Canadian society of intestinal research. In total, whole blood samples of 109 CD patients were available for analysis. All patients had active disease at sampling time. Response to anti-TNF therapy was defined as a steroid-independent change from active disease (HBI > 4 at baseline) to an inactive state of the disease (HBI < 4 at 6 months after start of therapy). Accordingly, we defined non-response to anti-TNF therapy as the disease being consistently active at both time points under investigation (HBI > 4 at baseline and 6 months).

UC patient collection: blood samples from UC patients of the "whole blood collection" were recruited as part of a larger panel of patients with IBD, recruited in the Comprehensive Centre for Inflammation Medicine (CCIM), at the University Hospital Schleswig-Holstein, Campus Kiel (Germany). Samples

for the present study were randomly selected among the available samples stored in the PopGen biobank. In total, whole blood samples of 36 UC patients were available.

Healthy control collection: blood samples were obtained from the PopGen control sample, consisting of a community-based sample from the Kiel area (Northern Germany) and of blood donors from the University Hospital Schleswig-Holstein, Campus Kiel(*2*). In total, 1316 individuals were recruited at baseline (between 2005 and 2007) and were re-examined at the second examination cycle; between 2010 and 2012; n=952)(*3, 4*). Participants filled-in questionnaires on health and nutrition, received a basic clinical examination and provided blood samples for research. For the present analyses, individuals with major diseases such as cancer, neurological and inflammatory diseases were excluded. In total, 99 whole blood samples were available. These samples were matched to the CD patients regarding age and sex.

2.5 ml of blood were drawn and stored in PAXgene tubes (Qiagen), following the manufacturer's instructions. Samples were stored at -80°C. Total RNA was isolated using the PAXgene blood miRNA kit (Qiagen) with automatic isolation using the Qiacube machine. 1,200 ng of total RNA were used for TCR library preparation.

Supplementary Table S2: Whole blood collection phenotype information. Information about disease phenotype was provided via the Montreal classification. Information on the disease activity was available through the Harvey-Bradshaw index (HBI) for CD patients and though the colitis activity index (CAI) and partial Mayo score for UC patients. Information about steroid therapy at sampling time was also provided.

	Sample group		
	CD	**UC**	**Healthy**
Basic phenotype			
Samples	109	36	99
Gender, Male/Female (%)	43/65 (40/60%)	21/15 (58/42%)	41/58 (41/59%)
Age, mean±std	40±12	37±13	48±13
Montreal classification			
Age at diagnosis in years			
A1, <16) (%)	7 (6%)	–	–
A2, 17-40 (%)	74 (70%)	–	–
A3, >40 (%)	25 (24%)	–	–
Behaviour			
B1, inflammatory (%)	51 (47%)	–	–
B2, stricturing (%)	21 (19%)	–	–
B3, penetrating (%)	36 (34%)	–	–
Location			
L1, ileal (%)	29 (27%)	–	–
L2, colonic (%)	23 (21%)	–	–
L3, ileocolonic (%)	56 (52%)	–	–
Disease extent			
E1, proctitis (%)	–	3 (3%)	–
E2, left-sided UC (%)	–	16 (47%)	–
E3, extensive UC (%)	–	15 (44%)	–
Info not available (NA)	1	2	–
Disease activity			
HBI, mean±std	11±5	–	–
Mild activity, 5≤HBI≤7 (%)	37 (34%)	–	–
Moderate activity, 8≤HBI≤16 (%)	57 (53%)	–	–
Severe activity, 17≤HBI≤∞ (%)	14 (13%)	–	–
Partial Mayo score, mean±std	–	3.3±2	–
CAI, mean±std	–	6.7±4.3	–
Info not available (NA)	1	0	–
Medications			
Steroids, yes/no (%)	7/102 (6/94%)	14/22 (39/61%)	–
Treatment naïve yes/no/NA (%)	77/14/18 (70/13/17%)	0/0/36 (0/0/100%)	–
Responders to anti-TNF yes/no/NA (%)	46/60/3 (42/55/3%)	–	–

Surgery sample collection

Samples were collected in collaboration with the *Clinic for General and Thoracic Surgery* of the University Hospital Schleswig-Holstein, Kiel, Germany. The study was approved by the ethics committee of the medical faculty of the Christian-Albrecht University Kiel, Kiel, Germany (ethics vote: D553/16). A total of 37 patients were recruited (11 CD, 13 UC, 13 CRC). It is important to notice that for CRC tissue samples, macroscopically normal, non-tumour tissue was utilized. Sample details are provided in **Supplementary Table S5**. Information on disease activity scores or treatment with biologics prior to operation was unfortunately not available. However, we did not see a correlation with these parameters in the whole blood collection.

Sample collection included **(1)** one 9 ml whole blood EDTA tube from which PBMCs were isolated by density gradient centrifugation (Ficoll, GE Healthcare) and **(2)** resected intestinal mucosa. Tissues were cut into pieces of approximately 30–50 mg, snap-frozen and stored at -80°C. Details on the number of collected samples and intestinal resection location for each patient are available in **Supplementary Table 3**, together with patients' basic phenotypic information.

Total RNA was isolated using the RNeasy mini kit from Qiagen following the manufacturer's instructions. Up to 3,200 ng of total RNA were used for TCR library preparation.

Supplementary Table S5: Surgery collection phenotype information. Information about the numbers of collected samples for each tissue is provided together with information about intestinal location of the sample and about therapy with steroids or immunosuppressants at time of surgery. NA: no information available.

	CD	UC	CRC
Basic phenotype			
Samples	11	13	13
Gender, Male/Female (%)	4/7 (36/64%)	7/6 (54/46%)	8/5 (62/38%)
Age, mean±std	40±15	52±16	69±12
Sample types			
Intestinal samples	11	13	9
Mesenteric lymph nodes	4	5	0
PBMCs	11	12	12
Intestine + PBMCs datasets	11	12	8
Resected tissue location			
Large intestine	4	13	13
Small intestine	7	0	0
Medications			
Steroids, yes/no/NA	4/5/2	7/6/0	0/0/13
Immunosuppressants, yes/no/NA	5/5/1	2/10/1	0/0/13

Surgery sample collection: sorted T cell populations

Additional 21 patients (9 CRC, 7 CD, 5 UC) were recruited for the surgery collection (**Supplementary Table S6**). PBMCs were isolated from peripheral EDTA blood samples by density gradient centrifugation (Ficoll, GE Healthcare).

To generate single cell suspensions from intestinal mucosa, resected tissue was washed twice in sterile PBS. Mucosa was separated from the muscle layer and fat and cut into small pieces. Mucosa pieces were digested in digestion buffer containing 1xHBSS+calcium+magnesium buffer (Gibco) supplemented with 2% human AB serum (Sigma Aldrich), 100 IU/ml penicillin, 100 µg/ml streptomycin, 0.25 µg/ml amphotericin B (Antibiotic Antimycotic Solution, Sigma Aldrich), 1mg/ml Collagenase NB (Nordmark Biochemicals) and 200µg/ml DNAse I (AppliChem) for 75min at 37°C under continuous rotation at a ratio 3ml/g of tissue. After digestion, the supernatant was filtered using a 100µm filter, followed by a 70µm filter. Remaining tissue pieces were squeezed through a sterile sieve, filtered and combined with the supernatant. Cells were pelleted by centrifugation for 10min at 500xg, followed by a red blood cell lysis (Red Blood Cell Lysis Solution, Miltenyi Biotec).

Supplementary Table S6: Sorted surgery collection phenotype information. Information about the numbers of collected samples for each tissue is provided together with information about intestinal location of the sample.

	Sample group		
	CD	**UC**	**CRC**
Basic phenotype			
Samples	7	5	9
Gender, M/F (%)	3/4 (43/57%)	1/4 (25/75%)	8/1 (89/11%)
Age, mean±std	39±11	42±17	69±17
Sample types			
Intestine + PBMCs datasets	6	5	9
Resected tissue location			
Large intestine	2	5	8
Small intestine	5	0	1

To isolate different T cell populations, $CD4^+$ T cells were isolated by MACS (CD4 MicroBeads, Miltenyi Biotec) and $CD4^+$ and $CD4^-$ cells were stained with the following antibodies according to manufacturers' protocols:

$CD4^+$ cells: CD3-AlexaFluor700 (OKT-3), CD4-BV421 (OKT-4) (both Biolegend), CD45RO-FITC (UCHL1), CCR7-PE-Vio770 (REA108), CD25-PE (REA613), CD127-APC (REA614), CD8-PerCP (BW135/80), CD14-PerCP (Tük4), CD20-PerCP (Tük4) (all MiltenyiBiotec).

$CD4^-$ cells: CD3-PE (BW264/56), CD8-APC (BW135/80), CD14-PerCP (Tük4), CD20-PerCP (Tük4) (all MiltenyiBiotec). Propidiumiodide was used to exclude dead cells. Cells were sorted on a BD FACS ARIA into the following populations:

$CD4^+$ T memory ($CD3^+CD4^+CD45RO^+CD127^+CD25^-CD8^-CD14^-CD20^-$)

$CD4^+$ T naïve ($CD3^+CD4^+CD45RO^-CCR7^+CD127^+CD25^-CD8^-CD14^-CD20^-$)

$CD4^+$ Treg ($CD3^+CD4^+CD45RO^+CD127^-CD25^+CD8^-CD14^-CD20^-$)

$CD8^+$ T cells ($CD3^+CD8^+CD14^-CD20^-$)

Total RNA was isolated using Trizol and was used for TCR library preparation.

Genotyping & HLA imputation

DNA extraction was performed by the DNA laboratory of the Institute of Clinical Molecular Biology (Christian-Albrechts-University of Kiel, Germany) using a Chemagic 360 from PerkinElmer (Waltham, Massachusetts, U.S.) with the low volume kit cmg 1491 and the buffy coat kit cmg-714 (Chemagen, Baesweiler, Germany) according to the manufacturer's protocol. Genotyping was conducted by the Institute of Clinical Molecular Biology's DNA Laboratory and Genotyping Core Facilities, employing Illumina's (Illumina Inc., San Diego, U.S.) Global Screening Array-24 Multi Disease (GSA) Version 1.0 A1 (whole blood collection: CD cases) or at the at the Regeneron Genetics Center using the same array (whole blood collection: healthy controls). The GSA chip highly covers the HLA region, with 9,876 variants on the extended HLA on chromosome 6 from 25Mb to 34Mb.Due to the small sample size numbers, we did not perform a special QC on the variants in the HLA region. The overall genotyping rate in both studies was > 98%. Based on the SNP data we excluded 2 individuals not matching to a European ancestry, none of the individuals in the case or control cohort were related with an identity-by-descent > 0.185, all recorded phenotypic and genotypic gender information matched. HLA genotypes were imputed at all classical HLA class I (HLA-*A, -B, -C)* and HLA class II (HLA-*DRB1/3/4/5*, -*DQA1*, -*DQB1*, *DPA1*, -*DPB1*) loci at full context 4-digit resolution using our previously published in-house trans-ethnic HLA imputation reference(*5*) based on the Illumina Immunochip, which we retrained using only the positions present on both the Illumina Immunochip and GSA Version 2.0 B1 which is highly similar to the content of the GSA Version 1.0 A1, and HIBAG(*6*) (https://hibag.s3.amazonaws.com/download/IKMB_HMG/index.html). Overall, 98% of all SNPs per locus out of 4,574 SNPs present in the HLA reference dataset overlapped between the reference and the GSA data.

SNPs not matching between the reference and the respective imputed dataset (based on their alleles) were excluded, ATCG variants with a MAF frequency of >40% or those not matching the reference (+-strand annotation assumed for both) were deleted from the dataset. Post imputation data were not cleaned according to any imputation threshold. However, measures such as the marginal probability of alleles(*7*) and similarity of alleles based on the SNP overlap between classifiers(*5*) were taken into consideration for the interpretation of results.

TCR library preparation and sequencing

Starting from total RNA, molecular-barcoded TCR cDNA libraries were prepared as previously described(*8*), with minor modifications for both TCRα and TCRβ chains. Briefly, cDNA synthesis was

performed using SMARTScribe reverse transcriptase (Clontech, Takara) using primers for the TCRα and TCRβ constant region. A unique molecular identifier (UMI), and a sample barcode of 6 nucleotides, were introduced via template-switching. cDNA synthesis was carried out for 60 minutes at 42°C. cDNA was then treated with Uracil DNA-Glycosylase (UDG, from New England Biolabs) and incubated for 30 min at 37°C. Samples were subsequently purified with the QIAquick PCR purification kit (Qiagen) and eluted in 50 or 100 µl deionized water. Purified cDNA was then amplified with 2 consecutive PCRs, respectively 18 and 12 cycles, with purification after each PCR using MagSi-NGSprep Plus (MagnaMedics). Illumina compatible adapters and sample-specific barcodes were added during the second PCR. Quality and concentration of the libraries were measured with TapeStation D1000 (Agilent) and Qubit (Thermofisher). Libraries were pooled using 5 ng per library and sequenced on Illumina HiSeq2500 with a single-index Rapid Run of 2×100 bp (whole blood) or on Illumina NovaSeq 6000 SP 2×150 bp flow cell (surgery). Custom sequencing primers were added to the Illumina primers. For the sorted surgery collection samples, libraries were prepared using human TCR profiling kit (MiLaboratory), according to the manufacturers' protocol. Libraries were sequenced on Illumina MiSeq 2×150 bp or NovaSeq 6000 SP 2×150 bp.

TCR data pre-processing

PCR and sequencing error correction were performed through identification and selection of unique molecular identifiers using the software MiGEC(*9*), version 1.2.6. Filtered sequences were aligned on a TCR gene reference, clonotypes were identified, grouped and CDR3 sequence was identified using the software MiXCR(*10*), version 2.1.1 (first sample collection: blood collection) or version 3.0.14 (second collection: surgery collection). Clonotype tables containing clonotype counts, frequencies, CDR3 nucleotide and amino acid sequences and V(D)J genes were obtained and used for further analysis.

Summary data on all analysed samples, including the total number of identified clonotypes per sample, are available in **Supplementary Table S3A**. A total of 444 samples were analysed for bulk TCR repertoire.

TCR data analysis

For statistical analysis, the Wilcoxon-Mann-Whitney U test was used unless otherwise indicated. Samples with less than 200 identified unique clonotypes were excluded from the analysis. TCRs found at a count inferior to 2 (singletons) or encoding for out of frame sequences or TCRs containing stop codons, were also excluded from further analyses.

Proportion analysis of TCR groups (CDR3length_Vgene_Jgene): For each disease group (CD, UC, healthy), the total number of TCR sequences being part of each TCR group (CDR3length_Vgene_Jgene) was calculated and divided by the number of total TCR sequences present

in each disease group in order to normalize for differences in TCR number among disease groups. The proportion of TCRs of each TCR group was then compared between disease groups, either CD against UC, or against healthy controls, for both TCR alpha and TCR beta.

Fisher test on identified TCR groups of interest: using specific CDR3 and VJ gene combinations identified through the analysis of repertoire proportions (12aa_TRAV1-2_TRAJ33, 15aa_TRAV12-1_TRAJ6), we performed one-sided Fisher's exact test on the single TCRs of these groups to identify TCR alpha clonotypes respectively decreased (12aa_TRAV1-2_TRAJ33) or increased (15aa_TRAV12-1_TRAJ6) in CD patients versus healthy individuals, from the whole blood and surgery collections. We applied multiple testing correction through the Benjamini-Hochberg (BH) method(*11*), in order to control expected proportion of false discoveries amongst the rejected hypotheses (false discovery rate, FDR).

Analysis on CAIT sequences: TCR sequences from the identified group of CDR3length = 15 amino acids, V gene = TRAV12-1 and J gene = TRAJ6 that were found to be more abundant in CD patients as compared to controls in the whole blood sample collection were used for the logo plot which defined the CAIT clonotype motif CVV**A*GGSYIPTF. All sequences of the selected TCR group, carrying the CAIT motif were selected for further analysis. For each sample, the number of CAIT sequences present in the sample as well as the sum of relative abundances (cumulative abundance) of CAIT sequences were calculated.

For the network plots, CAIT sequences of each sample were used. Their relative abundance was divided in discrete groups as indicated in the figure legends and defined the size of the vertexes of the network. Each sample was represented by nodes connected by edges. Separated clusters represent distinct samples/individuals. Figure 2E was generate using the R igraph package(*12*).

Association of CAIT clonotypes with phenotype and genotype traits: A generalized linear model was fitted using as dependent variable either the cumulative abundance or the number of CAIT clonotypes in each sample and as independent variables the clinical parameters which were, for CD patients: age (as from the Montreal classification), sex, smoking behaviour, disease location (as from the Montreal classification), disease activity (based on HBI), prediction response to anti-TNF therapy, and for healthy controls: age, sex and smoking behaviour. A similar model was fitted for each analysed HLA allele using as independent covariates the dosage of each allele, where 0 indicated absence, 1 heterozygosity and 2 homozygosity for the analysed allele. The used formulas for each model are described in **Supplementary Table S4**.

TCR sequence and structural analysis

The protein structural model of the exemplary CAIT TCR (TRAV12-1- CVVNLASGGSYIPTF - TRAJ6- / TRBV7-9 – CASSTRELANTIYF - TRBJ1-3) was constructed using the TCRmodel webserver(*13*) and visualized using PyMOL.

Single-cell analysis

Blood was drawn from three CD patients, 2 females and 1 male of age 29-, 49- and 36-year-old respectively, who were shown to have high abundance of CAIT TCRs during bulk TCR analysis, as well as from three age and sex matched healthy controls. PBMCs were isolated by gradient centrifugation from EDTA blood tubes. CD3+ cells were enriched by MACS (Miltenyi Biotec).
For single-cell analysis, $CD3^+$ (VioBle) $CD45RO^+$ (PeVio770) memory T cells were sorted in $CD8^+$ (VioGreen) and $CD4^+$ (APCVio770) subpopulations. Cells were removed from the sorting chamber into pre-coated low-bind collection tubes and centrifuged for 5 min at 400x g, 4°C. Cells were re-suspended in PBS plus 0,04% BSA to a concentration of 1000 cells/ul.

Single-cell suspensions were loaded on a Chromium Chip G (10x Genomics) according to manufacturer's instructions for processing with the Chromium Next GEM Single Cell 5' Library and Gel Bead Kit v1.1. Depending on the number of cells available for each patient, a maximum of 20,000 cells were loaded for each reaction. TCR single-cell libraries were subsequently prepared from the same cells with the Chromium Single Cell V(D)J Enrichment Kit, Human T Cell. Libraries were sequenced on the Illumina NovaSeq 6000 machine with 2×100 bp for gene expression, aiming for 50,000 reads per cell and 2×150 bp and 5000 reads per cell for TCR libraries.

Single-cell T cell receptor repertoire clonotype tables were generated using the VDJ command of the Cellranger software, version 3.1.0 from 10xGenomics and using the VDJ reference version 2.0.0. Clonotype tables were filtered to include only cells which passed quality filtering in the gene expression analysis. In addition, clonotypes were stringently filtered for possible doublets by removing clonotypes (i) found in 1 cell only and containing more than 1 TCR alpha and 1 TCR beta (ii) containing more than 1 TCR alpha and no TCR beta sequence (iii) containing more than 1 TCR beta and no TCR alpha sequence (iv) containing more than 2 TCR alpha or more than 2 TCR beta sequences.

Gene expression matrices were generated through the COUNT command of Cellranger v3.1.0. from 10xGenomics and using the reference GRCh38 version 3.0. Data were analyzed using the Seurat v3.2.3 R package(*14*). Cells with more than 400 but less than 3500 detected RNA features per cell and less than 8% mitochondrial RNA were retained for further analysis. Only genes present in at least 1% of the cells were considered. Also, only cells with a detected TCR were used. TCR VDJ genes were removed from gene expression counts to allow for an unbiased analysis. Data from multiple individuals were merged and batch effect correction by experimental day and sequencing run was performed using the Harmony package(*15*). 20 dimensions were used for performing principal component analysis (PCA) and uniform manifold approximation and projection (UMAP) in 3 dimensions. Clusters were identified using 0.4 resolution. Positive and negative marker genes were identified using Seurat FindMarkers function and the MAST method(*16*), considering only genes found in at least 25% of cells of each cluster.

MAIT cells were identified by their TCR alpha sequences (12 aa CDR3, TRAV1-2, TRAJ33/12/20), as well as iNKT cells (15aa CDR3, TRAV10, TRAJ18), while CAIT cells were defined from TCR alpha of 15aa, TRAV12-1, TRAJ6 and CDR3 motif CVV**A*GGSYIPTF. Proportion of CAIT cells carrying certain TRBV genes was calculated based on the number of unique CAIT clonotypes identified, independently from the number of cells carrying each clonotypes and originated by the same original clonotype by clonal expansion.

Immunophenotyping by flow cytometry

Human mononuclear cells were isolated from whole blood by standard density gradient centrifugation (Biocoll, Biochrom, Berlin Germany). Intestinal tissue was incubated in HBSS media with 20uL/mL DNAse, 1mg/mL collagenase, 1% penicillin/streptomycin and 0,2% human AB-serum for 30 min in a shaking incubator at 37°C. Biopsies were then dissociated by mechanical pressure and strained through a 70µm filter. Cells were washed with RPMI medium. Cells from blood and tissue freshly after isolation were stained on the surface using an antibody cocktail. All flow cytometry plots are gated on CD14-, CD20-, CD3+ lymphocytes after dead cell and doublet exclusion.

Whole blood and intestinal tissue biopsy sample collection for flow cytometry immunophenotyping

Crohn's disease and Ulcerative colitis patients: Samples were collected in collaboration with the *Comprehensive Center for Inflammation Medicine (CCIM)* of the University Hospital Schleswig-Holstein, Kiel, Germany. A total of 25 CD patients and 13 UC patients were recruited (including 2 from the '*Whole blood collection used for TCR repertoire sequencing').* Patient information is provided in **Supplementary Table S8**. An additional collection of patients with matching peripheral blood and intestinal biopsy samples was collected, 14 CD patients and 6 UC patients (**Supplementary Table S9**). Therefore, in total we analysed blood samples for 39 CD patients (25 + 14) and 19 UC patients (13 + 6).

Colon cancer patients as disease controls: Fifteen biopsy samples from healthy tissue of colorectal carcinoma patients (CRC) were obtained in collaboration with the department of general surgery of the University Medicine Rostock, Germany. Sample details are provided in **Supplementary Table S9.**

Healthy human blood: Samples were collected from 18 volunteer workers at the University Hospital Schleswig-Holstein, Kiel, Germany (**Supplementary Table S8**).

The study was approved by the ethics committee of the medical faculty of the Christian-Albrecht University of Kiel, Kiel, Germany. All study participants signed a written consent.

Human mononuclear cell staining for flow cytometry immunophenotyping

Mononuclear cells were washed with PBS + 2mM EDTA + 0.5% BSA (PEB buffer) and then stained with a cell surface antibody cocktail for CD3 -Alexa Fluor 700 (OKT3, Biolegend), CD4 APC-Vio700

(M-T466, MiltenyiBiotech), CD8 VioGreen (REA734, MiltenyiBiotech), CD14 PerCP (TÜK4, MiltenyiBiotech), CD20 PerCP (LT20, MiltenyiBiotech), CD161 Brilliant Violet 421 (DX12, BD Biosciences), IL-18Rα -APC (REA1095, MiltenyiBiotech), IL-18Rα -PE (REA1095, MiltenyiBiotech), TRAV12-1 FITC (F1, Invitrogen), TRAV1-2 PE-Vio700 (REA179, MiltenyiBiotech), MR1 5-OP-RU APC (NIH), at room temperature for 30 min in the dark. Cells were washed again and resuspended in 300uL of PEB buffer. Propidium Iodide (MiltenyiBiotech) was used to exclude dead cells. Data were acquired on an LSR Fortessa (BD Bioscience, San Jose, CA, USA). Flow cytometry data were analysed using FlowJo (Treestar, Ashland, OR, USA) software.

Supplementary Table S8: Immunophenotyping blood sample collection phenotype information. Information about disease phenotype was provided via de Montreal classification. Disease score was available through the Harvey-Bradshaw index. NA corresponds to "information not available".

	Sample group		
	CD	**UC**	**Healthy**
Basic phenotype			
Samples	25	13	18
Gender, M/F (%)	14/11 (56/44%)	8/5 (60/40%)	8/10 (45/55%)
Age, mean±std	43±16	43±12	31±8
Montreal classification			
Age at diagnosis in years			
A1, <16 (%)	4 (16%)	–	–
A2, 17-40 (%)	13 (52%)	–	–
A3, >40 (%)	4 (20%)	–	–
Info not available (NA)	4		
Behaviour			
B1, inflammatory (%)	9 (36%)	–	–
B2, stricturing (%)	6 (24%)	–	–
B3, penetrating (%)	7 (28%)	–	–
Info not available (NA)	3	–	–
Location			
L1, ileal (%)	11 (44%)	–	–
L2, colonic (%)	1 (4%)	–	–
L3, ileocolonic (%)	9 (36%)	–	–
Info not available (NA)	4	–	–
Severity			
S0, Clinical remission	–	5 (41%)	–
S1, Mild	–	2 (16%)	–
S2 Moderate	–	–	–
S3 Severe	–	–	–
Info not available (NA)		6	–
Extensity			
E1 Ulcerative proctitis	–	–	–
E2 Left-side UC	–	4 (33%)	–
E3 Extensive	–	3 (25%)	–
Info not available (NA)	–	6	
Disease activity			
HBI, mean±std	3 ±3	–	–
Clinical remission, HBI≤4 (%)	10 (40%)		
Mild activity, 5≤HBI≤7 (%)	6 (24%)		–
Moderate activity, 8≤HBI≤16 (%)	1 (4%)	–	–
Info not available (NA)	8	–	–

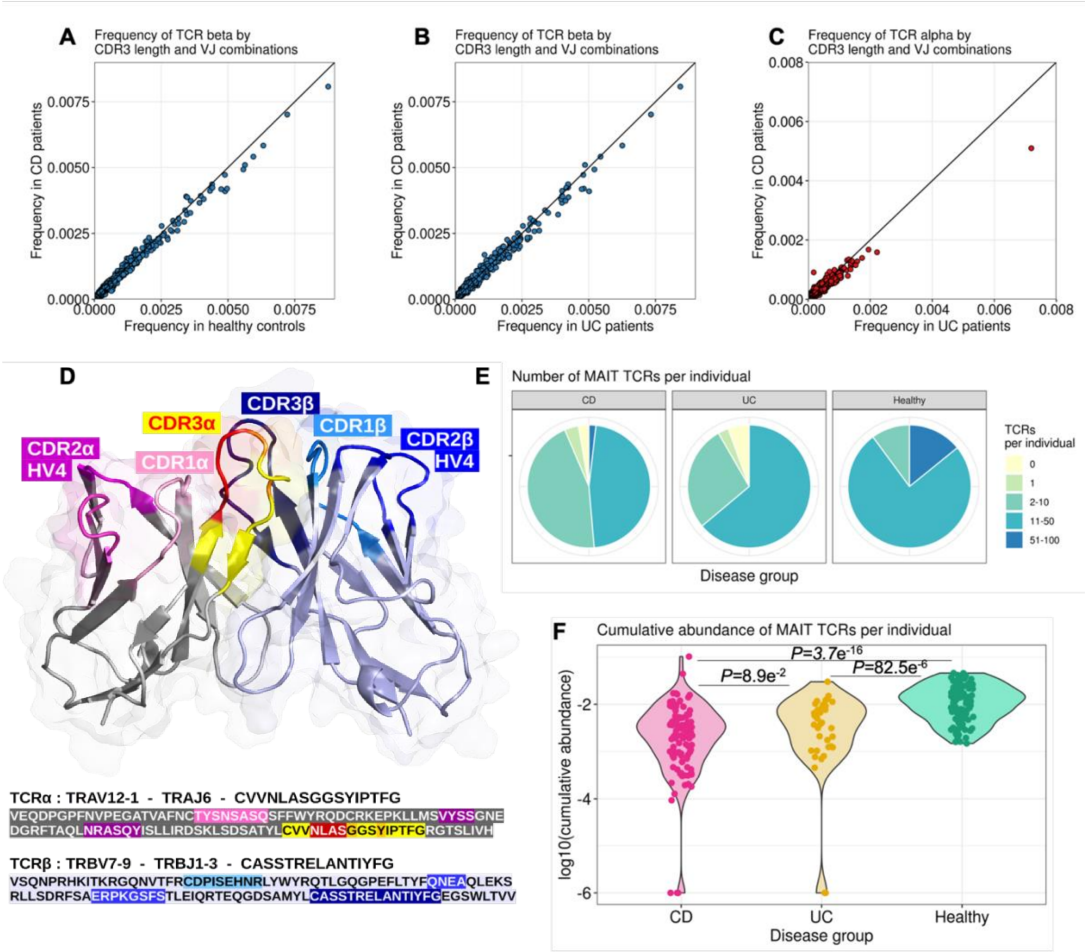
Supplementary Table S9: Immunophenotyping blood and intestinal tissue sample collection phenotype information. Information about disease phenotype was provided via de Montreal classification. Disease score was available through the Harvey-Bradshaw index. NA corresponds to "information not available".

	Sample group		
	CD	**UC**	**CRC**
Basic phenotype			
Samples	14	7	19
Gender, M/F (%)	7/7 (50/50%)	6/1 (85/15%)	11/8 (57/43%)
Age, mean±std	44±18	31±8	68±14
Tissue			
Blood	14	7	0
Gut (colon biopsy)	14	7	19
Montreal classification			
Age at diagnosis in years			
A1, <16) (%)	–	–	–
A2, 17-40 (%)	–	–	–
A3, >40 (%)	–	–	–
Info not available (NA)	14		
Behaviour			
B1, inflammatory (%)	7 (50%)	–	–
B2, stricturing (%)	2 (14%)	–	–
B3, penetrating (%)	1 (7%)	–	–
Info not available (NA)	4	–	–
Location			
L1, ileal (%)	5 (35%)	–	–
L2, colonic (%)	3 (21%)	–	–
L3, ileocolonic (%)	1 (7%)	–	–
Info not available (NA)	5	–	–
Severity			
S0, Clinical remission	–	–	–
S1, Mild	–	1 (14%)	–
S2, Moderate	–	3 (42%)	–
S3, Severe	–	–	–
Info not available (NA)	–	3	–
Extensity			
E1 Ulcerative proctitis	–	3 (42%)	–
E2 Left-side UC	–	2 (28%)	–
E3 Extensive	–	–	–
Info not available (NA)	–	2	–
Disease activity			
Mild activity, 5≤HBI≤7 (%)	1 (7%)	3 (42%)	–
Moderate activity, 8≤HBI≤16 (%)	–	2 (28%)	–
Severe activity, 16≤HBI≤∞ (%)	–	1 (14%)	–
Info not available (NA)	13	1	–

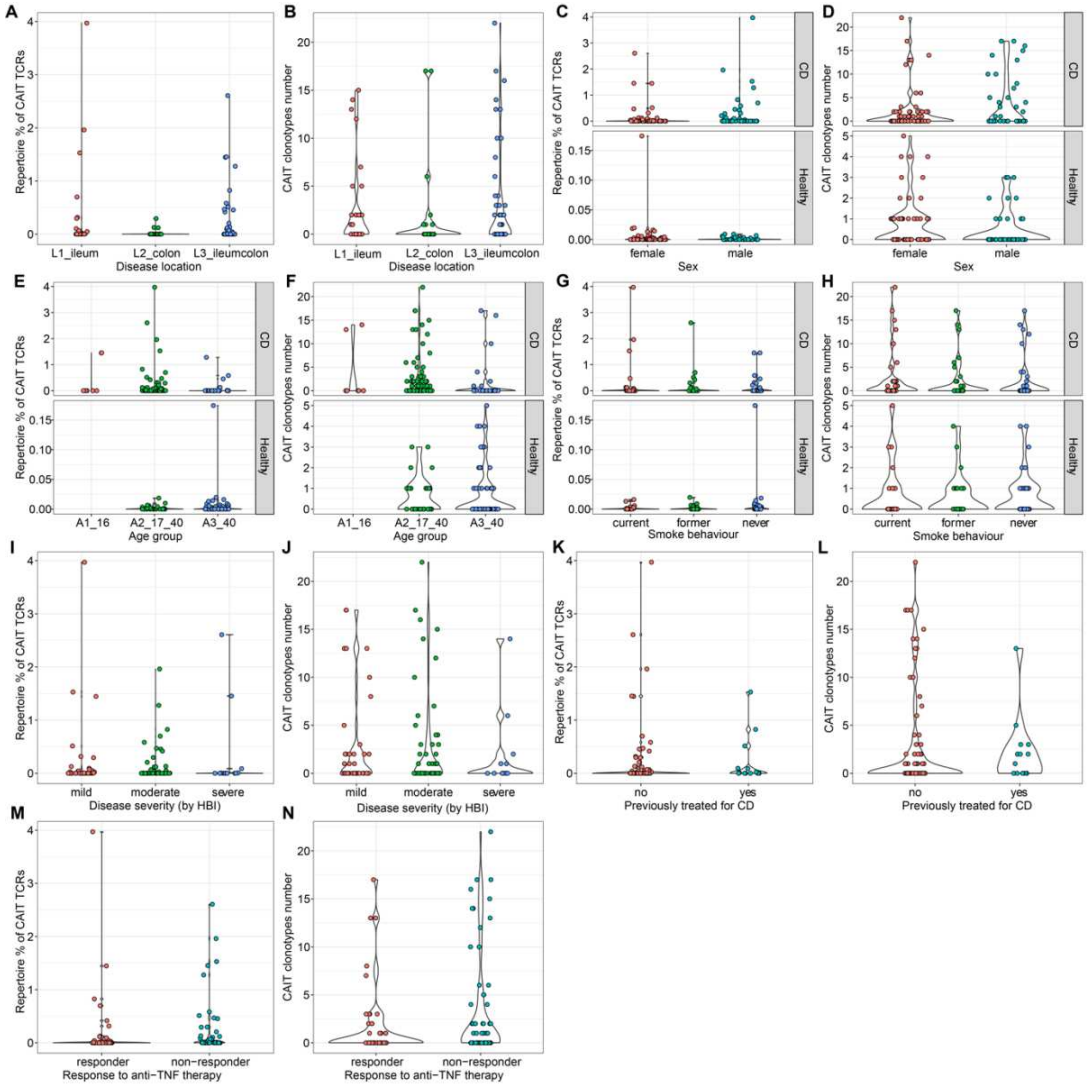
References

1. R. F. Harvey, J. M. Bradshaw, A SIMPLE INDEX OF CROHN'S-DISEASE ACTIVITY. *The Lancet* **315**, 514-514 (1980).
2. S. Waniek *et al.*, Association of Vitamin E Levels with Metabolic Syndrome, and MRI-Derived Body Fat Volumes and Liver Fat Content. *Nutrients* **9**, (2017).
3. M. Koch *et al.*, Dietary patterns associated with magnetic resonance imaging–determined liver fat content in a general population study. *The American Journal of Clinical Nutrition* **99**, 369-377 (2013).
4. U. Nöthlings, M. Krawczak, PopGen. *Bundesgesundheitsblatt - Gesundheitsforschung - Gesundheitsschutz* **55**, 831-835 (2012).
5. F. Degenhardt *et al.*, Construction and benchmarking of a multi-ethnic reference panel for the imputation of HLA class I and II alleles. *Human Molecular Genetics* **28**, 2078-2092 (2018).
6. X. Zheng *et al.*, HIBAG—HLA genotype imputation with attribute bagging. *The Pharmacogenomics Journal* **14**, 192-200 (2014).
7. F. Degenhardt *et al.*, Trans-ethnic analysis of the human leukocyte antigen region for ulcerative colitis reveals shared but also ethnicity-specific disease associations. *Hum Mol Genet*, (2021).
8. M. V. Pogorelyy *et al.*, Persisting fetal clonotypes influence the structure and overlap of adult human T cell receptor repertoires. *PLOS Computational Biology* **13**, e1005572-e1005572 (2017).
9. M. Shugay *et al.*, Towards error-free profiling of immune repertoires. *Nature Methods* **11**, 653-655 (2014).
10. D. a. Bolotin *et al.*, MiXCR: software for comprehensive adaptive immunity profiling. *Nature Methods* **12**, 380-381 (2015).
11. Y. Benjamini, Y. Hochberg. (WileyRoyal Statistical Society, 1995), vol. 57, pp. 289-300.
12. N. T. Csardi Gabor, The igraph software package for complex network research. *InterJournal, Complex Systems* **5**, 1–9 (2006).
13. R. Gowthaman, B. G. Pierce, TCRmodel: high resolution modeling of T cell receptors from sequence. *Nucleic Acids Res* **46**, W396-w401 (2018).
14. A. Butler, P. Hoffman, P. Smibert, E. Papalexi, R. Satija, Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature Biotechnology* **36**, 411-420 (2018).
15. I. Korsunsky *et al.*, Fast, sensitive and accurate integration of single-cell data with Harmony. *Nature Methods* **16**, 1289-1296 (2019).
16. G. Finak *et al.*, MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome biology* **16**, 278-278 (2015).

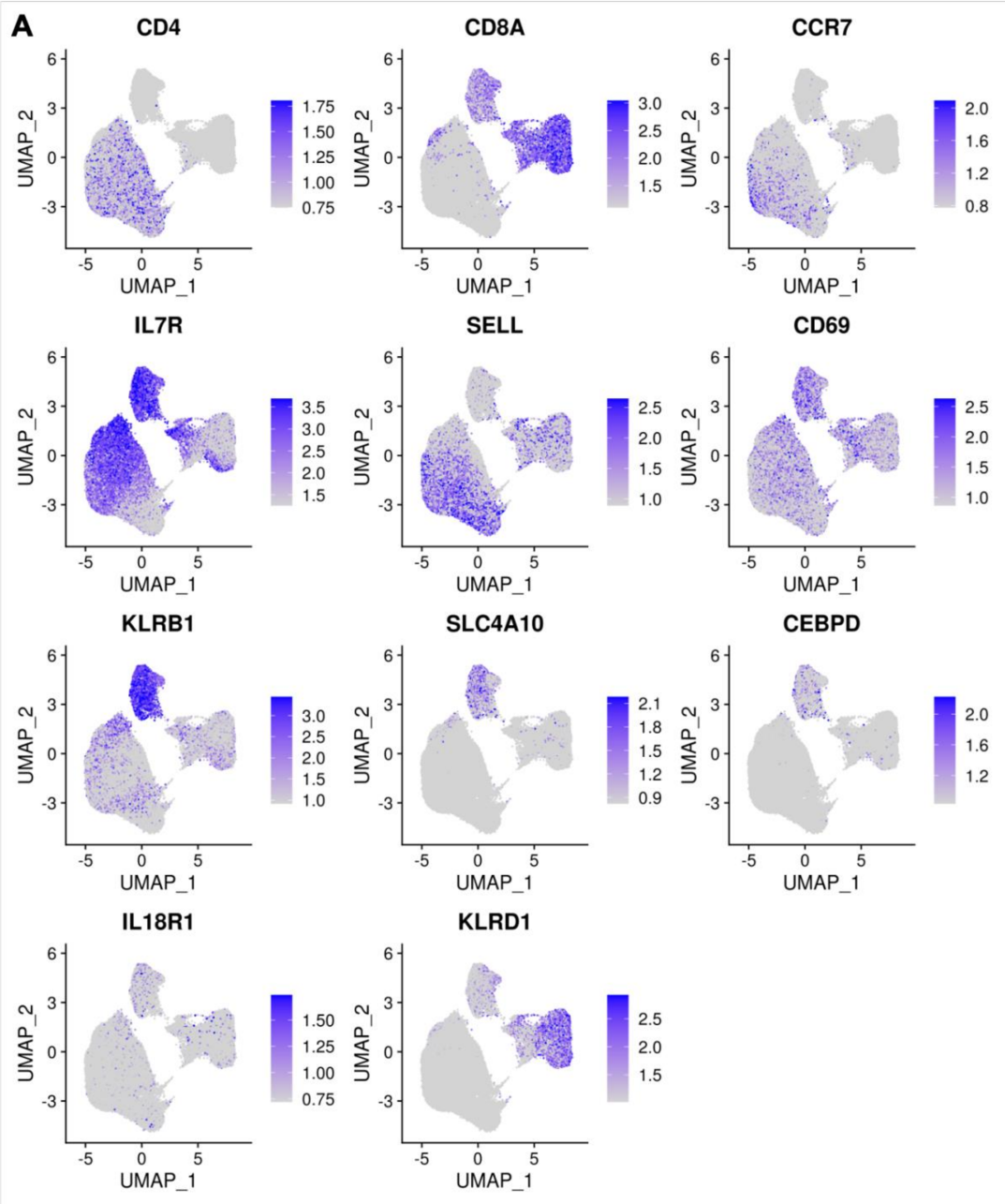
Supplementary Figures



Supplementary Figure S1: TCR groups comparison among disease groups, representative CAIT TCR structure and presence and abundance of MAIT clonotypes in the whole blood collection. **(A)** Scatterplot of the frequency of all CDR3 lengths and VJ gene combinations in CD patients and healthy controls for TCR beta. **(B)** UC patients and CD patients for TCR beta **(C)** UC patients and CD patients for TCR alpha. **(D)** Protein structural model of the TCR receptor TRAV12-1-TRAJ6-CVVNLASGGSYIPTF / TRBV7-9-TRBJ1-3-CASSTRELANTIYF. CDR1, CDR2, CDR3 and CDR2-HV4 loops are highlighted as indicated in boxes and colors as indicated in the sequence panel below. TCRα non-CDR regions are colored grey, TCRβ in light blue. **(E)** Pie chart showing the proportion of individuals carrying different numbers of MAIT clonotypes. **(F)** Cumulative abundance of MAIT clonotypes per individual. Differences between disease groups have been assessed using Mann-Whitney U test, followed by FDR multiple comparison correction.

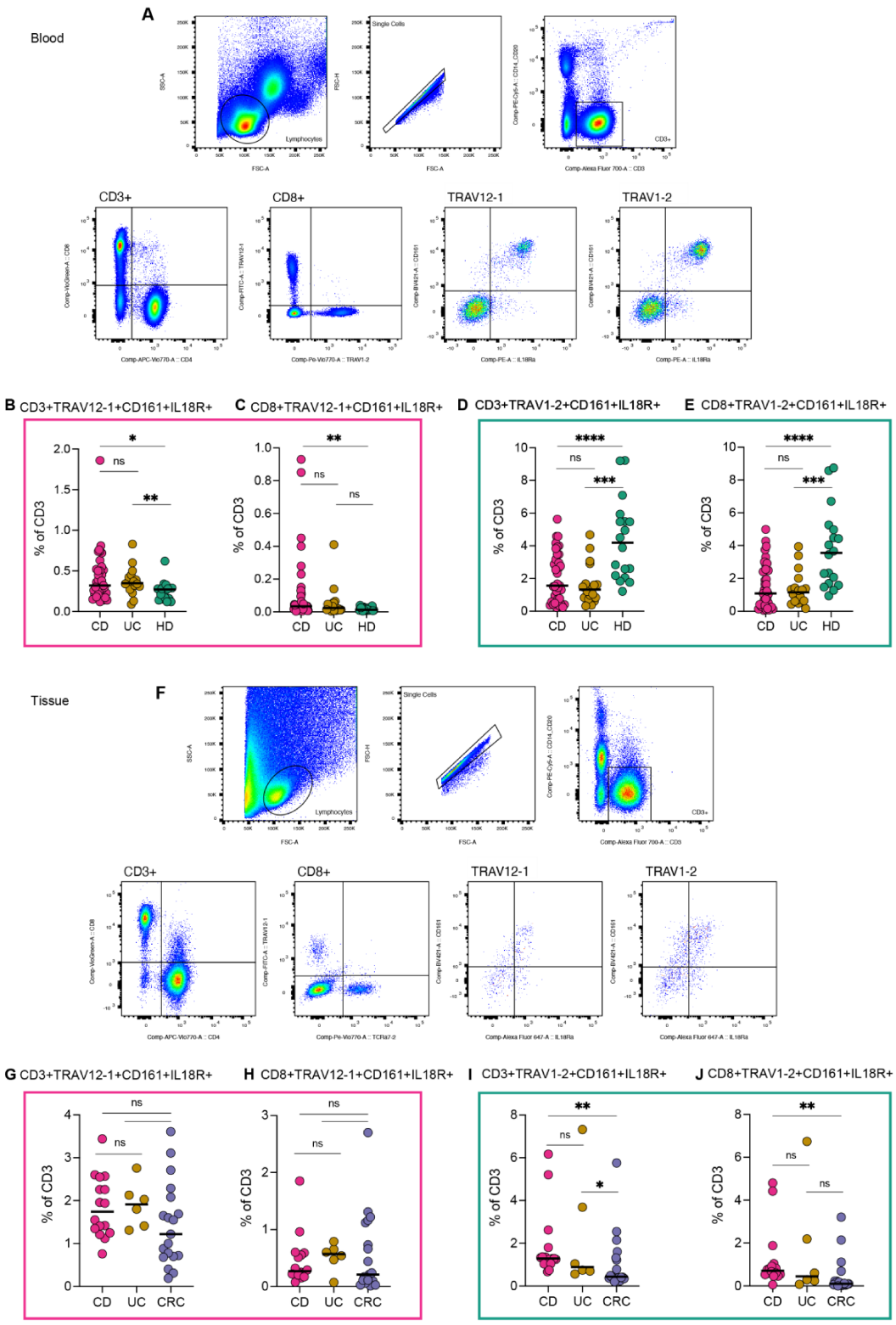


Supplementary Figure S2: Correlation of CAIT clonotypes with phenotype and clinical parameters. (A) CAIT repertoire % and disease location as for the Montreal classification in CD patients **(B)** Number of CAIT clonotypes and disease location as for the Montreal classification in CD patients **(C)** CAIT repertoire % and sex for CD patients and healthy controls **(D)** Number of CAIT clonotypes and sex for CD patients and healthy controls **(E)** CAIT repertoire % and age groups as for the Montreal classification, for CD patients and healthy controls **(F)** Number of CAIT clonotypes and age groups as for the Montreal classification, for CD patients and healthy controls **(G)** CAIT repertoire % and smoking behaviour, for CD patients and healthy controls **(H)** Number of CAIT clonotypes and smoking behaviour, for CD patients and healthy controls **(I)** CAIT repertoire % and disease activity as per the Harvey-Bradshaw index (HBI). Mild: 5≤HBI≤7. Moderate: 8≤HBI≤15. Severe: 16≤HBI≤∞, for CD patients. **(J)** Number of CAIT clonotypes and disease activity as per the Harvey-Bradshaw index (HBI). Mild: 5≤HBI≤7. Moderate: 8≤HBI≤15. Severe: 16≤HBI≤∞, for CD patients. **(K)** CAIT repertoire % and status of previous treatment for CD (treated/yes and untreated/no). **(L)** Number of CAIT clonotypes status and of previous treatment for CD (treated/yes and untreated/no). **(M)** CAIT repertoire % at baseline and response to anti-TNF treatment, in CD patients. **(N)** Number of CAIT clonotypes at baseline and response to anti-TNF treatment, in CD patients. Differences between disease groups have been assessed using Mann-Whitney U test, followed by FDR multiple comparison correction.



Supplementary Figure S3: Distribution of Seurat cluster marker genes in single-cell data (A) Distribution of marker genes among single-cell UMAP plot.

Rosati et al. - A novel unconventional T cell population enriched in Crohn's disease



Rosati et al. - A novel unconventional T cell population enriched in Crohn's disease

Supplementary figure S4: Immunophenotyping of TRAV1-2$^+$ or TRAV12-1$^+$ CD161$^+$IL18R$^+$ cells. (A) Flow cytometry plots showing gating strategy fort the immunophenotyping analysis done in peripheral blood from one representative CD patient. All stainings were performed on freshly isolated PBMCs. After exclusion of doublets and dead cells, all analyses were performed on CD14-CD20-CD3+ lymphocytes. Stains of CD4 and CD8 cells were used to investigate subpopulations of CAIT cells.

(B) PBMCs were gated on CD3+ lymphocytes and the percentage of TRAV12-1$^+$CD161$^+$IL18R$^+$ cells (which include CAIT cells) among CD3+ is indicated.

(C) PBMCs were gated on CD3+CD8+ lymphocytes and the percentage TRAV12-1$^+$CD161$^+$IL18R$^+$ (which include CAIT cells) cells among CD3+ is indicated.

(D) PBMCs were gated on CD3+ lymphocytes and the percentage of TRAV1-2$^+$CD161$^+$IL18R$^+$ cells (which include MAIT cells) among CD3+ is indicated.

(E) PBMCs were gated on CD3+CD8+ lymphocytes and the percentage TRAV1-2$^+$CD161$^+$IL18R$^+$ cells (which include MAIT cells) among CD3+ is indicated.

(F) Flow cytometry plots of cells isolated from intestinal mucosa of one representative CD patient.

(G) Intestinal cells were gated on CD3+ lymphocytes and the percentage of TRAV12-1$^+$CD161$^+$IL18R$^+$ cells (which include CAIT cells) among CD3+ is indicated.

(H) Intestinal cells were gated on CD3+CD8+ lymphocytes and the percentage TRAV12-1$^+$CD161$^+$IL18R$^+$ (which include CAIT cells) cells among CD3+ is indicated.

(I) Intestinal cells were gated on CD3+ lymphocytes and the percentage of TRAV1-2$^+$CD161$^+$IL18R$^+$ cells (which include MAIT cells) among CD3+ is indicated.

(J) Intestinal cells were gated on CD3+CD8+ lymphocytes and the percentage TRAV1-2$^+$CD161$^+$IL18R$^+$ cells (which include MAIT cells) among CD3+ is indicated.

Sample size of peripheral blood samples was n=39 CD patients, n=19 UC patients and n=18 healthy controls. For matched samples of blood and intestinal biopsies there were n=14 CD patients, n=7 UC patients and n=19 CRC control patients. Mann-Whitney U test, ns: $P > 0.05$, * $P \leq 0.05$, ** $P \leq 0.01$, *** $P \leq 0.001$, **** $P \leq 0.0001$.