A conserved population of MHC II-restricted, innate-like, commensal-reactive T cells in the gut of humans and mice

Carl-Philipp Hackstein[1,2], Dana Costigan[3], Linnea Drexhage[2,9], Claire Pearson[4], Samuel Bullers[4], Nicholas Ilott[4], Hossain Delowar Akther[1,2], Yisu Gu[4], Michael E.B. FitzPatrick[2], Oliver J. Harrison[5,6], Lucy C. Garner[2], Elizabeth H. Mann[4], Sumeet Pandey[2], Matthias Friedrich[2,4], Nicholas M. Provine[2], Holm Uhlig[7], Emanuele Marchi[2], Fiona Powrie[4], Paul Klenerman[1,2*], Emily E. Thornton[4,8*]

[1]Peter Medawar Building for Pathogen Research, University of Oxford

[2]Translational Gastroenterology Unit, Nuffield Department of Medicine, University of Oxford

[3]MRC Human Immunology Unit, MRC Weatherall Institute of Molecular Medicine, University of Oxford

[4]Kennedy Institute of Rheumatology, NDORMS, University of Oxford

[5]Center for Fundamental Immunology, Benaroya Research Institute, 1201 9th Ave, Seattle, WA 98101, USA.

[6]Department of Immunology, University of Washington, 750 Republican St, Seattle, WA 98108, USA.

[7]Translational Gastroenterology Unit, and Biomedical Research Centre, and Department of Paediatrics, University of Oxford, Oxford, OX39DU, UK.

[8]current affiliation: MRC Human Immunology Unit, MRC Weatherall Institute of Molecular Medicine, University of Oxford and Nuffield Department of Medicine, University of Oxford

[9]current affiliation: Sir William Dunn School of Pathology, University of Oxford

*co-senior author

lead contact emily.thornton@imm.ox.ac.uk

30

31

Abstract

Interactions with commensal microbes shape host immunity on multiple levels and are recognized to play a pivotal role in human health and disease. In this study, we show that MHC-II restricted, commensal-reactive T cells in the colon of both humans and mice acquire transcriptional and functional characteristics typically associated with innate-like T cells, including the expression of the key transcription factor PLZF and the ability to respond to cytokines including IL-12, IL-18 and IL-23 in a TCR-independent manner. These **M**HC-II restricted, **i**nnate-like, **c**ommensal-reactive T cells (T_{MIC}) are endowed with a polyfunctional effector potential spanning classic Th1- and Th17-cytokines, cytotoxic molecules as well as regulators of epithelial homeostasis and represent an abundant and conserved cell population in the human and murine colon. T cells with the T_{MIC} phenotype were increased in ulcerative colitis patients and their presence aggravated pathology in DSS-treated mice, pointing towards a pathogenic role in colitis. Our findings add T_{MIC} cells to the expanding spectrum of innate-like immune cells positioned at the frontline of intestinal immune surveillance, capable of acting as sentinels of microbes and the local cytokine milieu. (187 words)

Introduction

The mucosal surfaces of the intestine are colonized by myriad commensal microbes that outnumber the host's own cells and play crucial roles in metabolizing certain nutrients as well as limiting the growth of potentially pathogenic microbes[1, 2]. In this unique environment, a delicate balance is required to restrict antimicrobial immune responses to a minimum necessary to contain the microbes within the intestinal lumen while avoiding excessive inflammatory processes that would damage the host tissue[3]. MHC II-restricted CD4+ T cells that are responsive to microbial antigens are a major immune cell population in the intestine and were shown to play key roles in both homeostasis and chronic inflammation[4]. Considerable efforts have been made in the past to unravel how interactions with intestinal microbes shape the phenotype and function of these immune cells, including experiments in gnotobiotic mice[5, 6, 7], tetramer-based approaches analysing antigen-specific endogenous cells[8, 9, 10] and experiments utilizing TCR-transgenic mice to study the evolution of the response to microbial antigens[8, 11, 12, 13, 14, 15]. Using these tools, previous studies demonstrated that the type of antigen, as well as the milieu during activation, has a huge impact on local T cell differentiation, which results in a diverse continuum of phenotypes spanning from regulatory T cells (Treg)[11, 16] to Th1 and Th17 cells[8, 10].

Cytokine-mediated signals play a crucial role in inducing, regulating and maintaining T cells in tissues[17, 18, 19, 20]. IL-18R1 expression is more frequent on lamina propria CD4 T cells compared to their counterparts in lymphoid tissues[21] and signalling mediated by IL-18, IL-12 and type I interferon plays important roles in shaping T cell phenotypes and functions in the intestine[21, 22]. This cytokine-responsiveness is a feature conventional MHC-restricted T cells share with other immune cells enriched in mucosal sites, including innate-like T cells, which can also recognize bacterial products through their TCRs. Interestingly, most human colonic CD4 T cells also express the C-type lectin CD161[23], a key marker of human innate-like T cells such as the semi-invariant Vδ2 population[24, 25] as well as with iNKT and MAIT cells[26]. The association between high CD161 expression and the ability of T cells to respond to

cytokine stimulation in a TCR-independent, innate-like manner is not restricted to these well-defined subsets and can been observed in small populations across all T cell lineages[27]. Such a capacity for cytokine-responsiveness in an environment as rich in stimulating agents as the gut could have important consequences in both health and disease.

In this study, we seek to understand whether microbe-reactive cells in the gut may be able to span the MHC-II restricted conventional T cell population and cytokine-responsive innate-like T cells. We describe that in the human colon, microbe-reactive CD4 T cells have high CD161 expression, which is associated with the acquisition of key features of innate-like T cells including PLZF-expression and the ability to respond to cytokines including IL-12, IL-18, and IL-23 independently of TCR-signals even though they are MHC-II-restricted and express a diverse TCR repertoire. CD161^{hi} CD4 T cells accounted for the vast majority of CD4 T cells responding to several common commensal microbes placing these cells at the forefront of antimicrobial responses during homeostasis and inflammation. Interestingly, the transcriptional signature defining these MHC II-restricted, innate-like and commensal reactive T cells (T_{MIC}) was expressed independently of established T_{RM} signatures, suggesting a separate differentiation pathway. In mice, the transcriptional and functional characteristics of human T_{MIC} cells are mirrored by a unique population of MHC II-restricted double-negative alpha/beta T cells. Strikingly, the comparison of the commensal-reactive Cbir-TCR transgenic T cells with a helicobacter-reactive TCR transgenic, whose target is absent from the flora, revealed that only the commensal-reactive cells acquire a T_{MIC} phenotype in the murine gut, demonstrating its dependency on microbial antigen availability. Both human and murine T_{MIC} cells displayed a mixed Th1/Th17 effector profile, and T_{MIC} cells contributed to pathology in a murine colitis model.

Results

Human microbe-reactive CD4 T cells express high levels of CD161

The gut contains a broad range of different microbe-reactive CD4 T cells, which produce TNF when exposed to their respective microbe *in vitro*[28] (Figure 1, A and B). As we intended

to focus on conventional CD4 T cells, we excluded cells expressing TCR chains associated with known unconventional T cells populations and focussed on CD4 T cells negative for Vα7.2 (expressed by MAIT and GEM T cells), Vα24-Jα18 (expressed by iNKT cells) and TCRγδ (Supplemental Figure 1, A). Studying the commensal-reactive human colonic CD4 T cells responding to *E.coli*, *S.aureus* and *C.albicans*, we noticed that the majority of TNF-positive CD4 cells expressed above-average levels of the C-type lectin CD161 (Figure 1, A). While CD161-expressing colonic CD4 T cells as a whole were described as a population of cells with Th17 characteristics before[23], we were intrigued by the fact that anti-microbial responses seemed to be enriched within cells expressing the highest level of this marker and therefore wanted to study the role of CD161-expressing CD4 T cells in the gut in more detail. To that end, we subdivided the colonic CD4 population into CD161-negative, $CD161^{int}$ and $CD161^{hi}$ cells, defining $CD161^{hi}$ based on CD161-expression levels observed on a discrete subset of CD4 T cells co-expressing high levels of CD161 with another NK-cell marker, CD56 (Figure 1, C). The rationale behind this gating strategy was that CD56 expression can be used to pull out a subset MAIT cells with higher sensitivity towards cytokine stimulation from the larger CD8 T cell population[29].As CD56 expression is not a uniform feature of the entire MAIT population, we decided to study both double-positive (DP, $CD161^{hi}$CD56+) and $CD161^{hi}$ CD56- CD4 T cells in most assays. TNF-expressing microbe-responsive cells were found among both, $CD161^{hi}$CD56- and DP CD4 T cells, and the frequency of TNF-positive, microbe-responsive cells was significantly higher in both populations compared to either $CD161^{-}$ or $CD161^{int}$ cells. (Figure 1, D). Compared to each other, there was a non-significant trend for higher percentages of TNF-positive cells within the DP population.

In sum, our data show that the vast majority of human colonic microbe-responsive CD4 T cells displays a $CD161^{hi}$ phenotype with varying expression of CD56.

MHCII-restricted CD161^{hi} colonic CD4 T cells share transcriptional and functional characteristics with innate-like T cells

Further phenotypic profiling revealed that almost all microbe-reactive CD161^{hi} CD4 T cells expressed IL18Rα, while expression of IFNγ varied between the different microbe-specific populations (Figure 1, E and F). Strikingly, our experiments also revealed that PLZF, the key transcription factor regulating development and function of innate-like T cells, was expressed in microbe-responsive cells compared to the bulk of non-responding TNF- cells and was also more highly expressed in microbe-reactive cells compared to CD4 T cells producing TNF in response to staphylococcal enterotoxin B (SEB, Figure 1, G and H).

These findings prompted us to test whether these anti-microbial responses were dependent on the conventional CD4 T cell restriction element MHC II or if there was any role for unconventional antigen-presenting molecules like MR1 and CD1d. MHC II blocking massively reduced or abrogated the microbe-induced TNF production by human colonic CD4 T cells (Figure 1, I and J), supporting previous work showing these commensal-reactive cells are MHCII restricted[28]. In contrast, blocking of CD1d and MR1, the molecules presenting antigens to type I and II NKT cells and MAIT cells respectively, had no significant effects on anti-microbial responses. Similar, the addition of neutralizing antibodies against IL-12 and IL-18 failed to block TNF production (Figure 1, I and J).

Taken together, this indicates that commensal-responsive CD161^{hi} CD4 T cells, despite expression of several markers associated with innate-like T cells (CD161, higher levels of cytokine receptors, PLZF) are a population of MHCII-restricted CD4 T cells.

An intriguing aspect of our findings is that this innate-like phenotype was found on cell populations responding to microbes that, while all being considered commensals in the human intestine, differ notably from each other and included a gram-negative proteobacterium (*E.coli*), a member of the gram-positive firmicutes phylum (*S.aureus*) and the yeast *C.albicans*. As only a small fraction of the entire CD161^{hi} subset responded to each of these microbes (Figure 1 A

and D), we hypothesized that the CD161^{hi} population would contain further cell populations with the same phenotype, presumably responsive to microbes not tested in this study.

To test this idea, we phenotyped bulk CD161^{hi} CD4 as defined earlier (Figure 1, C) in more detail. Like the microbe-specific subsets, bulk CD161^{hi} CD4s expressed higher levels of PLZF (Supplemental Figure 1, B and C) and showed higher expression of IL18Rα compared to their intermediate and negative counterparts (Supplemental Figure 1, D). Strikingly, RNAseq of human intestinal CD4 T cells showed that a transcriptional core signature (Supplemental Table 1) previously identified in MAIT cells as well as other innate-like T cell populations in human blood[27], was strongly enriched in CD161^{hi} CD4 T cells as well (Supplemental Figure 1, F), suggesting that these CD4s indeed share transcriptional features with established innate-like T cell subsets.

The ability to mount TCR-independent responses to cytokine stimulation is a key functional property of innate-like T cells and hence, we assessed how the different colonic CD4 T cell subsets responded to IL-12/18 stimulation. Bulk CD161^{hi} CD4 were able to respond to combined IL-12 and IL-18 stimulation in the absence of additional TCR-stimulation (Figure 1, K) and accounted for the majority of responding CD4 T cells (Figure 1, L). As reported for CD56+ and CD56- MAIT cells[29], DP CD4 T cells showed a higher percentage of IFNγ-production than their CD161^{hi} (CD56^{-}) counterparts, but overall the responses of both subsets resembled the cytokine-induced, TCR-independent responses of CD161^{hi} innate-like T cells like MAIT, iNKT and Vδ2+ γδT cells.

Innate-like T cells including γδ T[30, 31, 32, 33], iNKT[34, 35], MAIT[36, 37, 38] and H2-M3 restricted T cells[39] have been shown to be involved in tissue repair in different barrier organs.

Having established that human CD4 CD161^{hi} T cells harbour populations of microbe-reactive T cells with innate-like characteristics localized in the colonic lamina propria, we also assessed whether they would also express tissue repair- associated factors upon stimulation. The regulation of tissue repair is a complex process involving many different

effector molecules and pathways. Hence, we used a multi-modal single cell sequencing approach to analyse the expression of a established gene list of tissue repair-associated factors[37, 39] in colonic CD4 T cells that had been stimulated with either cytokines (IL-12 and IL-18), plate-bound anti-CD3 antibodies or a combination of both overnight (Supplemental Figure 2). Gene counts of repair-associated factors were barely affected in CD161^{hi} CD4 T cells that had received cytokine stimulation (Supplemental Figure 2, A), while TCR or combined stimulation induced increased expression compared to unstimulated control cells (Supplemental Figure 2, B and C). In line with that, gene set enrichment analyses (GSEA) showed that indeed the tissue repair gene set was enriched in TCR and TCR + cytokine-stimulated, but not cytokine-only stimulated CD161^{hi} CD4 T cells (Supplemental Figure 2, D). Leading edge genes driving the enrichment included CSF1 and 2, the genes encoding m-CSF and GM-CSF, growth factors like VEGFA and AREG, the gene encoding Amphiregulin (Supplemental Figure 2, E). To conclude, human CD161hi CD4 T cells show the same TCR-dependent potential for the production of tissue repair-associated factors previously found in innate-like T cells.

TCR-sequencing in three donors revealed that bulk CD161^{hi} CD4 T cells use a diverse range of private TCRα and -β chains in line with the idea that this population consists of a pool of MHC II- restricted cells responding to a vast range of different microbes. Interestingly, while TCRα diversity was comparable to the repertoire diversity found in CD161^{-} CD4 T cells in both CD161^{hi} subsets regardless of CD56 expression, DP CD4 T cells showed some bias in their TCRβ usage (Supplemental Figure 2, F). However, no public TCRβ sequences could be found, suggesting that the DP phenotype can arise in CD161^{hi} CD4 T cells with different TCRs and this population might be the results of private clonal expansion.

Additional phenotyping revealed an enrichment of RoRγt – expressing cells among CD161^{hi} CD4 T cells (Supplemental Figure 3, A and B). To explore possible Th17-functionality, we

stimulated CD161^{hi} CD4 T cells with IL-23, a cytokine playing a pivotal role in gut-associated inflammatory diseases[40, 41, 42, 43]. Interestingly, IL-23 alone induced production of IFNγ and GzmB by a small subset of cells in a TCR-independent manner (Supplemental Figure 3, C and D). CD161^{hi} CD4 T cells also could produce the Th17 cytokines IL-17A, IL-17F and IL-22 to a limited extent (Supplemental Figure 3, D). The production of the latter cytokines however was strictly dependent on an additional TCR-stimulus.

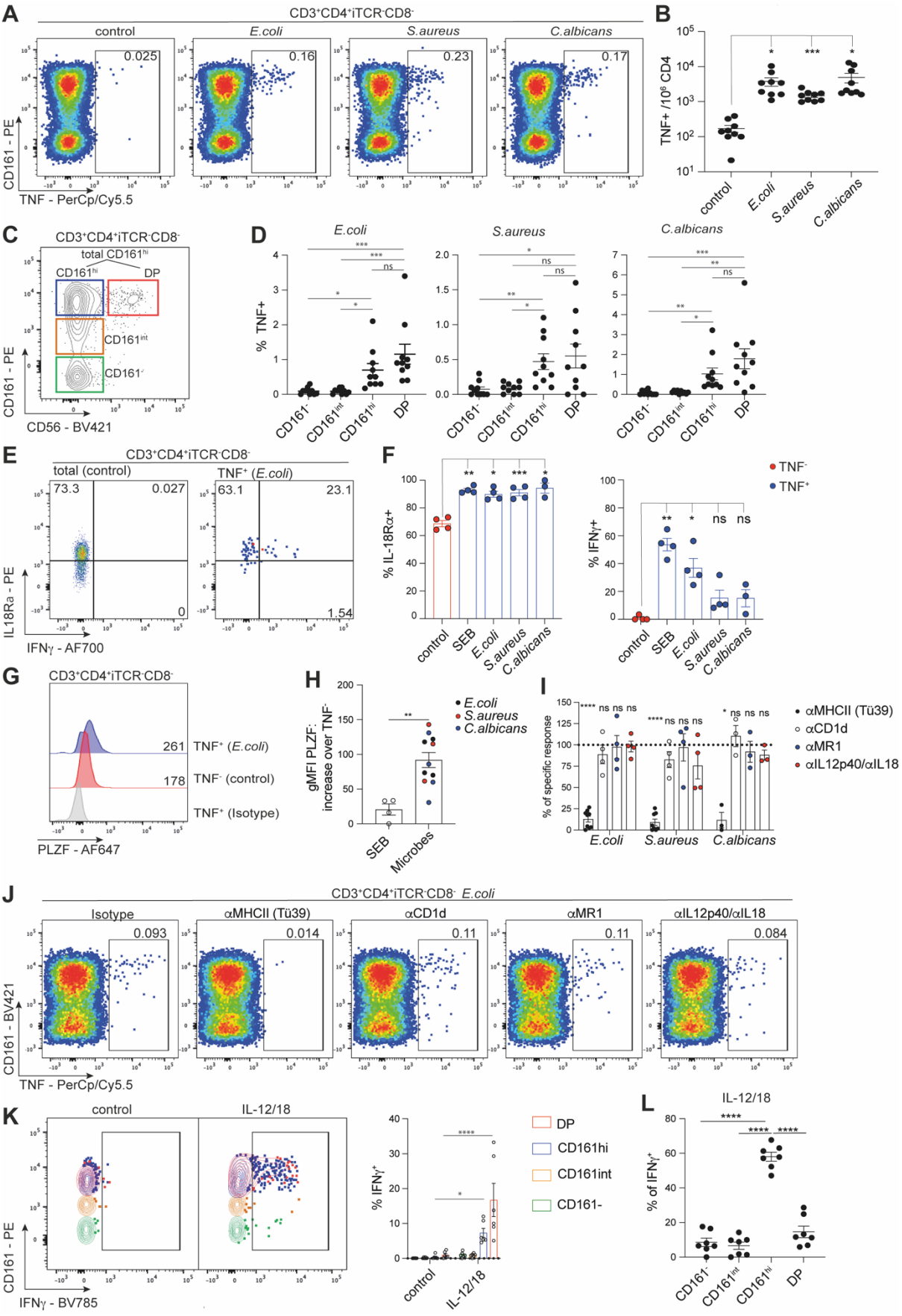


Figure 1. Human microbe-reactive CD4 T cells are CD161^{hi} and display a polyfunctional effector phenotype with innate-like features. See also Supplemental Figures 1 and 2.

(A) Flow cytometric data depicting the expression of TNF by iTCR (Vα7.2, Vα24-Jα18, TCRγδ)-negative colonic CD4 T cells after 8 hours of *in vitro* incubation in medium (control) or with the indicated heat-killed bacteria or *C. albicans*. (B) Scatterplot showing the absolute number of TNF-positive CD4 T cells per million colonic CD4 T cells. (C) Flow cytometric pictures showing the expression of CD161 and CD56 by colonic CD4 T cells after cells expressing any TCR associated with innate-like T cells (iTCR: Vα7.2, Vα24-Jα18, TCRγδ) and CD8 had been gated out. (D) Frequencies of TNF-expressing CD161^{-}, CD161^{int}, CD161^{hi} and double-positive (DP) CD4 T cells after 8 hours of stimulation with the indicated microbes. (E) Expression of IL18Rα and IFNγ on TNF^{-} or TNF^{+} CD4 T cells upon stimulation with *E.coli.* (F) Percentage of IL18Rα and IFNγ-positive cells among TNF^{-} or TNF^{+} CD4 T cells upon stimulation upon stimulation with SEB or the indicated microbes. (G) Expression of PLZF in TNF^{-} or TNF^{+} CD4 T cells upon stimulation with *E. coli.* (H) Difference in PLZF expression between TNF^{-} CD4 T cells compared to TNF^{+} CD4 T cells responding to either SEB or different microbes. (I) Microbe-specific TNF responses upon blockade of MHCII, CD1d, MR1 or IL12/18 as percentage of the response observed in the presence of isotype antibodies (100%, dotted line). Statistics were calculated comparing the response values for each blocking condition to the normalised response in each donor. (J) *E. coli* -induced TNF-expression by colonic CD4 T cells in the presence of the indicated blocking antibodies or an isotype control. (K) Expression of IFNγ by colonic CD4 T cells after 20 hours of *in vitro* stimulation with an IL-12 + IL-18 cytokine-cocktail. (G) Percentages of IFNγ-positive cells in the indicated cells from the experiment shown in (K). Data points were pooled from independent experiments using one or two human samples each. ns = not significant, $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, $^{****}p < 0.0001$; repeated measures ANOVA with Dunnett's multiple comparisons test (B, L), Friedman tests with Dunn's multiple comparisons tests (D), 2-way ANOVA with Sidak's multiple comparisons test (K). Mean ± SEM is shown.

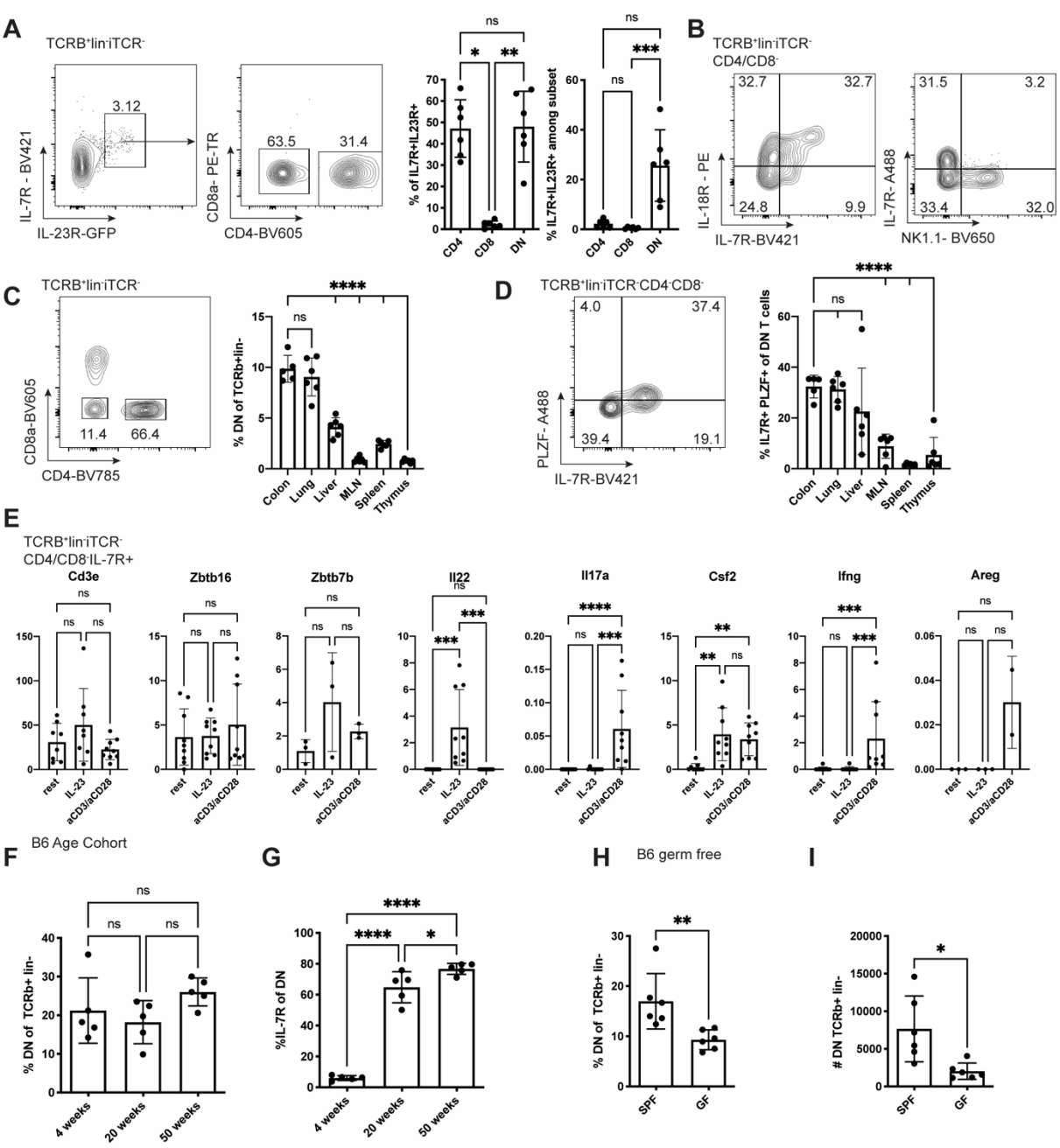
Innate-like MHCII-restricted CD161^{hi} colonic CD4 T cells display an effector memory phenotype and express a transcriptional program different from the one found in classic T_{RM} cells

Flow cytometric analyses showed that the majority of CD161^{hi} CD4 T cells displayed an effector memory phenotype (Tem, Supplemental Figure 3, E), in contrast to CD161^{-} and CD161^{int} CD4 T cells that also harbour major Tcm and naïve populations. Given the tissue-origin and characteristics displayed by CD161^{hi} CD4 T cells, we wondered if their transcriptional phenotype would correlate with the acquisition of genetic signatures associated with tissue-resident memory T cells (T_{RM}). To explore this, we obtained a list of CD161^{hi} signature genes by comparing gene expression in CD161^{hi} and DP CD4 T cells with CD161^{-} CD4 T cells. Between the genes that were significantly upregulated (log2 foldchange >2, adjusted p-value < 0.05, Supplemental Table 2) in CD161^{hi} and DP CD4 T cells there was an overlap of 26 genes, including KLRB1 (CD161), IL23R, IFNGR1, GPR65 and the transcription factors RORA and BHLHE40. Next, we assessed the expression of two previously published T_{RM} gene signatures[44, 45] in the CD4 cells from two publicly available intestinal single cell datasets[46, 47]. We then compared the expression of both T_{RM} signatures with the corresponding expression of our CD161^{hi} signature gene module on the single cell level. Expression of the two different T_{RM} signatures showed a strong positive correlation in both datasets, as expected (Supplemental Figure 3, F). On the other hand, the CD161^{hi} signature module correlated only weakly with both T_{RM} signatures (Supplemental Figure 3, G and H). Based on this, we concluded that the acquisition of the effector profile and innate-like features in CD161^{hi} CD4 T cells represents a unique feature of microbe-reactive CD4 T cells and not an inherent component of the T_{RM} - associated gene signatures in the colon.

Taken together our data suggested the potential for microbe-reactive T cells to take on an innate-like phenotype with broad cytokine responsiveness in the gut. However, since

276 experiments with *ex vivo* human cells are limited, we decided to determine whether an
277 equivalent subset of gut associated MHC-II restricted, innate-like, commensal-reactive
278 T cells (T_{MIC}) exist in mice to allow more extensive mechanistic *in vivo* experimentation.
279
280 **MHC-II restricted cytokine and microbe-reactive T cells in mice are CD4/CD8 double-**
281 **negative**
282 Given the innate-like phenotype of the microbe-reactive T cells found in human colon and
283 the absence of an easily identifiable mouse ortholog for CD161, we used IL-23R reporter
284 mice to identify a population of cytokine-responsive conventional (TCRalpha/beta+, not
285 staining positive for CD1d or MR1 tetramers) T cells in the murine lamina propria (Figure
286 2A). The IL23R-reporter signal was associated with expression of IL7R, which was
287 subsequently used as a surrogate marker in wildtype animals. When we looked further to
288 define this IL23R+IL7R+subset, which should contain Th17 cells, we found that a majority
289 did not express either CD4 or CD8a on the surface (Figure 2A), similar to cells described in
290 the peritoneum of the IL-23R reporter[48]. CD4/CD8 double negative (DN) cells have been
291 described in TCR transgenic lines[49] and settings of chronic inflammation such as lupus[50] and
292 spondyloarthropathy[51] but their function has remained enigmatic. The subset of DN cells in
293 the colon that expressed IL-7R also expressed IL-18 receptor but not NK1.1, indicating a
294 population that is broadly cytokine-responsive (Figure 2B). Gene expression analysis and
295 intracellular staining demonstrated that the absence of CD4 and CD8 was not due to
296 collagenase digestion or internalization (Supplemental Figure 4, A and B). While we used
297 MHCII blocking antibody to demonstrate MHCII restriction of the innate-like cells from human
298 tissue, we were able to use the B2M knockout mouse to eliminate the possibility that we
299 were studying a known innate-like T cell population. While the B2M knockout showed a
300 slight decrease in DN cell numbers, the overall phenotype of the cells was similar
301 (Supplemental Figure 4, C), demonstrating that these cells are not dependent on MHC-I,
302 CD1d or MR1. To determine whether the presence of this DN T cell population was specific
303 to the gut, we examined mucosal and non-mucosal tissue sites. While lymphoid organs were

304 almost devoid of DN T cells, the barrier sites contained abundant DN T cells (Figure 2C) with
305 the lungs containing a similarly sizable population of innate-like T cells (Figure 2D).
306
307 To determine whether these cells were phenotypically and functionally similar to human
308 cytokine-responsive cells, we performed flow cytometry and *in vitro* stimulations. As with the
309 human cells, mouse colonic DN T cells contain a PLZF-expressing population (Supplemental
310 Figure 4, D). As expected, these cells expressed CD3e, Zbtb7b (ThPOK), a transcription
311 factor associated with Th cells, and Zbtb16 (PLZF), which do not change upon TCR or IL-23
312 stimulation. By contrast, IL-22 production was stimulated by IL-23 treatment while IL-17A
313 and IFNg production were stimulated by TCR stimulation. GM-CSF was produced in both
314 conditions. These data suggest that the DN T cell population are similar to the cytokine and
315 microbe-responsive population found in the human colon and have the ability to tune their
316 response based on the cytokine milieu and/or antigen stimulation (Figure 2E).
317
318 In order to determine whether these cells are likely to be the microbe-responsive T cells of
319 interest in the murine gut, we examined mice from different ages, supposing that older mice
320 would have encountered more microbes and mild insults that lead to microbe translocation
321 into the lamina propria. While the percent of the overall T cell population with the CD4/CD8
322 DN phenotype did not change with age (Figure 2F), the proportion with a cytokine-
323 responsive phenotype, marked by IL-7R, increased markedly between 4 and 20 weeks of
324 age and still more by 50 weeks of age (Figure 2G). Conversely, germ free mice, devoid of
325 microbes, had lower percentages and absolute numbers of DN T cells (Figure 2H and I).



326

327 **Figure 2 Mouse cytokine-responsive CD4-/CD8- mucosal T cells respond to microbial stimuli.**

328 **See also Supplemental Figure 4.**

329 Mouse immune cells were isolated from various tissues and analysed by flow cytometry and/or

330 stimulated *in vitro*

331 (A) Flow cytometric analysis of colonic lamina propria leukocytes from IL-23RGFP/+ mice gated

332 TCRb+lin-iTCR- (MR1tet-CD1dtet-TCRgd-CD11c-CD11b-B220-) showing IL-7R and IL-23R

333 coexpressing cells (left). The gated IL-7R+/IL-23R+ population is shown for CD4 and CD8a

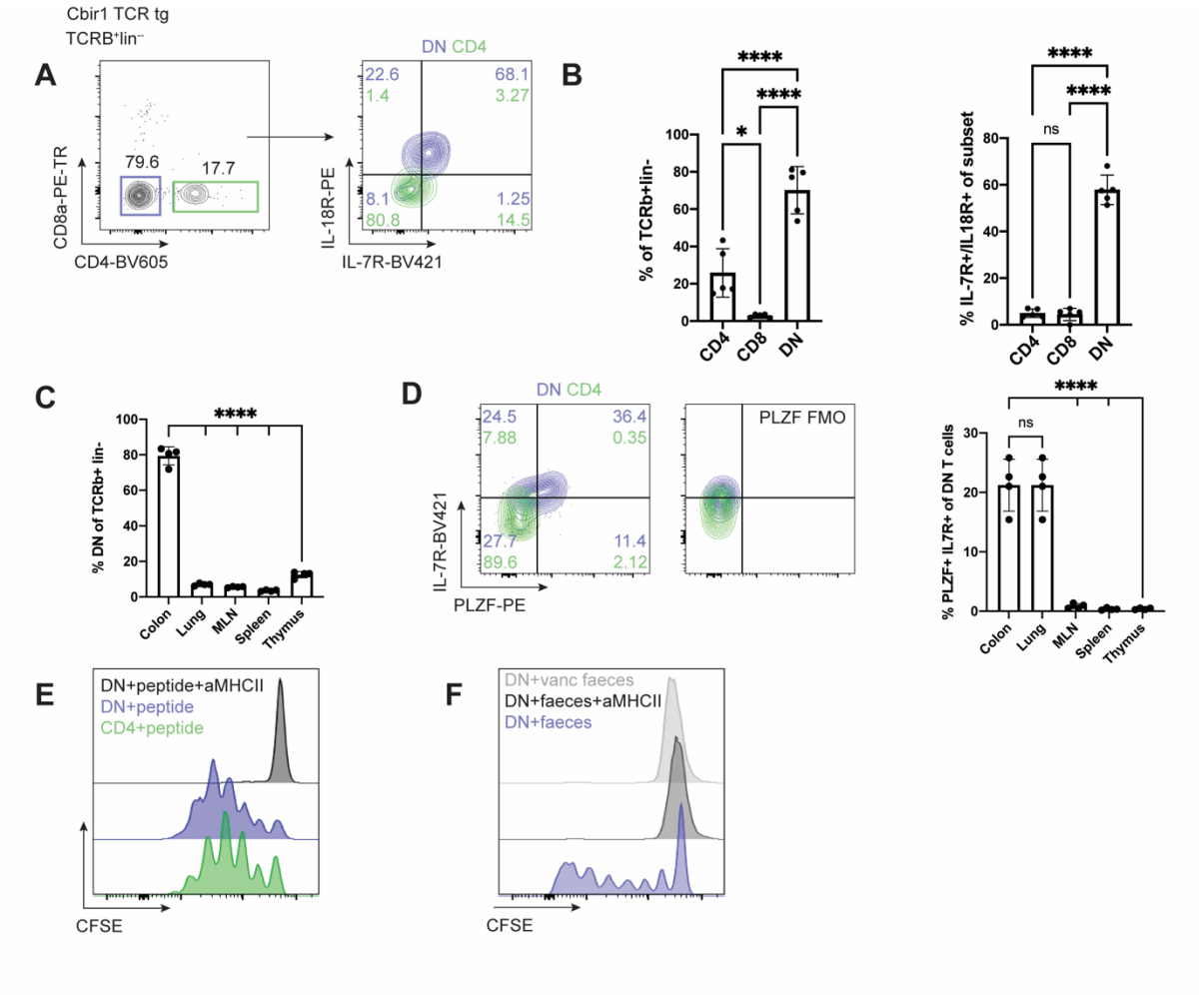
expression followed by quantification of the percent of the IL7R+/IL23R+ population is either CD4+, CD8a+ or CD4/CD8a double negative (DN). Quantification of gated TCRb+lin- CD4, CD8a, or DN cells that have the IL-7R/IL-23R positive phenotype. Each point represents one individual mouse (n=6). This experiment is representative of three independent experiments. (B) Example flow cytometry plot to demonstrate the TCRb+lin-iTCR-CD4/CD8- cells expressing IL-7R also express the IL-18R (left) and IL-7R and NK1.1 are not coexpressed in DN T cells (right). Plots are from different exeriments but representative of two experiments each, n=5. (C) Example FACS plot (colon) and quantification of DN phenotype in mouse colon, lung, liver, MLN, spleen, and thymus. Each point represents one individual mouse. n=5, representative of two independent experiments. ANOVA with post-test comparing each mean to colon. (D) Example FACS plot (colon) and quantification of T cells innate-like phenotype (IL-7R+/PLZF+) in mouse colon, lung, liver, MLN, spleen, and thymus. Each point represents one individual mouse. n=5, representative of two independent experiments. (E) FACS sorted TCRb+lin-iTCR-CD4/CD8- cells were isolated from mouse colon and stimulated in vitro with either aCD3/aCD28 or IL-23 or rested in complete media for 24 hours when RNA was extracted and qPCRs performed for Cd3e, Zbtb16 (PLZF), Zbtb7b (ThPOK), Il22, Il17a, Csf2 (GM-CSF), and Ifng. Genes were normalized to *Hprt*. Each point represents one mouse. Data represents two combined experiments, n=9 in total (F) Quantification of FACS data from DN TCRb+lin- cells from colon LPLs from mice aged 4, 20, and 50 weeks. (G) Quantification of FACS data of IL-7R expression on the DN T cell population of 4-, 20-, and 50-week-old mice. Each point represents an individual mouse. n=5 , representative of two independent experiments. (H) Quantification of FACS data of DN TCRb+lin- cells from colon LPLs from specific pathogen free (SPF) or germ free (GF) mice. (I) Absolute cell numbers of the DN cell population in the colonic LPL fraction of SPF and GF mice. Each point represents and individual mouse. n=6, representative of three independent experiments.

MHCII-restricted microbe reactive cells develop an innate-like phenotype in the gut

In order to explore further the function of gut-associated microbe-reactive T cells and to determine whether the innate-like phenotype can be observed in a known MHCII restricted, microbe-reactive T cell, we turned to the Cbir1 TCR transgenic mouse[12]. The Cbir1 transgenic T cell recognizes a flagellar antigen from Clostridium cluster XIVa, and when crossed onto the Rag1 knockout background (Rag-/-), the transgenic mouse only contains T

cells of one specificity without the possibility of recombination. Hence, these mice represent an ideal model to study the phenotype and functions of commensal-reactive MHCII-restricted T cells *in vivo*. As observed in B6 mouse colons, the colon of the Cbir1 mouse contained a population of CD4/CD8 negative cells (Figure 3A and B), but in the context of the TCR transgenic, approximately 70% of the T cells had this unique phenotype. As seen with the other microbe-reactive cells from human and mouse, these cells expressed IL-18R and IL-7R (Figure 3A and B), suggesting they are broadly cytokine-responsive in addition to being microbe-reactive. Unlike Tmics in B6 mice, Cbir1 T cells are specific for a single commensal that is found in the gut. The prevalence of the DN phenotype is generally restricted to the colon in this setting suggesting antigen presence is important for development and/or retention of these cells (Figure 3C). Like DN T cells in wildtype animals, Cbir DN T cells also expressed PLZF in the colon (Figure 3D); however, a proportion of the DN cells in the lung also had PLZF expression suggesting there may be some crosstalk between the tissues. Analysis of the thymus from B6 and Cbir1 mice demonstrated a very small but appreciable population of T_{MIC} cells in the thymus (Supplemental Figure 3E) suggesting that cells may develop in the thymus similar to other innate-like T cells and populate mucosal sites where they are able to persist. By contrast, another MHCII-restricted TCR transgenic mouse strain *(Hh-TCR Rag-/-)* specific for a microbe, *Helicobacter hepaticus,* whose antigen is absent from our animal facility did not contain many DN T cells (Supplemental figure 4, E and F), suggesting that antigen stimulation is likely required for development or retention in the gut. To verify that DN T cells still responded to their antigen in a TCR/MHCII-dependent manner, cells were stimulated *in vitro* with peptide loaded antigen presenting cells (APCs). Sorted DN Cbir1 transgenic T cells indeed responded to their cognate peptide, and their proliferation could be blocked with an MHCII blocking antibody (Figure 3E). To further demonstrate that proliferation to the native antigen was MHCII dependent, APCs were fed faeces from steady-state or vancomycin treated mice overnight before T cell introduction, and proliferation was again blocked by MHCII blocking antibodies (Figure 3F). This demonstrated that our facility contained the antigen the T cells

392 respond to and that proliferation was MHCII dependent. Taken together, these data suggest
393 T_{MIC} cells can develop in the mouse.
394



395 Figure 3. Cbir1, MHCII restricted, microbe-reactive TCR transgenic develop an innate-like phenotype in the gut

396 **Figure 3. Cbir1, MHCII restricted, microbe-reactive TCR transgenic develop an innate-like**
397 **phenotype in the gut**

398 LPLs from Cbir1 TCR transgenic line crossed to Rag-/- mice were analysed directly *ex vivo* or
399 stimulated *in vitro*.

400 (A) Flow cytometric analysis of TCRb+lin- cells from the colon lamina propria of Cbir1 Rag-/- mice
401 stained for CD4 and CD8a. (left) and IL-7R and IL-18R stain of the gated CD4 (green) and DN (blue)
402 cell populations (right) (B) Quantification of flow cytometry data showing the percent of the TCRb+lin-
403 population that are CD4, CD8a, and DN (left), and percent of each population that have the IL-18R/IL-

7R expressing phenotype (right). Each point represents an individual animal. N=5, representative of two independent experiments. (C) Quantification of FACS data of the percent of T cells with a DN phenotype from Cbir1 colon LPL, lung, MLN, spleen, and thymus. Each point represents an individual animal. N= 4, representative of two independent experiments. ANOVA with post-test comparing each mean to colon. (D) Representative flow cytometry data (left) and quantification (right) of IL-7R and PLZF expression within the CD4 and DN T cell subsets with fluorescence minus one (FMO) control (right). (E) Representative flow cytometry histograms of sorted CD4 or DN splenic Cbir1 Rag-/- T cells labelled with CFSE and incubated for five days with peptide pulsed bone marrow derived DCs (BMDCs). (F) Representative flow cytometry histograms of sorted DN T cells from Cbir Rag-/- mice stained with CFSE and incubated for five days with BMDCs fed faeces of steady-state or vancomycin treated mice from the same animal facility.

Mouse and human T_{MIC} cells share a transcriptional program

In order to determine how similar human and murine T_{MIC} cells are, we performed RNA sequencing on sorted intestinal T cell populations of interest and several control populations. We merged human and murine RNA-sequencing datasets based on common gene ids as described before[37, 52]. This approach allowed us to initially focus on genes present in both datasets, enabling us identify cell populations with similar transcriptional features based on clustering and principal component analyses (PCA). To that end, we merged a human dataset containing four CD4 T cell populations differing in their expression of CD161 (Supplemental Figure 4, A), a murine dataset containing colonic ab TCR CD4 and double-negative T cells (Supplemental Figure 4, B) as well as iNKTs and a selection of Immgen-derived T cell datasets[53] as comparators. Strikingly, $CD161^{int}$ and $CD161^{-}$ human CD4 T cells clustered together with murine CD4s based on the first two components of our PCA (Figure 4, A) or hierarchical clustering (Supplemental Figure 4, C), while $CD161^{hi}$ CD4s clustered closer to murine iNKTs and DN T cells, confirming the presence of transcriptional features shared with innate-like T cells described before and the existence of a transcriptional program shared between human and murine T_{MIC} cells. Since the main source

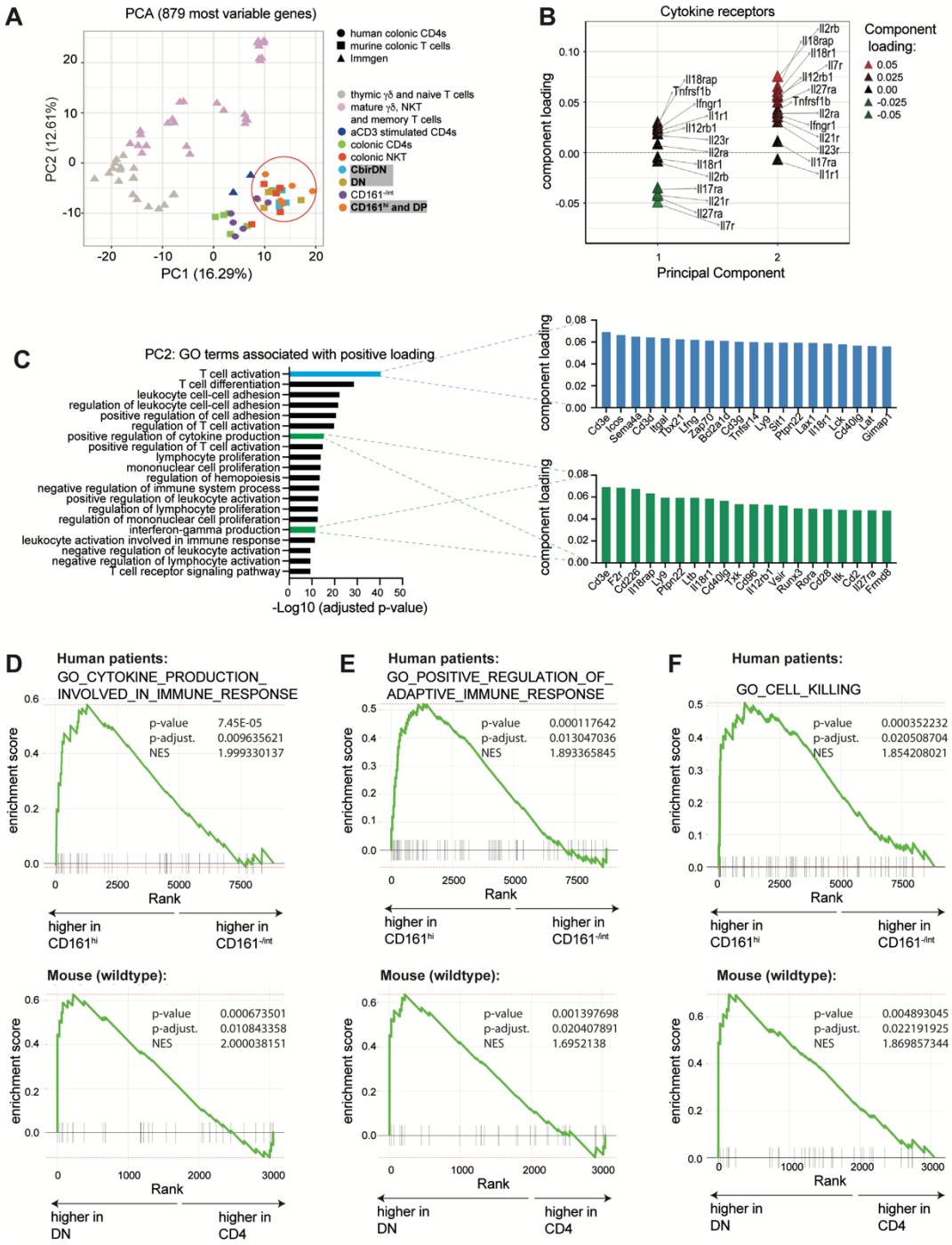
of variance (PC1) in our merged dataset separated most the Immgen-derived samples from ours, with the exception of samples which had undergone *in vitro* αCD3 treatment, we wondered if the inclusion of these data might have a confounding impact on the way human and murine colonic T cells clustered amongst each other. Restricting our analysis to human and murine colonic T cells however lead to similar results, as murine DN T cells and human CD161^{hi} CD4 T cells were still split apart from their respective counterparts by PCA and clustered together (Supplemental Figure 5, D).

To explore the shared features of microbe- and cytokine-responsive T cells found in mouse and man, we analysed which genes were driving the clustering in our PCA (Figure 4 A, Supplemental Figure 5, C, Supplemental Table 3). In line with the fact that PC1 in our merged dataset separated samples originating from spleen, lymph nodes and the thymus from activated and colon-derived T cells, a GO-term analysis of the genes positively associated with it indicated that processes linked to activation, differentiation and adhesion of T cells were the main drivers. Overall, GO: 0042110 “T cell activation” turned out to be the GO term most significantly associated with PC1 (Supplemental Figure 5, E, Supplemental Table 4). On the other hand, across the entire merged dataset PC2 separated more mature, effector-like T cell types (iNKTs, γδT cells, CD161^{hi} CD4 T cells and DN) cells from others. Interestingly, a PCA-loading analysis revealed that most cytokine-receptor genes contained in the dataset were positively associated with this PC (Figure 4, B), providing a molecular basis for the observed cytokine-responsiveness of the human and murine cell subsets towards IL-23, IL-12 and IL-18 and suggesting that it might extend beyond those measured by flow cytometry. This was further corroborated by the fact that 2 of the 20 GO terms (Supplemental Table 4) most significantly associated with the genes driving this PC are linked to cytokine responsiveness and production (GO:0001819 “positive regulation of cytokine production” and 0032609 “interferon-gamma production”) and include the genes

458 *Il18rap*, *Il12rb*, *Il18r1* and *Ifngr1* (Figure 4, C). Interestingly, GO: 0042110 "T cell activation"

459 again was the most positively associated term, albeit driven by other genes than for PC1.

460



461

462 **Figure 4: Human and murine T_{MIC} cells share transcriptional features and show an enrichment**

463 **of effector-related genes. See also Supplemental Figure 5 and Supplemental Figure 6.**

Colonic human CD161^{hi} and murine DN T cells as well as several control populations were sorted and subjected to bulk RNA-sequencing. The depicted analyses were performed after filtering for non- and lowly expressed genes.
(A) Principal component analysis performed on the 866 most variable genes (IQR > 0.75); plotted are the scores of the first two principal components (PC). Genes were obtained from a merged dataset consisting of human and murine T cell populations derived from experiments described in this study or obtained from Immgen as indicated. Human and murine datasets were merged based on orthologue genes. (B) Loading analysis of the principal components from (A), illustrating the contribution of genes encoding cytokine-receptors to the overall variance described by PC1 and 2. (C) Analysis showing the top 20 Gene Ontology (GO) terms enriched in the genes positively contributing to PC2 from (A). For the highlighted GO-terms, the respective top 20 genes from PC2 are shown. (D - F) GSEA plots depicting the enrichment of the three indicated GO-terms in human (CD161^{hi} CD4) and murine (DN) T_{MIC} cells compared their respective control populations.

An important limitation of the data merging method we employed is that genes not present in all of the individual datasets or genes that do not have a direct ortholog in humans or mice respectively, get removed during the merging process. To test whether the characteristic features we identified in our merged dataset are also present in the complete original datasets we separately compared human CD161^{hi} CD4 T cells to their CD161$^{int/-}$ counterparts and murine wildtype or Cbir DN to murine CD4 T cells. Based on differential gene expression analyses, we obtained lists of genes associated with human CD161^{hi} CD4 T cells (human T_{MIC} gene module, Supplemental Table 2) or DN T cells in mice (murine T_{MIC} gene module, Supplemental Table 5). Expression of these gene modules was tested in the respective other species by GSEA, revealing enrichment of the murine T_{MIC} gene module in human CD161^{hi} CD4 T cells and of the human T_{MIC} gene module in murine wildtype and Cbir DN T cells respectively (Supplemental Figure 6, A and B). Genes driving the enrichment of the T_{MIC} gene modules included the human and murine orthologues of IL23R as well as those of the transcription factors RORA, ID2 and BHLHE40. Other notable genes distinguishing human and murine T_{MIC} cells from CD161$^{int/-}$ or CD4 T cells respectively were

the motility factor S100A4, the pH-sensor GPR65 and the GPI transamidase PIGS (Supplemental Figure 6, C and D). Additional GSEAs revealed that human CD161^{hi} CD4 and murine DN T cells show an enrichment in GO terms and genes associated with cytokine production and regulation of the immune response (Figure 4, D and E and Supplemental Figure 5, E). We also found that both cell populations also showed an enrichment of the GO term 0001906 “cell killing” (Figure 4, F), not identified in our merged dataset. This finding is in line with the capability of CD161^{hi} CD4 T cells to produce GzmB, as seen earlier (Figure 1, E), confirming and expanding the array of potential effector functions that can be attributed to human and murine T_{MIC} cells. In contrast to this, GO:0042110 “T cell activation” did not show a significant enrichment in any of the analysed populations (Supplemental Figure 6, F). This is in line with the fact that it was associated with both PC1 and 2 in our analysis, suggesting that genes associated with this term are to some degree upregulated in colonic T cells in general.

T_{MIC} cells are present in the gut in human and murine colitis

Given the large range of potential effector functions that human and murine T_{MIC}s possess, we next wanted to explore their functional impact in a disease setting.

In biopsies obtained from inflamed tissue of ulcerative colitis (UC) patients, we found that CD4 T cells in general increase in numbers. The largest increase was seen for the CD161^{-} and CD161^{int} subsets, but we also noted that cells with a CD161^{hi} phenotype are present in equal or even slightly elevated numbers per gram of tissue compared to samples obtained from non-inflamed or normal tissue (Figure 5, A). We also observed persistence of this cell type in two mouse models of colitis (Supplemental Figure 7, A) and the analysis of a previously published dataset containing single cell data from healthy controls and IBD patients[46] showed higher expression of a gene module containing the T_{MIC}-associated genes *ZBTB16*, *KLRB1*, *IL18R1* and *IL23R* in CD4 T cells from non-inflamed and inflamed tissue samples compared to healthy controls (Supplemental Figure 7, B), supporting the idea that T_{MIC} cells are present in inflamed colitic tissue. Further, compared to cells isolated from non-

inflamed human tissue, T_{MIC} cells in UC showed higher expression of inhibitory receptors like TIGIT, TIM3, LAG3 and CD39 (Figure 5, B), even more pronounced than their $CD161^{-}$ and $CD161^{int}$ counterparts (Supplemental Figure 7, C). These observations suggest that human T_{MIC} cells are triggered in the context of IBD, which was also supported by increased expression of CTLA4 and TIGIT in CD4 T cells from inflamed IBD tissue when comparing cells expressing the aforementioned T_{MIC}-associated gene module (Supplemental Figure 7, D).

Taken together with the high effector potential and microbe-responsiveness displayed by T_{MIC} cells, these findings suggest that T_{MIC} cells might be involved in UC pathology.

T_{MIC} cells contribute to colon pathology in murine models of colitis

To determine whether microbe-reactive T cells can contribute to colonic pathology, we returned to the Cbir1 transgenic mouse model, which recognizes a microbe present in the normal mouse microbiota. While transferring isolated tissue-resident cells into another host to study their behaviour would be the ideal experiment, as with other tissue-resident cell populations, DN cells isolated from the colon would not repopulate the gut of a recipient mouse, even in the case of Rag-/- where there is no competition from endogenous T cells (Supplemental Figure 8, A-C). To get around this issue and study the innate-like T cell population in isolation, we compared Cbir1 Rag-/- mice treated with anti-CD4 to deplete Th and lymphoid tissue inducer cells (~90% DN T cells) with anti-CD4 treated Rag-/- (no T cells) mice (Supplemental Figure 8, D and E). To understand whether these cells could contribute to pathogenesis when they encounter their antigen in a proinflammatory environment, we used the dextran sulphate sodium (DSS) model of colitis, which results in barrier breakdown and bacterial translocation. When challenged with DSS, CD4-depleted Cbir1 mice lost significantly more weight than their CD4-depleted Rag-/- littermates at the peak of disease (Figure 5 C and D). Hallmarks of systemic disease was associated with a trend toward increased colon pathology marked by epithelial damage and crypt abscesses

(Figure 5F and G). Systemic inflammation was further marked by increased splenomegaly in the Cbir Rag-/- mice, bearing T_{MIC} (Figure 5, E). To understand the differences in disease on a tissue level, RNA was isolated from colonic tissue, and qPCR was performed for genes of interest associated with T_{MIC}. The presence of T_{MIC} in the Cbir mice resulted in significantly increased levels of *Tnf, Il1b, and Csf2* (GM-CSF) (Figure 5, H) but not other pro-inflammatory cytokines or the repair associated gene *Areg* (Figure 5, I) suggesting a specific response module in the context of DSS. Taken together, these data suggest that we have identified an as-yet unappreciated population of commensal and cytokine-responsive T cells that can contribute to intestinal pathogenesis in the context of barrier breakdown.

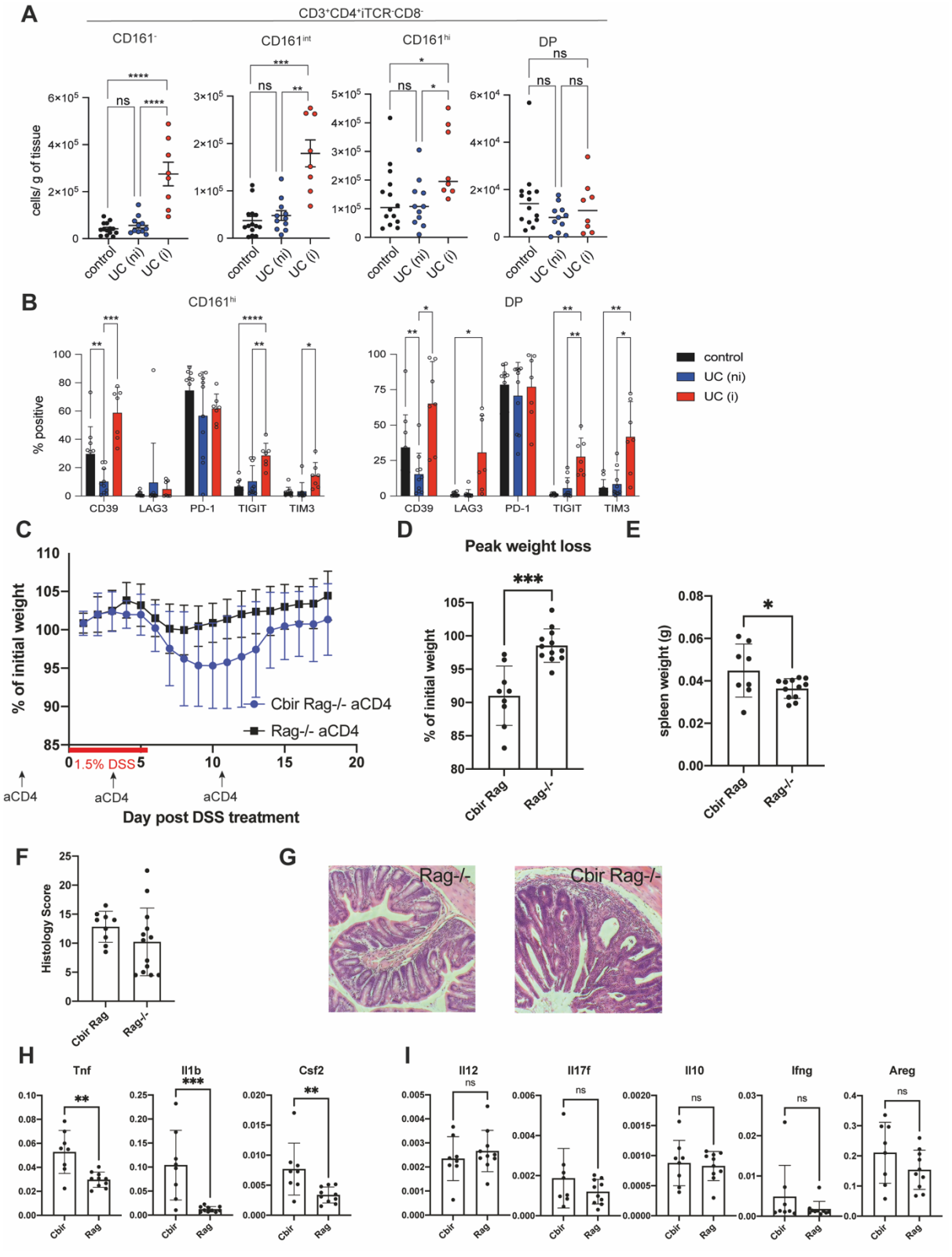


Figure 5: T_{MIC} cells are present in the inflamed tissue of human UC patients and exacerbate pathology in a murine colitis model. See also Supplemental Figure 6 and Supplemental Figure 8.

(A) LPMCs were isolated from resections of normal tissue from colorectal cancer patients (control) or from biopsies obtained from UC patients. UC samples were classified as "non-inflamed" (UC(ni)) or inflamed (UC(i)) based on their UCEIS score. All samples were weighed, the total numbers of iTCR (Vα7.2, Vα24-Jα18, TCRγδ)-negative CD4 T cells expressing $CD161^{-}$, $CD161^{int}$, $CD161^{hi}$ and DP CD4 T cells in each sample recorded by flow cytometry and then normalized to the weight of the respective sample in grams. (B) Bar plots summarizing the expression of CD39, LAG3, PD-1, TIGIT and TIM3 as determined by flow cytometry on the human T_{MIC} cells from (A). (A,B) Data points were pooled from independent experiments using one or two human samples each. (C) Weight loss curve of Cbir Rag-/- or Rag-/- littermates pre-treated with anti-CD4 antibody and given 1.5% DSS in drinking water for 6 days. (D) Quantification of peak weight loss for each animal in the Cbir Rag-/- and Rag-/- controls. (E) Spleen weights at DSS experiment endpoint. (F) Quantification of colon histology score at DSS experiment endpoint. Each point represents one individual mouse. n=9 Cbir, n=12 Rag-/-, showing a combination of two independent experiments. (G) Example H&E staining showing unique features of epithelial damage in Cbir Rag-/- after DSS. (H and I) qPCR of gene expression relative to HPRT for colonic tissue isolated on day 18 after DSS administration.

$^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$, $^{****}p < 0.0001$; Kruskal-Wallis test with Dunn's multiple comparisons test (A), mixed effects analysis with Tukey's multiple comparisons test (B). Mean ± SEM is shown.

Discussion

Commensal-immune interactions are key in maintaining homeostasis in the intestine. Using human tissue samples and mouse models, our data demonstrate the acquisition of innate-like features by commensal-reactive, MHC-II-restricted T cells in the colon. These characteristics place human and murine T_{MIC} cells on the expanding array of T cell populations bridging innate and adaptive immunity. T_{MIC} lie between established innate like T cell subsets such as MAIT and iNKT cells and classic tissue resident memory T cells as a population expressing innate-like factors such as PLZF and high levels of CD161 (in

humans), while featuring a diverse, MHC-II restricted TCR repertoire at the same time. Located at mucosal barriers, T_{MIC} cells are likely to play an important role in microbial surveillance and shaping the local cytokine milieu.

T_{MIC} cells are potent producers of effector molecules including several key Th17-cytokines. This is in line with previous studies that described the broader population of human colonic CD161+ CD4 T cells as a Th17 population in the context of Crohn's disease[23] and observed IL-17 production by murine DN T cell in response to intracellular pathogens[48] and human DN T cells in lupus[50]. Our findings expand on this, as we demonstrate that especially the $CD161^{hi}$ subset is not restricted to Th17 functions but displays a dual Th1/Th17 effector profile much like human innate-like T cells[54] and is capable of producing factors associated with cytotoxicity. Notably, despite the differences in CD4 expression, these functional properties were conserved in murine T_{MIC} cells. As we were analysing cells derived from normal, uninflamed tissue in this study, it is conceivable, that in the context of human inflammatory disease, T_{MIC} cells would get skewed more towards a Th17-phenotype while our findings represent a more homeostatic state representative of healthy tissue. Indeed, several recent publications have found that in human intestinal CD4 T cells, effector phenotypes exist as a gradient across the population rather than as distinct subsets[55, 56, 57], supporting that idea that many of these cells have the potential to produce a wide range of effector molecules.

Mucosal tissues are rich in T_{RM} cells which arise in response to infection or vaccination and confer local protection in the event of re-encounter with a given pathogen. Interestingly, analyses of single cell sequencing data revealed that the expression of an innate-like gene expression profile does not correlate tightly with the expression of established T_{RM} gene signatures, indicating that the T_{MIC} phenotype constitutes a special status and not part of the general T_{RM} program in the colon. While the pathways leading to T_{MIC} formation are not fully understood at the moment, it is worth pointing out that most T_{RM} models make use of acute

infections such as LCMV and HSV or precisely controlled challenges with model antigens or vaccines to induce T_{RM} cells[44, 45, 58, 59], meaning that antigen exposure is intense and transient. In contrast, T cells responding to commensal microbes would be triggered with limited amounts of antigen but multiple times over a prolonged period of time or even constitutively. While future studies are required to fully address this question, this model would be supported by the observed differences between microbe-reactive Cbir- or Hh-TCR transgenic T cells. While the former respond to a ubiquitous commensal antigen and acquire a T_{MIC} phenotype in the gut, the latter, responding to a bacterium not normally present, display a typical CD4 phenotype at steady-state and maintain a conventional CD4 phenotype when transferred into an infected host[11]. Further work with other microbe-reactive TCR transgenics[9] may shed light on the ability of other microbes to drive this phenotype.

Existing data indicate that the local factors specific to mucosal tissues play an essential role in inducing the T_{MIC} phenotype as Cbir T cells located in non-mucosal organs do not have an innate-like phenotype, neither do other commensal-specific T cells in thymus[60]. This suggests that T_{MIC} cells are not a predefined lineage or an accident of altered selection in T cell transgenic mice but rather the result of adaption to local conditions. This is supported by the finding that murine T_{MIC} cells can be found in all mucosal tissues examined. While future experiments are required to identify which antigen-presenting cells, cytokines and signalling mechanisms are required for T_{MIC} induction, we hypothesize that local myeloid populations in the lamina propria could be involved. These antigen presenting cells have the potential to stimulate T_{MIC} cells through TCR and IL-23R simultaneously while cytokines from epithelial cells such as IL-18 may tune T_{MIC} cells to produce a different set of responses.

Our results confirmed several other studies reporting the colitogenic potential of microbe-reactive T cells[60, 61, 62] and are further supported by the fact that human T_{MIC} cells are present in inflamed tissue of ulcerative colitis patients in normal or even enhanced numbers. However, the role T_{MIC}s play in the lamina propria likely goes beyond that. Innate-like T cells

646 including MAIT cells were recently shown to possess tissue repair capacities[37, 38, 39] in
647 humans and mice and since T_{MIC} cells share transcriptional features with these populations,
648 T_{MIC} cells could perform similar function under the right circumstances. In MAITs, whether
649 activation is induced by TCR or cytokine-signalling is critical in determining if they express
650 repair-associated factors or a pure pro-inflammatory program respectively[36, 37]. T_{MIC} cells
651 might operate in a similar way and in the steady-state, triggered in a limited fashion by
652 commensal-dependent TCR-mediated signals, could actually contribute to tissue
653 homeostasis and the repair of limited injuries. In contrast, in the context of wide-spread
654 inflammation, e.g. after a major breach of the epithelial barrier like in the DSS colitis model
655 we tested, cytokine- and combined cytokine- and TCR-mediated activation would likely
656 dominate, leading to massive production of Th1 and Th17 effector molecules contributing to
657 colitis.
658
659 T cells with a T_{MIC} phenotype are not a rare cell population in the lamina propria and are
660 especially abundant in humans. Given their high effector and colitogenic potential as well as
661 their responsiveness to microbes, future studies addressing how exactly these cells are
662 induced and regulated and how they behave in the steady state are needed. Fully
663 understanding their function could lead to new therapeutic targets in relevant human
664 diseases like IBD and checkpoint induced colitis and could have important implications for
665 patients undergoing any kind of microbiome therapy or suffering from conditions associated
666 with alterations in the microbiome.
667

668 **Acknowledgments**

669 C-PH was supported by a Research Fellowship (403193363) from the Deutsche
670 Forschungsgemeinschaft (DFG, 2018-2020), the Wellcome Trust (222426/Z/21/Z, awarded to PK)
671 and a Medical Sciences Internal Fund (Pump-Priming: award 0009784).
672 HHU acknowledges research support from the Helmsley Charitable Trust and the NIHR Biomedical
673 Research Centre Oxford.

FP was supported by Wellcome Trust (095688/Z/11/Z and 212240/Z/18/Z).

PK was supported by the Wellcome Trust (222426/Z/21/Z), the National Institute for Health Research (NIHR) Biomedical Research Centre (BRC), a NIHR Senior Fellowship and the National Institutes of Health (NIH, U19 AI082630).

ET was supported by Wellcome Trust (095688/Z/11/Z and 212240/Z/18/Z, awarded to FP), Nuffield Department of Medicine, MRC core grant reference MC_UU_00008, and the University of Oxford COVID Rebuilding Research Momentum Fund (CRRMF).

We would like to thank Luke Barker for supporting animal husbandry and experimentation throughout the project. We thank Charles O. Elson III for the Cbir1 mouse strain, Dan Littman for the Hh7-2 mouse strain, and Dan Cua for the IL-23RGFP mouse strain. We want to thank Dr Ida Parisi, Dr Bryony Stott and Miss Rhiannon Cook in the Kennedy Institute of Rheumatology Histology Service for tissue processing and staining. We thank the Oxford Genomics Centre at the Wellcome Centre for Human Genetics (funded by Wellcome Trust grant reference 203141/Z/16/Z) for the generation and initial processing of sequencing data. The following reagent(s) was/were obtained through the NIH Tetramer Core Facility: CD1d and MR1 tetramers. The MR1 tetramer technology was developed jointly
by Dr. James McCluskey, Dr. Jamie Rossjohn, and Dr. David Fairlie, and the material was produced by the NIH Tetramer Core Facility as permitted to be distributed by the University of Melbourne. We want to thank Helen Ferry for flow cytometry panel design and cell sorting at the EMD/TGU flow cytometry facility and Neil Ashley for RNA-sequencing at the WIMM Single Cell Genomics Facility conducted as part of a Human Immune Discovery Initiative (HIDI) project awarded to C-PH.
We acknowledge the contribution of the Oxford BRC Gastrointestinal Biobank supported by the NIHR BRC. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health. We thank members of the Oxford TGU biobank, especially J. Chivenga, A. Isherwood, R. Williams and M. Cabrita for facilitating the collection of patient samples.
We further acknowledge the contribution of the members of the Oxford IBD Investigators consortium: Dr Carolina Arancibia, Dr Adam Bailey, Professor Ellie Barnes, Dr Noor Bekkali, Dr Elizabeth Bird-Lieberman, Dr Oliver Brain, Dr Barbara Braden, Dr Jane Collier, Professor James East, Dr Lucy Howarth, Professor Paul Klenerman, Professor Simon Leedham, Dr Rebecca Palmer, Dr Fiona

Powrie, Dr Astor Rodrigues, Dr Francesca Saffioti, Professor Alison Simmons, Dr Peter Sullivan, Professor Holm Uhlig, Professor Jack Satsangi, Dr Philip Allan, Dr Timothy Ambrose, Dr Jan Bornschein, Dr Jeremy Cobbold, Dr Emma Culver, Dr Michael Pavlides and Dr Alissa Walsh

This work benefited from data collected by the ImmGen consortium[53].

Author Contributions

C-PH contributed to conceptualization, data curation, formal analysis, investigation, methodology, project administration, validation, visualization, writing-original draft, review and editing. DC contributed to investigation, validation, formal analysis, and writing - review and editing. LD contributed to investigation, validation, and formal analysis. CP provided conceptualization, resources, investigation, project administration, and writing - review and editing. SB contributed to investigation, project administration, and writing - review and editing. NI contributed to data curation, formal analysis, and writing- review. HAD contributed to investigation. YG contributed to investigation and writing-review and editing. MEBF contributed to investigation and writing - review and editing. OJH contributed to investigation, resources, and writing-review and editing. LCG contributed to investigation, methodology and writing - review and editing. EHM contributed to investigation and writing - review and editing. SP contributed to investigation. MF contributed to investigation and writing - review and editing. NMP contributed to supervision and writing - review and editing. HU contributed to supervision – review and editing. EM contributed to methodology, validation and visualization. Oxford IBD investigators provided resources and contributed to funding acquisition and project administration. FP contributed to conceptualization, funding acquisition, supervision, validation, and writing-review and editing. PK contributed to conceptualization, funding acquisition, supervision, validation, and writing-review and editing. ET contributed to conceptualization, data curation, formal analysis, investigation, methodology, project administration, supervision, validation, visualization, writing-original draft, review and editing.

Declaration of Interests

FP received consultancy or research support from GSK, Novartis, Janssen, Genentech and Roche. HHU has received research support or consultancy fees from Janssen, UCB Pharma, Eli Lilly, AbbVie, Celgene, OMass and MiroBio. PK has done consultancy for UCB and Medimab.

Materials and Methods

Experimental model and subject details

Human samples

Normal adjacent tissue from colorectal cancer (CRC) patients who were undergoing surgery was collected by the TGU biobank. Biopsies from ulcerative colitis (UC) patients were from patients attending the John Radcliffe hospital. All tissue samples were collected with appropriate patient consent and NHS REC provided ethical approval (reference number 16/YH/0247).

Characteristics of the CRC patients

CRC patients	n = 49
Age (years, average, SD)	68 ± 12
Sex (Male/Female)	28/21
Time since diagnosis (months, average, SD)	1 ± 4

Characteristics of the UC patients

UC patients		n = 20
Inflamed		n = 9
	Age (Average, SD)	38 ± 19

	Sex (Male/Female)	4/5
	Time since diagnosis (years, average, SD)	14 ± 12
	UCEIS score (average, range)	3.7, [2-6]
Uninflamed		n = 11
	Age (Average, SD)	53 ± 11
	Sex (Male/Female)	4/7
	Time since diagnosis (years, average, SD)	17 ± 13
	UCEIS score (average, range)	0.1, [0-1]

Mice

Mice were bred and maintained in the University of Oxford specific pathogen free (SPF) animal facilities. Experiments were conducted in accordance with local animal care committees (UK Scientific Procedures Act of 1986). B2Mko mice were obtained from The Jackson Laboratory and maintained by Oliver Harrison at the Benaroya Research Institute. Mice were routinely screened for the absence of pathogens and were kept in individually ventilated cages with environmental enrichment. The C57BL/6 age cohort was purchased from Charles River. Cbir1 TCR transgenic mice were a kind gift from Charles Elson III. *Il23r*$^{gfp/+}$ were obtained from Daniel Cua (Merck Research Laboratories, Palo Alto, USA). Hh7-5 TCR transgenic mice were a kind gift from Dan Littman. Backcrosses to B6J were verified by SNP analysis performed by Transnetyx. Mice were age and sex matched, using equal numbers of each sex where possible with the exception of DSS experiments. DSS experiments were performed with females to reduce cage numbers and potential confounding pathology from fighting. Mice were divided into cages, ensuring all genotypes were present in each cage, and experimenters were blinded to genotype for the duration of the experiment.

Data availability

All RNA sequencing datasets are available via permanent link below. Datasets will be made publicly available upon publication.

Human datasets: https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-11440. Reviewers can have anonymous access through the login Username: Reviewer_E-MTAB-11440 Password: uvEHoooe

Mouse datasets: https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-11397. Reviewers can have anonymous access through the login Username: Reviewer_E-MTAB-11397 Password: HE269zxc

The Rhapsody single cell data have been uploaded to the Gene Expression Omnibus repository (accession number: GSE2017159).

Code availability

R scripts allowing the reproduction of the data will be made available upon request.

Method details

Processing of CRC tissue resections

Fat and muscle tissue was superficially removed using sterile scissors and forceps before the mucosa was washed in 1mM DTT (Sigma-Aldrich) dissolved in PBS (GIBCO) supplemented with 40µg/ml genticin (Thermo Fisher), 10µg/ml ciprofloxacin, 0.025µg/ml amphotericin B and 100 U/ml penicillin, 0.1mg/ml streptomycin (all from Sigma-Aldrich) for 15 min in a shaker set to 200 rpm at 37° Celsius. To remove epithelial cells, samples were subsequently washed three times in the same PBS-based buffer containing 5mM EDTA (VWR). Finally, specimens were weighted, cut into small pieces, and

789 either directly digested or stored on 1ml freezing medium (90% FCS, 10% DMSO) at -80° for later
790 usage (0.2 – 0.4 g per vial).

791 **Direct digestion of resection-derived tissue**

792 For the assays aiming to detect microbe-reactive CD4 T cells, resection-derived tissues were directly
793 digested by transferring the dissected tissue into a RPMI 1640-based digestion medium containing
794 10% FCS (Sigma-Aldrich), 40µg/ml genticin (Thermo Fisher), 10µg/ml ciprofloxacin, 0.025µg/ml
795 amphotericin B, 100 U/ml penicillin, 0.1mg/ml streptomycin (all from Sigma-Aldrich), 0.1 mg/ml
796 collagenase A and 0.01mg/ml DNase I (both from Roche). HEPES (10mM, GIBCO) was freshly
797 added to the buffer before the digestion. The specimens were incubated for 20 minutes in a shaker
798 set to 200 rpm at 37° Celsius and filtered over a 100µm Cell strainer to isolate liberated cells. This
799 process was then repeated several times until the tissue had been fully digested and no more
800 additional cells could be obtained. Liberated cells were collected and pooled in R10 medium (RPMI
801 1640 containing 10% FCS, 100 U/ml penicillin, 0.1mg/ml streptomycin and 2mM L-glutamine) and
802 stored at 4° Celsius. To enrich mononuclear cells, digests were centrifuged for 20 minutes in a two-
803 layer Percoll (Sigma-Aldrich) gradient and collected at the 40%/80% interphase.

804 **Generation of single cell suspensions from stored tissue and biopsies**

805 To remove the DMSO-containing freezing medium, resection-derived tissue samples were washed
806 over a 70µm Cell strainer with pre-warmed R10 medium. Washed resection specimens or biopsies
807 obtained from UC patients or healthy controls were transferred into gentleMACS C tubes (Miltenyi)
808 and 5 ml of and RPMI 1640-based digestion medium containing 1mg/ml collagenase D and
809 0.01mg/ml DNase I (both from Roche) was added. Samples were homogenized on the gentleMACS
810 (Miltenyi) with the brain01_02 program and incubated for 60 minutes in a shaker set to 200 rpm at 37°
811 Celsius. All specimens were then fully disrupted on the gentleMACS by running program B01 and the
812 content of the C tubes was filtered into R10 medium over 70µm cell strainers. Cell were washed two
813 times in R10 medium and again, mononuclear cells were enriched by density Percoll (Sigma-Aldrich)
814 centrifugation and collected at the 40%/80% interphase.
815

Detection of microbe reactive human CD4 T cells

Mononuclear cells from directly digested colonic resections were counted and adjusted to 10^7 cells/ml in R10 medium and plated out on 96U-plates, 10^6 cells per well. In some experiments, cells were incubated with 10µg/ml ULTRA-LEAF purified mouse anti human MHCII (Tü39) or IgG2ακ (MOPC-173) control antibodies (Biolegend) for 30 min at 37° Celsius to evaluate the MHC II-dependency of the observed responses. Next, cells were mixed with heat-killed microbes: *E.coli* Nissle (Ardeypharm GmbH), *S. aureus* (National collection of type cultures 6571) and *C.albicans* (InvivoGen) in a 1:10 ratio. After two hours of incubation at 37° Celsius, Brefeldin A (Invitrogen) was added to the cells to block cytokine release and then incubation as continued for 6 additional hours. Finally, cells were washed, pelleted by centrifugation and microbe-reactive cells were detected by FACS staining for CD154 (24-31) and TNF (MAb11). Both antbodies were purchased from Biolegend and used at 1:50.

TCR and cytokine stimulation of human CD4 T cells

To provide TCR stimulation to human CD4 T cells, NUNC Maxisorp plates (BioLegend) were coated with purified anti human CD3 antibodies (BioLegend) diluted to 1.25µg/ml in sterile PBS for 2h at 37° Celsius or at 4° Celsius overnight. Control wells were filled with PBS only. Plates were washed two times with PBS and one time with R10 medium before cells were added.
Mononuclear cells were adjusted to either $4x10^6$ cells/ml in R10 medium and 100µl cell suspensions were added to the appropriate well of the coated plates. Purified anti human CD28 antibodies (BioLegend) and added to all wells previously coated with anti CD3 antibodies (OKT3) to a final concentration of 1µg/ml. Recombinant human IL-23 (Miltenyi) was diluted to 400ng/ml in R10 medium and added to the cells to a final concentration of 50ng/ml.
For stimulation with IL-12 (Miltenyi) and IL-18 (BioLegend), cells were plated out at a concentration of 10^7 cells/ml in R10 medium on regular 96U plates and mixed with recombinant cytokines (final concentration 50ng/ml for both IL-12 and IL-18). In all experiments, some cells were left untreated to determine baseline expression of the analysed effector molecules.
Cells were incubated with the different combinations of antibodies or cytokines for 24 hours total. Brefeldin A solution (Invitrogen) was added to all well for the last 4 hours to prevent cytokine release.

FACS

Single cell suspensions were initially stained with the LIVE/DEAD Fixable Near IR Dead Cell dye (Invitrogen) diluted 1:1000 in PBS for 20 min at room temperature. Cells were then washed and stained in PBS supplemented with 0.5% FCS and 2mM EDTA with the antibodies purchased from Biolegend, BD, Miltenyi, Invitrogen or Thermo Fisher for 20-30 min at room temperature.
Prior to subsequent staining steps or acquisition, cells were fixed in 2% Formaldehyde (Sigma-Aldrich) for 10 min at room temperature. For intracellular cytokine staining, cells were permeabilized and stained with the appropriate antibodies using the BD Cytofix/Cytoperm Kit following the manufacturer instructions. Intranuclear staining for transcription factors was performed using the Foxp3/Transcription Factor Staining Buffer Kit (Invitrogen) according to the manufacturer's instructions. The following antibodies were used to stain human cells: CCR7 (G043H7), CD3 (UCHT1), CD4 (OKT4), CD8 (SK1), CD39 (A1), CD45RO (UCHL1), CD45RA (HI100), CD56 (NCAM16.2 and HCD56), CD154 (32-31), CD161 (HP3G10 and 191B8), GzmB (GB11 and QA16A02), ICOS (REA192), IFNγ (4S.B3), IL-17A (BL168), IL-17F (eBio18F1 and SHLR17), IL18Rα (H44), IL-22 (22URTI), LAG3 (7H2C65), PD-1 (29F.1A12), PLZF (R17-809), TCR Vα7.2 (3C10), TCR Vα24-Jα18 (6B11), TCRγδ (B1 and 6B11), TIGIT (A15153G), TIM3 (F38-2E2), TNF (MAb11), RoRγt (Q21-559).
The following antibodies were used to stain murine cells: CD45R/B220 (RA3-6B2), CD11b (M1/70), CD11c (N418), TCRγδ (GL3), CD45 (30-F11), CD127(A7R34), CD8a (53-6.7), CD4 (RM4-5), TCR β chain (H57-597), PLZF (Mags.21F7), RoRγt (Q31-378), CD25 (PC61), NK1.1 (PK136), CD218 (IL18Rα, P3TUNYA). PE- or APC-labelled mouse CD1d- and MR1- tetramers were obtained from the NIH Tetramer Core facility.

All cells were acquired on a LSR II or Fortessa (BD); data were analyzed using FlowJo software (Treestar).

Nucleic acid extraction and bulk TCR sequencing of human cells

TRIzol (Thermo Fisher) nucleic acid extraction was used to extract high purity RNA from sorted human colonic CD4 T cells. Briefly, after sorting, cells were centrifuged (500 g, 5 minutes), resuspended in 1 ml TRIzol, then frozen at -80°C until RNA extraction. For RNA extraction samples were brought to room temperature, mixed with 200 µl chloroform (Sigma-Aldrich) and centrifuged at

12500 rpm for 5 minutes. 500 µl of the aqueous phase was taken and RNA extracted using the RNAdvance Tissue Isolation kit (Agencourt). RNA was assessed for concentration and purity using the RNA Pico assay on a 2100 Bioanalyzer instrument (both Agilent). Bulk TCR repertoire sequencing was performed using the amplicon-rescued multiplex (ARM)-PCR method (iRepertoire Inc). Library generation was performed in-house according to the manufacturer's instructions. In brief, the extracted RNA, enzyme mix, and barcoded primer reaction mix were mixed on ice, followed by a combined RT-PCR and initial amplification step in the thermal cycler using the iRepertoire low-input protocol. Next, the products were purified using the kit's solid phase reverse immobilisation (SPRI) beads and ethanol washes, then eluted in water. This product was then combined with enzyme mix and universal primers for the second amplification step. The final product was purified again using SPRI beads and eluted in water. The quality, size distribution, concentration, and presence of contaminating primer dimers of the final product was assessed using several QC steps, including identification of a clear band of appropriate size on agarose gel electrophoresis, a spectral photometer (Nanodrop, Thermofisher Scientific), and the DNA 1000 kit using a 2100 Bioanalyzer instrument (Agilent). Libraries were quantified using the KAPA Library Quantification Kit (Roche) on a CFX96 Thermal Cycler instrument (Bio-Rad) before equimolar pooling. Samples were submitted to the Oxford Genomics Centre where a PhiX library spike-in was added (10%) due to the low diversity of the TCR library, before 300bp paired-end sequencing on an Illumina MiSeq instrument (WTCHG,University of Oxford) was run.

Bulk TCR repertoire analysis

Data processing of TCR repertoire libraries was performed using the iRepertoire analysis pipeline. In brief, reads were demultiplexed based on the 6-N molecular barcode associated with the sample. Low quality reads were trimmed (removing anything with a Phred score of less than 30), and R1 and R2 reads were overlapped and stitched. Only stitched reads where identity within the overlapped portions was 100% were included in downstream analysis. Reads were then mapped to the IMGT database, and only reads that map to reference sequences and contain canonical CDR3 motifs were included for further analysis. Finally, several filters (see irepertoire.com/irweb-technical-notes) were applied to remove sequencing artefacts, PCR artefacts, insertion, deletion, and substitution errors, and low frequency (n=1) reads. Initial data analysis was performed using the iRweb data analysis platform

(iRepertoire, Inc., USA). Additional analysis and generation of plots was performed using SeeTCR (friedmanlab.weizmann.ac.il/SeeTCR).

BD Rhapsody targeted single cell transcriptomics

Lamina propria mononuclear cells were isolated from three donors, stimulated with IL-12/18, plate-bound αCD3 or a combination of both as described above overnight. Samples were stained with oligonucleotide-conjugated Sample Tags from the BD Human Single-Cell Multiplexing Kit, a panel of 50 Abseq antibodies in BD stain buffer following the manufacturer protocol. Subsequently, cells were stained with a smaller panel of fluorescently labelled sorting antibodies (CD4, CD8, Vα7.2, TCRγδ, Vα24-Jα18, CD45) and LIVE/DEAD Fixable Near IR Dead Cell dye as for FACS experiments. 20.000-40.000 cells per donor and stimulatory condition were sorted on a BD ARIA III as live, $CD45^+$ $CD4^+$ $CD8^-$ $V\alpha7.2^-$ $V\alpha24\text{-}J\alpha18^-$ $TCR\gamma\delta^-$ in the Experimental Medicine Division Flow Cytometry Facility. Sorted cells were spun down (300g, 5min) and resuspended in 200µl chilled BD sample buffer and 20.000 cells were pooled together from all 12 samples and subsequently loaded onto a BD Rhapsody cartridge. Single cell capture and cDNA synthesis with a BD Rhapsody express system were performed using the manufacturer's reagents and protocols. In brief, the process included cell-capture with beads in microwell plate, followed by cell lysis, bead recovery, cDNA synthesis and library preparation using using the BD Rhapsody Targeted mRNA and Abseq Amplification kit. Separate libraries were prepared for sample tags, Abseq and targeted mRNA using a basic immune panel (BD Rhapsody Immune response Panel Hs) in combination with custom panel covering 145 additional genes (Supplemental Table 6). Importantly, the latter included genes associated with tissue-repair in a population of unconventional T cells in the murine skin[39] which were also be shown to be expressed in murine and human MAIT cells[36, 37]. Sequencing of the pooled libraries was completed on a NovaSeq6000 (Illumina, San Diego, CA) at Novogene (Cambridge, UK).

BD Rhapsody data analysis

The FASTQ-files generated from the Rhapsody experiment were uploaded onto the Seven Bridges Genomics online platform together with FASTA-files containing the sequence information about the

mRNA and Abseq targets and were subjected to the BD Rhapsody Targeted analysis pipeline. Data analysis was performed in SeqGeq (BD). Briefly, quality control was performed by gating out events with low gene expression, samples were demultiplexed using the Lex-BDSMK plugin to separate cells from the different stimulation conditions and CD161^{-}, CD161^{int} and CD161^{hi} cells were identified by gating using DNA-barcoded antibodies for CD161 and CD56. Gene expression profiles for the CD161^{hi} cells were exported and z-scores were calculated for tissue-repair associated genes upon cytokine, TCR or combined stimulation. Expression of the whole tissue-repair gene signature was analyzed by gene set enrichment analysis using the fgsea R package (see below).

Gene signature expression analysis in published datasets

Previously published colonic single cell data[46, 47] were re-analysed using version 4 of the Seurat R package [63] following the package's vignette. In brief, cells with abnormally high features numbers or high percentages of mitochondrial genes were removed, data were normalized (*method* = *LogNormalize* of the *NormalizeData* function) and scaled (ScaleData). Cell clusters were identified using the *FindClusters* and *RunUMAP* functions. To restrict the datasets to CD4 T cells, markers distinguishing the clusters were identified and visualized by the *FindAllMarkers* and *Featureplot* functions and based on the expression of CD3E, CD4, CD8B, CD8A and ZBTB7, the datasets were subsetted. The Smillie dataset was further subsetted to retain only cells annotated as "Healthy", excluding cells isolated from IBD patients.

The processed gene counts were extracted (*GetAssayData* function) and an expression-based ranking was built for all genes in each cell by using the *AUCell_buildrankings* function from the AUCell R package[64]. Files containing lists of genes associated with T_{RM} cells[44, 45] or CD161-expressing cells[27] were loaded and processed to be compatible with AUCell in R using the *getGMT* and *setGeneSetNames* functions from the GSEABase R package[65]. An additional control dataset (150 Genes random) was generated by randomly selecting 150 genes from the respective datasets. Areas under the curve (AU) values were calculated using AUCells *AUCell_calcAUC* function. To analyse the relationship between the different gene sets, the AUCell values were exported and the Pearson correlation was calculated in Prism.

Merging of human and murine RNA-seq datasets

The merging of sequencing data from different sources[52] and different species[36, 37] has been described before. In brief, a selection of RNAseq datasets from ImmGen[53] were merged with the human and murine RNA-sequencing datasets described in this study. In each dataset separately, zero-count and lowly expressed genes were initially removed from the raw read counts using the edgeR R package[66]. Raw counts were then log2-transformed using the *voom* function from the limma R package[67]. The murine orthologues of the genes present in the human dataset were identified using the *getLDS* function of the biomaRt package[68] and all datasets were merged based on common gene symbols. The *ComBat* function of the sva R package[69] was used to remove batch effects, as described previously[70]. The genes from the resulting merged dataset were filtered by variance (IQR > 0.75) and subjected to a principal component analysis and a hierarchical clustering analysis using the Euclidean distance metric.

PCA-loading and GO-term analyses of the merged dataset

The *biplot* function of the PCAtools R package[71] was used to generate PCA plots based on the most variable genes of the previously merged dataset. The *plotloadings* function from the same package was then used to obtain lists of the genes driving the clustering of the cell populations along the major principal components. Lists of genes positively associated with either principal component 1 or 2 were then subjected to a GO-term analysis (the *ont* argument was set to “BP”, p-value cut-off 0.01, q-value cut-off 0.05) to using the *enrichGO* function provided by the clusterprofiler R package[72] to predict biological processes associated with human CD161^{hi} CD4 and murine DN T cells.

Gene set enrichment analyses of the murine and human RNAseq-datasets

Raw gene count data from human and murine RNA-sequencing experiments were loaded into R and genes not or only lowly (less than 10 total counts or not expressed in all samples from at least one experimental group) expressed were removed using basic R commands and the *filterByExpr* function of the edgeR R package[66]. Filtered gene counts were then processed, transformed and normalized using the *DESeqDataSetFromMatrix*, *vst* and *DESeq* functions from the DESeq2 R package[73], respectively. The *results* function from the same package was used to generate lists of genes differentially expressed between human CD161hi and CD161int/- CD4 T cells or murine wt or Cbir DN and CD4 T cells respectively. To conduct gene set enrichment analyses (GSEA), a list of biological

processes was obtained using the msigdbr R package[74] and the *fgseaMultilevel* function of the fgsea R package[75] was used to perform GSEA and calculate p-values, BH-adjusted p-values and normalised enrichment scores (NES). The *plotEnrichment* function from the same package was used to visualize the enrichment curves of relevant processes.

Isolation of tissue leukocytes from mouse

Colon

Colon tissue was cut into ~1cm pieces and incubated (2x) in RPMI containing 1% BSA (Sigma Aldrich) and 5mM EDTA (Sigma Aldrich) in a 37C shaking incubator. Remaining tissue was incubated in RPMI containing 1% BSA, 15mM HEPES and 300U/ml of Collagenase VIII (Sigma-Aldrich, St Louis, MO) to digest the remaining tissue. Cell populations were purified by 37.5% Percoll (GE Healthcare, Little Chalfont, U.K.) gradient centrifugation (600 x *g*, 5 min). Lymphocytes were isolated from the pellet.

Lung

Lung tissue was minced to 1mm pieces using a scalpel and incubated in RPMI containing 1% BSA, 15mM HEPES and 300U/ml of Collagenase VIII (Sigma-Aldrich, St Louis, MO) in a 37C shaking incubator for approximately 30 min, pipetting to break up tissue halfway through. Lymphocytes were isolated from the single-cell suspension by gradient centrifugation with Ficoll-Hypaque (GE-Healthcare, 600 x g, 20 min).

Skin

Ears were harvested and stored on ice in PBS/BSA. Ears were mechanically split, finely minced, and digested in RPMI with BSA, collagenase D (Roche), and Liberase TM (Roche) at 37°C for 80 min. Leukocytes were separated by gradient centrifugation with Lymphoprep (StemCell Technologies, Inc.).

Lymphoid tissues

Lymphoid tissues (spleen and lymph nodes) were isolated from surrounding tissues with tweezers and maintained on ice in PBS/BSA. Tissues were processed to a single cell suspension by maceration through a 70µm mesh. Spleen samples were then incubated with 1ml ACK lysis buffer for 3 minutes to lyse red blood cells.

RNA sequencing

5 pools of 4 B6 mice were sorted for equal cell number (300) for CD127+ DN, MAIT, NKT, CD127+ CD4. 5 pools of 2 Cbir Rag-/- mice were sorted for Cbir DN population. Cells for RNA sequencing were isolated from tissue as described above and sorted with a FACSAria III directly into 300µl RLT (Qiagen). RNA was isolated using the RNAeasy micro kit (Qiagen). Quality control, library prep, and sequencing were performed at the Wellcome Trust Centre for Human Genetics, Oxford Genomics Centre using the SmartSeq2 protocol.

To obtain human data, CD4 T cells from three donors were sorted on a BD ARIA III sorter based on the expression of CD161 and CD56 as $CD161^{-}$, $CD161^{int}$, $CD161^{hi}CD56^{-}$ or $CD161^{hi}$ $CD56^{+}$ cells. For each donor and cell population a total of 800 cell was sorted in 4 batches (200 cells each) directly in PCR tubes containing 4µl SmartSeq2 lysis buffer. Reverse transcription was done directly from the sorted, lysed cells following the protocol published by Picelli *et al.*[76] in the MRC WIMM Sequencing Facility. cDNA libraries cDNA libraries were processed with a Nextera XT kit, using 8bp barcodes and were sequenced on a NextSeq500 sequencer.

Tissue qPCR

3mm pieces of colonic tissue were placed into RNA*later* (Qiagen) directly after sacrifice and stored at -20° Celsius until use. RNA was isolated using the RNAeasy mini kit (Qiagen) according to manufacturer's instrucitons. cDNA was synthesized using the Superscript III reverse transcription kit (Life Technologies). Quantitative real-time PCR for the candidate genes was performed using the Taqman system (Life Technologies) in duplicate and presented relative to *Hprt*.

DSS colitis

Mice were pre-treated with a depleting dose of anti-CD4 antibody (GK1.5, BioXCell, 0.5-1mg tested by batch) three days prior to DSS treatment and every 7 days for the duration of the experiment. 1.5% dextran sulfate sodium salt (36,000-50,000 M Wt, colitis grade, MP biomedical) was given in drinking water from day 0-5. Mice were weighed and monitored daily. Mice were culled and tissues harvested on day 19.

1. O'Hara, A.M. & Shanahan, F. The gut flora as a forgotten organ. *EMBO Rep* **7**, 688-693 (2006).

2. Ley, R.E., Peterson, D.A. & Gordon, J.I. Ecological and evolutionary forces shaping microbial diversity in the human intestine. *Cell* **124**, 837-848 (2006).

3. Hooper, L.V., Littman, D.R. & Macpherson, A.J. Interactions between the microbiota and the immune system. *Science* **336**, 1268-1273 (2012).

4. Sorini, C., Cardoso, R.F., Gagliani, N. & Villablanca, E.J. Commensal Bacteria-Specific CD4(+) T Cell Responses in Health and Disease. *Front Immunol* **9**, 2667 (2018).

5. Bauer, H., Horowitz, R.E., Levenson, S.M. & Popper, H. The response of the lymphatic tissue to the microbial flora. Studies on germfree mice. *Am J Pathol* **42**, 471-483 (1963).

6. Martin, R., Bermudez-Humaran, L.G. & Langella, P. Gnotobiotic Rodents: An In Vivo Model for the Study of Microbe-Microbe Interactions. *Front Microbiol* **7**, 409 (2016).

7. Chung, H. *et al.* Gut immune maturation depends on colonization with a host-specific microbiota. *Cell* **149**, 1578-1593 (2012).

8. Yang, Y. *et al.* Focused specificity of intestinal TH17 cells towards commensal bacterial antigens. *Nature* **510**, 152-156 (2014).

9. Ansaldo, E. *et al.* Akkermansia muciniphila induces intestinal adaptive immune responses during homeostasis. *Science* **364**, 1179-1184 (2019).

10. Hand, T.W. *et al.* Acute gastrointestinal infection induces long-lived microbiota-specific T cell responses. *Science* **337**, 1553-1556 (2012).

11. Xu, M. *et al.* c-MAF-dependent regulatory T cells mediate immunological tolerance to a gut pathobiont. *Nature* **554**, 373-377 (2018).

12. Cong, Y., Feng, T., Fujihashi, K., Schoeb, T.R. & Elson, C.O. A dominant, coordinated T regulatory cell-IgA response to the intestinal microbiota. *Proc Natl Acad Sci U S A* **106**, 19256-19261 (2009).

13. Lathrop, S.K. *et al.* Peripheral education of the immune system by colonic commensal microbiota. *Nature* **478**, 250-254 (2011).

14. Chai, J.N. *et al.* Helicobacter species are potent drivers of colonic T cell responses in homeostasis and inflammation. *Science immunology* **2** (2017).

15. Hsieh, S. *et al.* Polysaccharide Capsules Equip the Human Symbiont Bacteroides thetaiotaomicron to Modulate Immune Responses to a Dominant Antigen in the Intestine. *Journal of immunology* **204**, 1035-1046 (2020).

16. Nutsch, K. *et al.* Rapid and Efficient Generation of Regulatory T Cells to Commensal Antigens in the Periphery. *Cell Rep* **17**, 206-220 (2016).

17. Schenkel, J.M. & Masopust, D. Tissue-resident memory T cells. *Immunity* **41**, 886-897 (2014).

18. Iborra, S. *et al.* Optimal Generation of Tissue-Resident but Not Circulating Memory T Cells during Viral Infection Requires Crosspriming by DNGR-1(+) Dendritic Cells. *Immunity* **45**, 847-860 (2016).

19. Thome, J.J. *et al.* Spatial map of human T cell compartmentalization and maintenance over decades of life. *Cell* **159**, 814-828 (2014).

20. Mackay, L.K. *et al.* T-box Transcription Factors Combine with the Cytokines TGF-beta and IL-15 to Control Tissue-Resident Memory T Cell Fate. *Immunity* **43**, 1101-1111 (2015).

1110 21. Harrison, O.J. *et al.* Epithelial-derived IL-18 regulates Th17 cell differentiation and
1111 Foxp3(+) Treg cell function in the intestine. *Mucosal immunology* **8**, 1226-1236 (2015).
1112

1113 22. Bergsbaken, T., Bevan, M.J. & Fink, P.J. Local Inflammatory Cues Regulate
1114 Differentiation and Persistence of CD8(+) Tissue-Resident Memory T Cells. *Cell Rep*
1115 **19**, 114-124 (2017).
1116

1117 23. Kleinschek, M.A. *et al.* Circulating and gut-resident human Th17 cells express CD161
1118 and promote intestinal inflammation. *The Journal of experimental medicine* **206**, 525-
1119 534 (2009).
1120

1121 24. Provine, N.M. *et al.* Unique and Common Features of Innate-Like Human Vdelta2(+)
1122 gammadeltaT Cells and Mucosal-Associated Invariant T Cells. *Front Immunol* **9**, 756
1123 (2018).
1124

1125 25. Wragg, K.M. *et al.* High CD26 and Low CD94 Expression Identifies an IL-23
1126 Responsive Vdelta2(+) T Cell Subset with a MAIT Cell-like Transcriptional Profile.
1127 *Cell Rep* **31**, 107773 (2020).
1128

1129 26. Gao, Y. & Williams, A.P. Role of Innate T Cells in Anti-Bacterial Immunity. *Front*
1130 *Immunol* **6**, 302 (2015).
1131

1132 27. Fergusson, J.R. *et al.* CD161 defines a transcriptional and functional phenotype across
1133 distinct human T cell lineages. *Cell Rep* **9**, 1075-1088 (2014).
1134

1135 28. Hegazy, A.N. *et al.* Circulating and Tissue-Resident CD4(+) T Cells With Reactivity
1136 to Intestinal Microbiota Are Abundant in Healthy Individuals and Function Is Altered
1137 During Inflammation. *Gastroenterology* **153**, 1320-1337 e1316 (2017).
1138

1139 29. Dias, J., Leeansyah, E. & Sandberg, J.K. Multiple layers of heterogeneity and subset
1140 diversity in human MAIT cell responses to distinct microorganisms and to innate
1141 cytokines. *Proc Natl Acad Sci U S A* **114**, E5434-E5443 (2017).
1142

1143 30. Jameson, J. *et al.* A role for skin gammadelta T cells in wound repair. *Science* **296**, 747-
1144 749 (2002).
1145

1146 31. Boismenu, R. & Havran, W.L. Modulation of epithelial cell growth by intraepithelial
1147 gamma delta T cells. *Science* **266**, 1253-1255 (1994).
1148

32. Liuzzi, A.R. *et al.* Unconventional Human T Cells Accumulate at the Site of Infection in Response to Microbial Ligands and Induce Local Tissue Remodeling. *Journal of immunology* **197**, 2195-2207 (2016).

33. Chen, Y., Chou, K., Fuchs, E., Havran, W.L. & Boismenu, R. Protection of the intestinal mucosa by intraepithelial gamma delta T cells. *Proc Natl Acad Sci U S A* **99**, 14338-14343 (2002).

34. Liew, P.X., Lee, W.Y. & Kubes, P. iNKT Cells Orchestrate a Switch from Inflammation to Resolution of Sterile Liver Injury. *Immunity* **47**, 752-765 e755 (2017).

35. Tanno, H. *et al.* Invariant NKT cells promote skin wound healing by preventing a prolonged neutrophilic inflammatory response. *Wound Repair Regen* **25**, 805-815 (2017).

36. Hinks, T.S.C. *et al.* Activation and In Vivo Evolution of the MAIT Cell Transcriptome in Mice and Humans Reveals Tissue Repair Functionality. *Cell Rep* **28**, 3249-3262 e3245 (2019).

37. Leng, T. *et al.* TCR and Inflammatory Signals Tune Human MAIT Cells to Exert Specific Tissue Repair and Effector Functions. *Cell Rep* **28**, 3077-3091 e3075 (2019).

38. Constantinides, M.G. *et al.* MAIT cells are imprinted by the microbiota in early life and promote tissue repair. *Science* **366** (2019).

39. Linehan, J.L. *et al.* Non-classical Immunity Controls Microbiota Impact on Skin Immunity and Tissue Repair. *Cell* **172**, 784-796 e718 (2018).

40. Hue, S. *et al.* Interleukin-23 drives innate and T cell-mediated intestinal inflammation. *The Journal of experimental medicine* **203**, 2473-2483 (2006).

41. Uhlig, H.H. *et al.* Differential activity of IL-12 and IL-23 in mucosal and systemic innate immune pathology. *Immunity* **25**, 309-318 (2006).

42. Yen, D. *et al.* IL-23 is essential for T cell-mediated colitis and promotes inflammation via IL-17 and IL-6. *The Journal of clinical investigation* **116**, 1310-1316 (2006).

43. Kullberg, M.C. *et al.* IL-23 plays a key role in Helicobacter hepaticus-induced T cell-dependent colitis. *The Journal of experimental medicine* **203**, 2485-2494 (2006).

44. Milner, J.J. *et al.* Runx3 programs CD8(+) T cell residency in non-lymphoid tissues and tumours. *Nature* **552**, 253-257 (2017).

45. Beura, L.K. *et al.* CD4(+) resident memory T cells dominate immunosurveillance and orchestrate local recall responses. *The Journal of experimental medicine* **216**, 1214-1229 (2019).

46. Smillie, C.S. *et al.* Intra- and Inter-cellular Rewiring of the Human Colon during Ulcerative Colitis. *Cell* **178**, 714-730 e722 (2019).

47. FitzPatrick, M.E.B. *et al.* Human intestinal tissue-resident memory T cells comprise transcriptionally and functionally distinct subsets. *Cell Rep* **34**, 108661 (2021).

48. Riol-Blanco, L. *et al.* IL-23 receptor regulates unconventional IL-17-producing T cells that control bacterial infections. *Journal of immunology* **184**, 1710-1720 (2010).

49. Lacorazza, H.D., Porritt, H.E. & Nikolich-Zugich, J. Dysregulated expression of pre-Talpha reveals the opposite effects of pre-TCR at successive stages of T cell development. *Journal of immunology* **167**, 5689-5696 (2001).

50. Crispin, J.C. *et al.* Expanded double negative T cells in patients with systemic lupus erythematosus produce IL-17 and infiltrate the kidneys. *Journal of immunology* **181**, 8761-8766 (2008).

51. Sherlock, J.P. *et al.* IL-23 induces spondyloarthropathy by acting on ROR-gammat+ CD3+CD4-CD8- entheseal resident T cells. *Nat Med* **18**, 1069-1076 (2012).

52. Marchi, E., Lee, L.N. & Klenerman, P. Inflation vs. Exhaustion of Antiviral CD8+ T-Cell Populations in Persistent Infections: Two Sides of the Same Coin? *Front Immunol* **10**, 197 (2019).

53. Heng, T.S., Painter, M.W. & Immunological Genome Project, C. The Immunological Genome Project: networks of gene expression in immune cells. *Nat Immunol* **9**, 1091-1094 (2008).

54. Provine, N.M. & Klenerman, P. MAIT Cells in Health and Disease. *Annu Rev Immunol* **38**, 203-228 (2020).

55. Kiner, E. *et al.* Gut CD4(+) T cell phenotypes are a continuum molded by microbes, not by TH archetypes. *Nat Immunol* **22**, 216-228 (2021).

56. Cano-Gamez, E. *et al.* Single-cell transcriptomics identifies an effectorness gradient shaping the response of CD4(+) T cells to cytokines. *Nature communications* **11**, 1801 (2020).

57. Miragaia, R.J. *et al.* Single-Cell Transcriptomics of Regulatory T Cells Reveals Trajectories of Tissue Adaptation. *Immunity* **50**, 493-504 e497 (2019).

58. Mueller, S.N. & Mackay, L.K. Tissue-resident memory T cells: local specialists in immune defence. *Nature reviews. Immunology* **16**, 79-89 (2016).

59. Gebhardt, T. *et al.* Memory T cells in nonlymphoid tissue that provide enhanced local immunity during infection with herpes simplex virus. *Nat Immunol* **10**, 524-530 (2009).

60. Zegarra-Ruiz, D.F. *et al.* Thymic development of gut-microbiota-specific T cells. *Nature* **594**, 413-417 (2021).

61. Feng, T., Wang, L., Schoeb, T.R., Elson, C.O. & Cong, Y. Microbiota innate stimulation is a prerequisite for T cell spontaneous proliferation and induction of experimental colitis. *The Journal of experimental medicine* **207**, 1321-1332 (2010).

62. Cong, Y. *et al.* CD4+ T cells reactive to enteric bacterial antigens in spontaneously colitic C3H/HeJBir mice: increased T helper cell type 1 response and ability to transfer disease. *The Journal of experimental medicine* **187**, 855-864 (1998).

63. Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587 e3529 (2021).

64. Aibar, S. *et al.* SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* **14**, 1083-1086 (2017).

65. Morgan, M., Falcon, S. & Gentleman, R. GSEABase: Gene set enrichment data structures and methods. 2021. p. R package version 1.54.50.

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

67. Ritchie, M.E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).

68. Durinck, S., Spellman, P.T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc* **4**, 1184-1191 (2009).

69. Leek, J.T. *et al.* sva: Surrogate Variable Analysis. 3.40.0. ed; 2021. p. R package version

70. Johnson, W.E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118-127 (2007).

71. Blighe, K. & Lun, A. PCAtools: PCAtools: Everything Principal Components Analysis. 2021. p. R package version 2.4.0.

72. Yu, G., Wang, L.G., Han, Y. & He, Q.Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284-287 (2012).

73. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).

74. Dolgalev, I. msigdbr: MSigDB Gene Sets for Multiple Organisms in a Tidy Data Format. 2021. p. R package version 7.4.1.

75. Korotkevich, G., Sukhov, V. & Sergushichev, A. Fast gene set enrichment analysis. *bioRxiv* (2019).

76. Picelli, S. *et al.* Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc* **9**, 171-181 (2014).