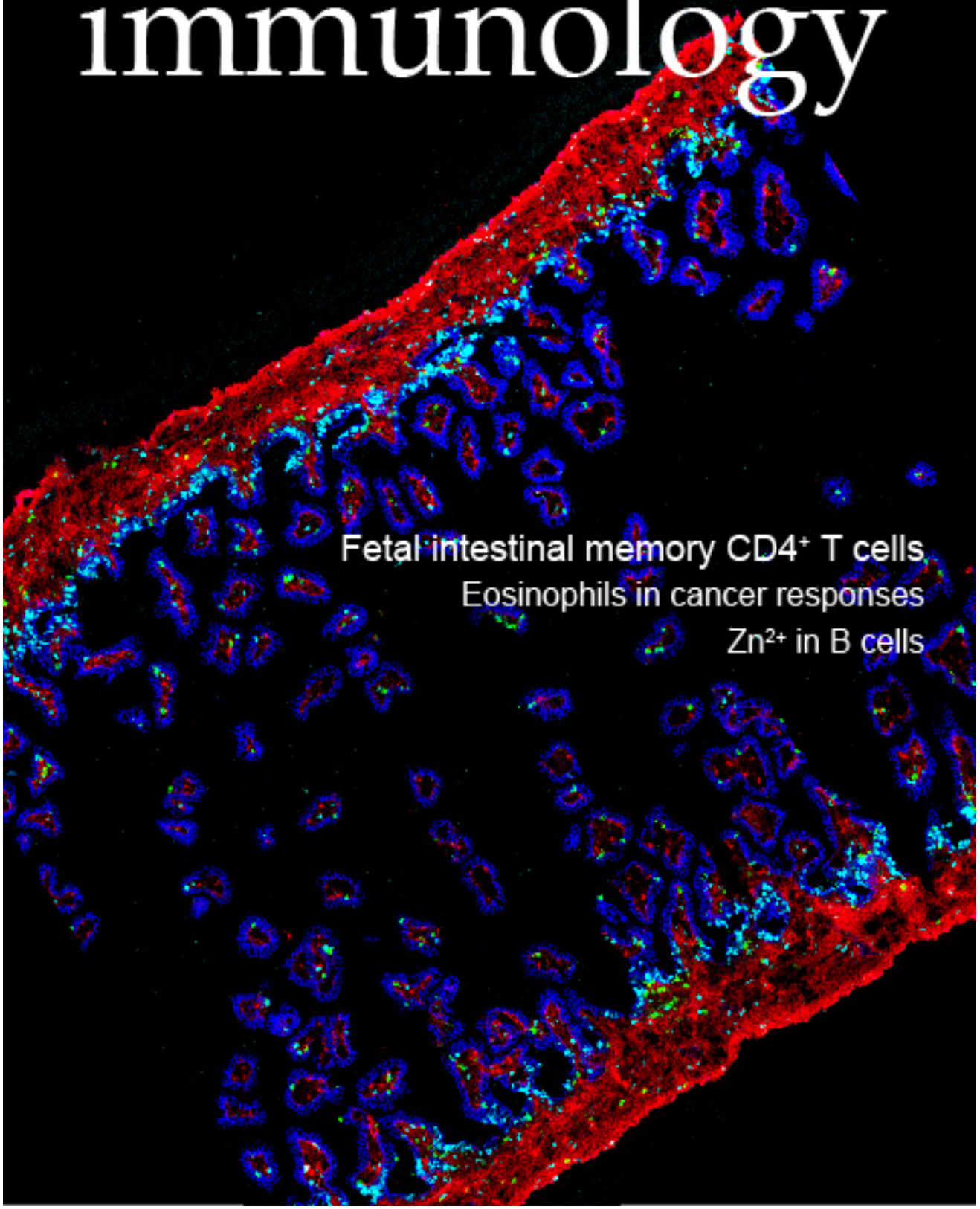
Development of the human fetal immune system: novel insights from high-dimensional single-cell technologies

Li, N.

MARCH 2019 VOL 20 NO 3
www.nature.com/natureimmunology

nature immunology



Fetal intestinal memory $CD4^+$ T cells

Eosinophils in cancer responses

Zn^{2+} in B cells

Chapter 3

Memory CD4$^+$ T cells are generated in the human fetal intestine

Na Li[1], Vincent van Unen[1], Tamim Abdelaal[2,3], Nannan Guo[1], Sofya A. Kasatskaya[4,5], Kristin Ladell[6], James E. McLaren[6], Evgeny S. Egorov[4], Mark Izraelson[4], Susana M. Chuva de Sousa Lopes[7], Thomas Höllt[2,8], Olga V Britanova[4], Jeroen Eggermont[9], Noel F.C.C. de Miranda[10], Dmitriy M. Chudakov[4,5,11,12,13,14], David A. Price[6,15], Boudewijn P.F. Lelieveldt[3,9], Frits Koning[1]

[1]Department of Immunohematology and Blood Transfusion; [2]Leiden Computational Biology Center, Leiden University Medical Center, Leiden, Netherlands; [3]Department of Pattern Recognition and Bioinformatics Group, Delft University of Technology, Delft, Netherlands; [4]Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Moscow, Russia; [5]Centre for Data-Intensive Biomedicine and Biotechnology, Skolkovo Institute of Science and Technology, Moscow, Russia; [6]Division of Infection and Immunity, Cardiff University School of Medicine, Cardiff, UK; [7]Department of Anatomy and Embryology, Leiden University Medical Center, Leiden, Netherlands; [8]Computer Graphics and Visualization Group, Delft University of Technology, Delft, Netherlands; [9]Department of Radiology, [10]Department of Pathology, Leiden University Medical Center, Leiden, Netherlands; [11]Central European Institute of Technology, Masaryk University, Brno, Czech Republic; [12]Department of Molecular Technologies, Pirogov Russian National Research Medical University, [13]MiLaboratory LLC, Skolkovo Innovation Centre, Moscow, Russia; [14]Privolzhsky Research Medical University, Nizhny Novgorod, Russia; [15]Systems Immunity Research Institute, Cardiff University School of Medicine, Cardiff, UK

Nature Immunology 20:3, 301-312 (2019)

ON THE COVER

Koning and colleagues used mass cytometry, single-cell RNA-seq and high-throughput TCR sequencing to characterize the CD4+ T cell compartment in the human fetal intestine

Abstract

The fetus is thought to be protected from exposure to foreign antigens, yet CD45RO$^+$ T cells reside in the fetal intestine. Here we combined functional assays with mass cytometry, single-cell RNA-sequencing and high-throughput T cell antigen receptor (TCR) sequencing to characterize the CD4$^+$ T cell compartment in the human fetal intestine. We identified 22 CD4$^+$ T cell clusters, including naive-like, regulatory-like and memory-like subpopulations, which were confirmed and further characterized at the transcriptional level. Memory-like CD4$^+$ T cells had high expression of Ki-67, indicative of cell division, and CD5, a surrogate marker of TCR avidity, and produced the cytokines IFN-γ and IL-2. Pathway analysis revealed a differentiation trajectory associated with cellular activation and proinflammatory effector functions, and TCR repertoire analysis indicated clonal expansions, distinct repertoire characteristics and interconnections between subpopulations of memory-like CD4$^+$ T cells. Imaging-mass cytometry indicated that memory-like CD4$^+$ T cells colocalized with antigen-presenting cells. Collectively, these results provide evidence for the generation of memory-like CD4$^+$ T cells in the human fetal intestine that is consistent with exposure to foreign antigens.

Introduction

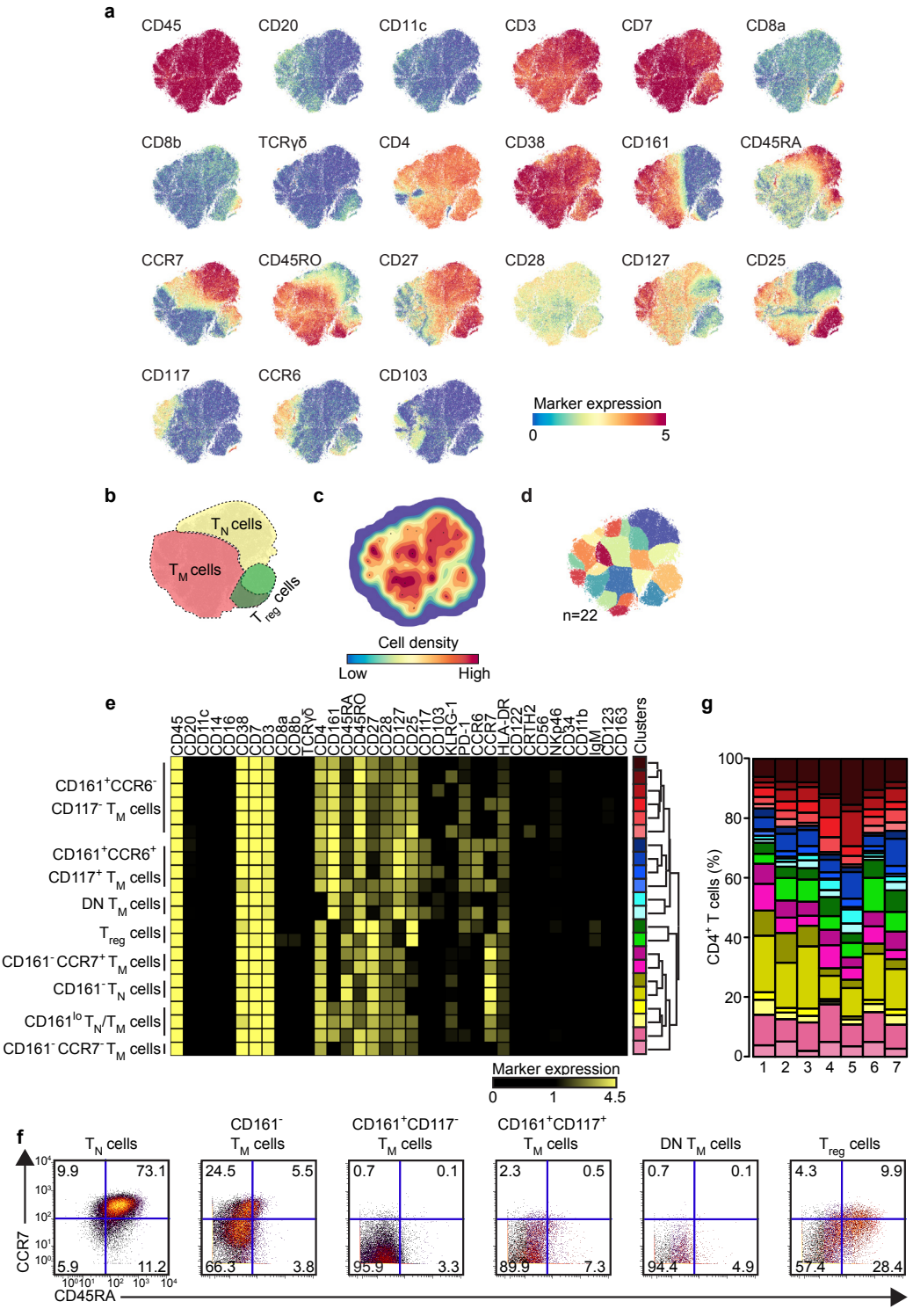
Adaptive immunity is founded on the selection and expansion of antigen-specific T cells from a clonally diverse pool of naive precursors[1]. Naive T cells recirculate among lymph nodes to survey the array of peptide epitopes bound to major histocompatibility complex (MHC) proteins on the surface of antigen-presenting cells (APCs), and functional recognition of a given peptide-MHC molecule is governed by various danger signals and specific engagement via the clonotypically expressed T cell antigen receptor (TCR). This triggers a program of differentiation and proliferation that results in the generation of effector T cells, which home to the site of the primary infection and contribute to pathogen clearance, and memory T cells, which remain in the circulation and mediate anamnestic responses to secondary infection. In the last decade, it has also become clear that tissue-resident T cells are commonly present at barrier sites, including the intestine[2].

Fundamental knowledge of adaptive immunity during early life remains sparse. The infantile intestine is known to harbor clonally expanded T cells[3], which were also identified in the human fetal intestine, but rarely in fetal mesenteric lymph nodes, fetal thymus or fetal spleen, suggesting compartmentalization[4]. In addition, a rare population of CD4+ T cells displaying a memory and proinflammatory phenotype has been identified in umbilical cord blood[5]. Although the dogma of a sterile womb has been challenged by reports of bacteria colonization in the placenta[6,7], amniotic fluid[8,9] and meconium[10], others have questioned these results[11]. Here we have combined functional studies with mass cytometry, RNA-sequencing (RNA-seq) and high-throughput TCR-sequencing to perform an in-depth analysis of the fetal intestinal CD4+ T cell compartment. Our results provide evidence for memory formation in the human fetal intestine, consistent with *in utero* exposure to foreign antigens.

Results

Human fetal intestinal CD4+ T cells are phenotypically diverse

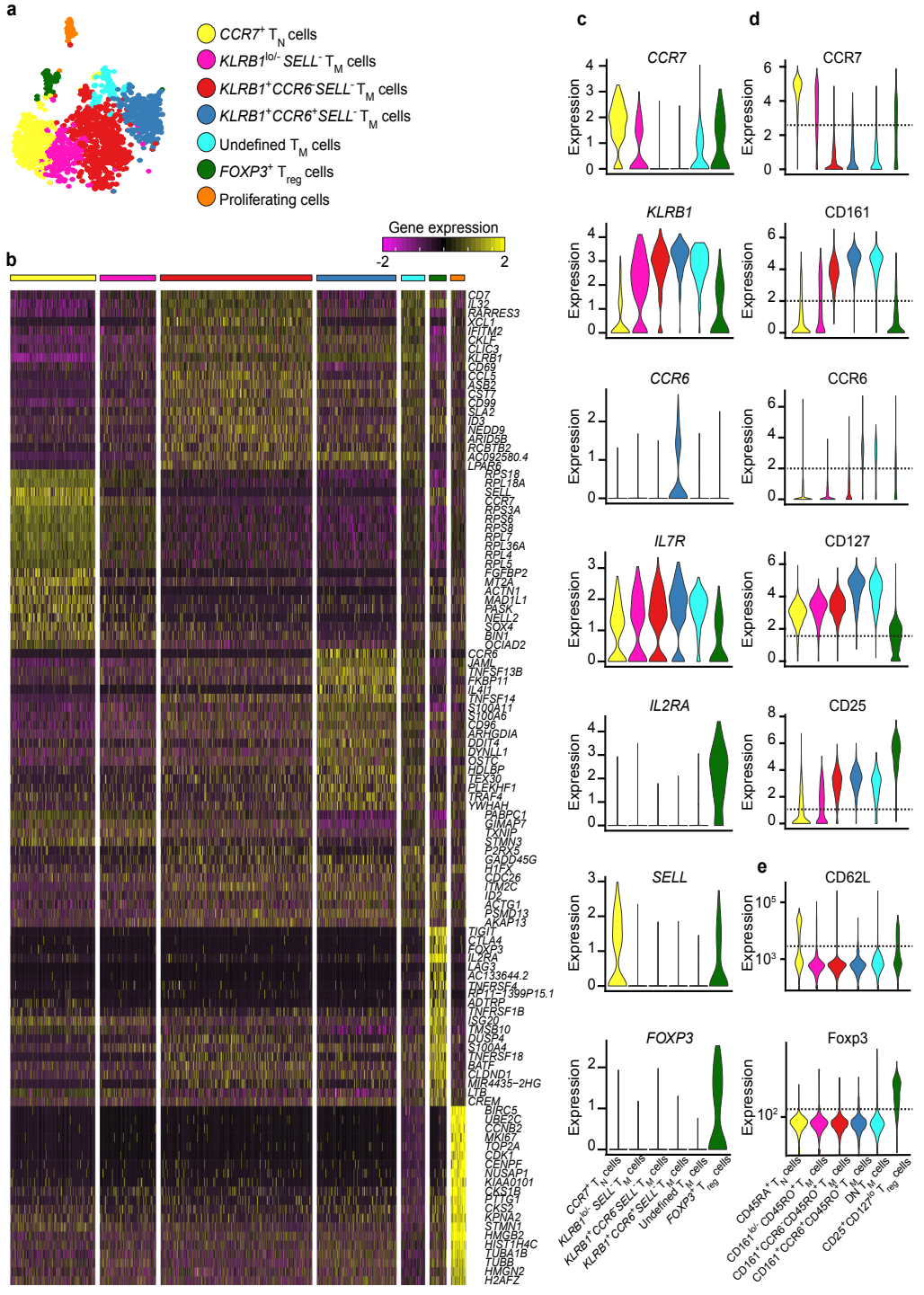
To explore the CD4+ T cell compartment in the human fetal intestine, we applied a mass cytometry panel comprising 35 antibodies (**Supplementary Table 1**) that was designed to capture the heterogeneity of the immune system to seven lamina propria samples aged 14-21 gestational weeks[12]. After data acquisition, we selected CD45+ immune cells (**Supplementary Fig. 1a**) and mined the dataset via hierarchical stochastic neighbor embedding (HSNE)[13]. At the overview level, HSNE


a
CD45
CD20
CD11c
CD3
CD7
CD8a
CD8b
TCRγδ
CD4
CD38
CD161
CD45RA
CCR7
CD45RO
CD27
CD28
CD127
CD25
CD117
CCR6
CD103
Marker expression
0
5
b
T N cells
T M cells
T reg cells
c
Cell density
Low
High
d
n=22
e
CD45
CD20
CD11c
CD14
CD16
CD38
CD7
CD3
CD8a
CD8b
TCRγδ
CD4
CD161
CD45RA
CD45RO
CD28
CD27
CD127
CD25
CD117
CD103
KLRG-1
PD-1
CCR6
CCR7
HLA-DR
CD122
CRTH2
CD56
NKp46
CD34
CD11b
IgM
CD123
CD163
Clusters
CD161+CCR6-
CD117- TM cells
CD161+CCR6+
CD117+ TM cells
DN TM cells
Treg cells
CD161-CCR7+ TM cells
CD161- TN cells
CD161lo TN/TM cells
CD161-CCR7- TM cells
Marker expression
0
1
4.5
g
CD4+ T cells (%)
100
80
60
40
20
0
1
2
3
4
5
6
7
f
TN cells
CD161- TM cells
CD161+CD117- TM cells
CD161+CD117+ TM cells
DN TM cells
Treg cells
CCR7
CD45RA
9.9
73.1
5.9
11.2
24.5
5.5
66.3
3.8
0.7
0.1
95.9
3.3
2.3
0.5
89.9
7.3
0.7
0.1
94.4
4.9
4.3
9.9
57.4
28.4

landmarks depicted the general composition of the immune system, with clear separation of the CD4+ T cell lineage (**Supplementary Fig. 1b**). We identified 110,332 CD4+ T cells, with an average of 15,761 events per fetal intestine, comprising 47.9% ± 9.6% of all immune cells. We then subjected HSNE-defined CD4+ T cells (**Supplementary Fig. 1b**) to t-distributed stochastic neighbor embedding (t-SNE)[14] in Cytosplore[15] to project their marker expression profiles onto a two-dimensional graph (**Fig. 1a and Supplementary Fig. 1c**). CD4+ T cells were characterized as CD45+CD3+CD4+CD7+ (**Fig. 1a**). Moreover, all CD4+ T cells were positive for the tissue-resident marker CD38 and approximately 50% of cells expressed CD161. 24.1% of the CD4+ T cell population co-expressed CD27, CD28, CD45RA and CCR7, indicative of a naive T cell (T_N) phenotype, whereas 64.5% expressed CD45RO, indicative of a memory T cell (T_M) phenotype (Fig. 1a,b). While all CD45RO+ T_M cells were CD28+, differential expression of CD25, CD27, CD103, CD117, CD127, CCR6 and CCR7 was observed on these cells (**Fig. 1a,b**), reflecting substantial phenotypic diversity.

We next applied Gaussian mean-shift clustering to the mass cytometry data using the t-SNE coordinates of the embedded CD4+ T cells (**Fig. 1a**). Based on cell density features (**Fig. 1c**), this identified 22 distinct CD4+ T cell clusters (**Fig. 1d**), each defined by a unique marker expression profile. Hierarchical clustering of the heatmap revealed eight major groups (CD161+CCR6-CD117- T_M cells, CD161+CCR6+CD117+ T_M cells, DN T_M cells, T_{reg} cells, CD161-CCR7+ T_M cells, CD161- T_N cells, CD161lo T_N/T_M cells and CD161-CCR7- T_M cells) (**Fig. 1e**). High expression of CD25 and a lack of CD127 distinguished two regulatory T (T_{reg}) cell clusters, with either a CD45RA+ T_N or a CD45RO+ T_M phenotype (**Fig. 1a,b,e**). CD161+CD4+ T cells branched into a CCR6-CD117-CD45RO+ T_M and a CCR6+CD117+CD45RO+ T_M cluster (**Fig. 1e**). Moreover, CD45RA+ T_N and CD45RO+ T_M cells were detected in both the CD161- and the CD161lo subpopulations. Additional diversity was observed for the expression of several activation markers, including CRTH2, HLA-DR,

Fig. 1. Mass cytometric analysis of fetal intestinal CD4+ T cells. **a**, t-SNE embedding of all CD4+ T cells (n = 110,332) derived from human fetal intestines (n = 7). Colors represent the ArcSinh5-transformed expression values of the indicated markers. **b**, t-SNE plot depicting the population cell border for T_N cells (dashed yellow line), T_M cells (dashed red line), and Treg cells (dashed green line). **c**, Density map describing the local probability density of cells, where black dots indicate the centroids of identified clusters using Gaussian mean-shift clustering. **d**, t-SNE plot showing cluster partitions in different colors. **e**, Heatmap showing median expression values and hierarchical clustering of markers for the identified subpopulations. **f**, Biaxial plots showing CD45RA and CCR7 expression on the indicated clusters analyzed by mass cytometry. The 22 clusters were merged into 6 phenotypic groups according to the heatmap shown in (**e**). **g**, Composition of the CD4+ T cell compartment in each fetal intestine represented by vertical bars, where the colored segment lengths represent the proportion of cells as a percentage of all CD4+ T cells in the sample. Colors as shown in (**e**)


a
CCR7+ TN cells
KLRB1lo/- SELL- TM cells
KLRB1+CCR6-SELL- TM cells
KLRB1+CCR6+SELL- TM cells
Undefined TM cells
FOXP3+ Treg cells
Proliferating cells
b
Gene expression
-2
2
c
d
e
CCR7
KLRB1
CCR6
IL7R
IL2RA
SELL
FOXP3
CD161
CD127
CD25
CD62L
Foxp3
Expression
CD45RA+ TN cells
CD161lo- CD45RO+ TM cells
CD161+CCR6-CD45RO+ TM cells
CD161+CCR6+CD45RO+ TM cells
DN TM cells
CD25+CD127lo Treg cells

KLRG-1 and PD-1, the latter especially within the CD45RO+ T_M cell clusters (**Supplementary Fig. 1c**). Of note, a small population of CD4-CD8a-TCRγδ- (DN) T_M cells clustered among CD4+ T cells in both the HSNE and t-SNE plots. Biaxial plots confirmed coexpression of CD45RA and CCR7 on T_N cells (**Fig. 1f**), whereas the CD161lo/-CD45RO+ T_M subpopulation contained both CCR7+ central memory T (T_{CM}) cells and CCR7- effector memory T (T_{EM}) cells (**Fig. 1f**). All other CD45RO+ T_M subpopulations harboured primarily T_{EM} cells. Quantification of cellular frequencies for the CD4+ T cell clusters per fetal intestine revealed highly similar compositions with all CD45RO+ T_M clusters detectable in all samples (**Fig. 1g**). In contrast, parallel analyses of CD4+ T cells isolated from three fetal livers and three fetal spleens from one shared and two additional fetuses aged 16-21 gestational weeks revealed a predominance of CD45RA+ T_N cells (**Supplementary Fig. 2a,b**). These results delineated a phenotypically diverse array of human fetal intestinal CD4+ T cells, most of which displayed features associated with antigen exposure.

Fetal CD4+ T cells display a memory gene expression profile

We next performed single-cell RNA-seq on flow-sorted fetal intestinal CD4+ cells from a lamina propria sample that was also included in the mass cytometry analysis. This yielded data for 1,804 CD4+ T cells, identifying cell-specific variable expression of 2,174 genes (Methods), which were further analyzed using the Seurat computational pipeline[16]. Unsupervised clustering revealed nine transcriptionally distinct subpopulations, seven of which corresponded to CD3+ T cell subsets while two displayed a gene expression profile matching *CD86*+*HLA-DR*+ APCs. The corresponding gene expression profiles of the seven T cell subsets were projected onto a single graph using t-SNE (**Fig. 2a**), and the top 20 upregulated genes were displayed in a heatmap (**Fig. 2b**). Five of the seven RNA-seq-identified CD4+ T cell subpopulations corresponded to the mass cytometry-defined CD4+ T cell major groups: *CCR7*+ T_N with CD45RA+ T_N, *KLRB1*lo/-*SELL*- T_M with CD161lo/-CD45RO+ T_M, *KLRB1*+*CCR6*-*SELL*- T_M with CD161+CCR6-CD45RO+ T_M, *KLRB1*+*CCR6*+*SELL*- T_M with CD161+CCR6+CD45RO+ T_M and *FOXP3*+ T_{reg} cells with CD25+CD127lo T_{reg} cells. The mass cytometry defined CD161- and CD161lo subpopulations (**Fig. 1e**) could

Fig. 2. Single-cell RNA-sequencing of fetal intestinal CD4+ T cells. a, t-SNE embedding of fetal intestinal CD4+ T cells (n = 1,804) showing seven transcriptionally distinct clusters, including *CCR7*+ T_N (n = 358), *KLRB1*lo/-*SELL*-T_M (n = 237), *KLRB1*+*CCR6*-*SELL*-T_M (n = 640), *KLRB1*+*CCR6*+*SELL*-T_M (n = 336), undefined T_M (n = 101), *FOXP3*+ T_{reg} cells (n = 71), and proliferating cells (n =61). Colors indicate different cell clusters. **b**, Heatmap showing the normalized single-cell gene expression value (Z-Score, purple-to-yellow scale) for the top 20 differentially upregulated genes in each identified cluster. Colors as shown in (**a**). **c–e**, Expression of the indicated genes in each identified cluster at (**c**) the RNA level (log-normalized) and (**d**,**e**) the protein level analyzed by (**d**) mass cytometry (CyTOF, ArcSinh5-transformed) or (**e**) flow cytometry, presented as violin plots. Dashed lines indicate background levels. Colors as shown in (**a**).

not be discriminated in the RNA-seq dataset. One additional RNA-seq-identified subpopulation corresponded to proliferating cells, based on the expression of genes associated with cell division (*CCNB2*, *CDK1* and *MKI67*) (**Fig. 2a,b**).

As CD45RA and CD45RO were not detectable, we used other markers to distinguish T_M from T_N clusters. To compare gene or marker expression among cell clusters, we used violin plots, displaying the mode average as the thickest section (**Fig. 2c-e**). Consistent with the mass cytometry data, RNA-seq-defined T_N cells were *KLRB1*$^-$*CCR7*$^+$*SELL*$^{+/-}$ (**Fig. 2c,d**), the latter confirmed by flow cytometry (**Fig. 2e and Supplementary Fig. 3a**). In the absence of T_M-associated markers, *SELL*$^-$ T_M cell populations were identified on the basis of differential expression of *KLRB1* and *CCR6* (**Fig. 2c,d**). Consistent with the mass cytometry data, expression of *KIT* (CD117) was restricted primarily to *KLRB1*$^+$*CCR6*$^+$*SELL*$^-$ T_M cells (**Supplementary Fig. 3b**). Moreover, the gene expression profile of the *IL2RA*$^+$*IL7R*^{lo}*FOXP3*$^+$ T_{reg} cell population (**Fig. 2c**) corresponded to the mass cytometry-defined CD25$^+$CD127^{lo} T_{reg} cells (**Fig. 2d**). In addition, several RNA-transcripts including *LAG3*, *TIGIT, CTLA4* and *TNFRSF18* (GITR) ascertained the identity of *FOXP3*$^+$ T_{reg} cells (**Fig. 2b**). Finally, the RNA-seq data revealed an undefined T_M cluster that was not identified by mass cytometry, but expressed genes similar to those detected in the *KLRB1*$^+$*CCR6*$^+$*SELL*$^-$ T_M subpopulation, such as *CD69*, *CCL5* and *JAML*. Cell population frequencies identified by mass cytometry and RNA-seq were comparable with the exception of mass cytometry-defined CD25$^+$CD127^{lo} T_{reg} cells and RNA-seq-defined *FOXP3*$^+$ T_{reg} cells (**Supplementary Fig. 3c**).

Compared with the *CCR7*$^+$ T_N population, *KLRB1*$^{lo/-}$*SELL*$^-$ T_M, *KLRB1*$^+$*CCR6*$^-$*SELL*$^-$ T_M, *KLRB1*$^+$*CCR6*$^+$*SELL*$^-$ T_M and undefined T_M subpopulations had high expression of the tissue-resident and activation-associated gene *CD69*, the differentiation-promoting gene *ANXA1* (Annexin A1), the chemokine-like factor *CKLF*, the cytokine *IL32*, the proliferation-associated gene *JUN* (C-Jun) and the adhesion molecule *JAML* (**Fig. 3a**). *CD40LG* (CD154), *TNFSF14* and *TGFB1* were specifically upregulated by *KLRB1*$^+$*CCR6*$^-$*SELL*$^-$ T_M, *KLRB1*$^+$*CCR6*$^+$*SELL*$^-$ T_M and undefined T_M clusters, while *CCL5* and *MAP3K8* kinase upregulated by *KLRB1*$^+$*CCR6*$^-$*SELL*$^-$ T_M and undefined T_M subpopulations. Moreover, *IL4I1* was specifically expressed by *KLRB1*$^+$*CCR6*$^+$*SELL*$^-$ T_M and undefined T_M cells. In addition, all fetal *SELL*$^-$ T_M subpopulations had high expression of the tissue-resident genes *ITGAE* (CD103) and/or *CD38* (**Fig. 3a**).

In agreement with the RNA-seq data, flow cytometry indicated that the activation markers CXCR3, CCR4, CD69 and CD226 were highly expressed on CCR7$^-$ T_{EM} cells (**Fig. 3b**). All CD4$^+$ T cells expressed CD95, with the highest expression on CD161$^+$ T_{EM} cells (**Fig. 3b**). Expression of CD31, a marker associated with recent

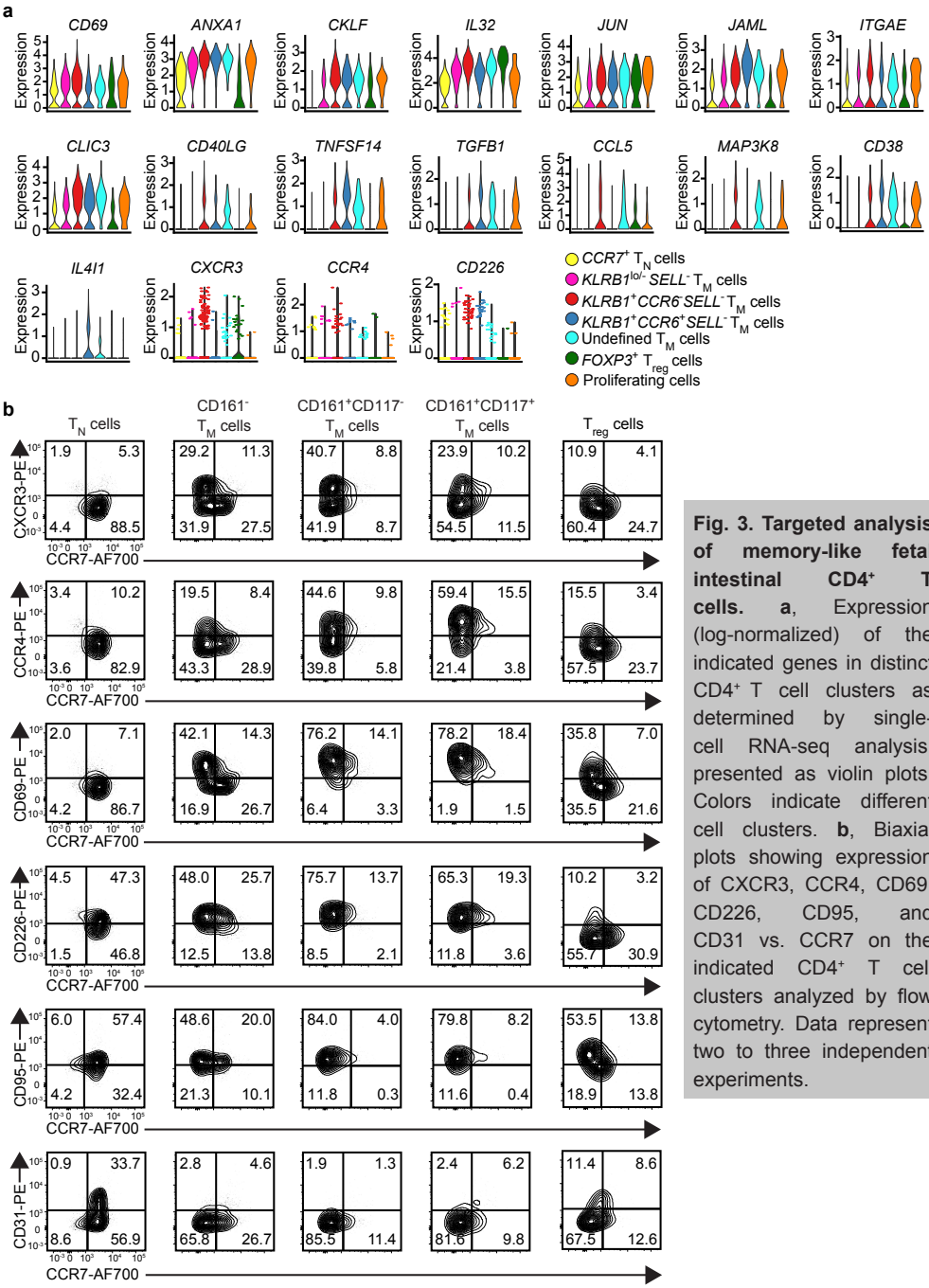


Fig. 3. Targeted analysis of memory-like fetal intestinal CD4+ T cells. **a**, Expression (log-normalized) of the indicated genes in distinct CD4+ T cell clusters as determined by single-cell RNA-seq analysis, presented as violin plots. Colors indicate different cell clusters. **b**, Biaxial plots showing expression of CXCR3, CCR4, CD69, CD226, CD95, and CD31 vs. CCR7 on the indicated CD4+ T cell clusters analyzed by flow cytometry. Data represent two to three independent experiments.

thymic emigrants[17], was highest on CD45RA$^+$ T_N cells (**Fig. 3b**).

Thus, RNA-seq confirmed the existence of distinct subpopulations of CD4$^+$ T cells and indicate that many genes associated with inflammation and tissue residency were upregulated by fetal CD4$^+$ T_M cells, consistent with antigen-driven functionality and maturation.

Computational analysis reveals a differentiation pathway of CD4$^+$ T cells

We next visualized the evolution of the t-SNE computation of the mass cytometry and RNA-seq data to reveal the ordering of single cells along putative differentiation trajectories[12,15]. At the onset of the mass cytometry data computation, where cells are grouped based on major shared features, CD25$^+$CD127^{lo} T_{reg} cells clustered separate from the other cells, whereas the other cell clusters were ordered in a linear fashion with the CD45RA$^+$ T_N cells next to the CD161$^{lo/-}$CD45RO$^+$ T_M cells, followed by the CD161$^+$CCR6$^-$CD45RO$^+$ T_M cells and the CD161$^+$CCR6$^+$CD45RO$^+$ T_M cells, consecutively (**Fig. 4a**). A similar phenotypic ordering was observed in parallel analyses of the RNA-seq data, although the *KLRB1*$^+$*CCR6*$^+$*SELL*$^-$ T_M subpopulation aligned differently, but remained connected with the *KLRB1*$^+$*CCR6*$^-$*SELL*$^-$ T_M cluster (**Fig. 4b**). Individual marker expression patterns at the middle of the t-SNE computation validated the ordering of the clusters and the comparability of the mass cytometry and RNA-seq data (**Supplementary Fig. 4a,b**). Similar patterns were identified using Diffusion map[18], VorteX[19] and principal component analysis (PCA)[20] (**Supplementary Fig. 4c–f**). Thus, this analysis reveals a putative differentiation pathway leading to T_M formation.

To extend our analysis of the gene expression profiles underlying this putative differentiation trajectory, we used the pseudotime algorithm in the Monocle toolkit[21,22], which calculates the ordering of individual cells based on single-cell expression profiles. Based on this analysis, *CCR7*$^+$ T_N cells were separated from *SELL*$^-$ T_M cells (**Fig. 4c**). When we clustered genes according to expression patterns along the pseudotime trajectory, cell-to-cell transitioning could be explained by the kinetics of 1,376 variable genes, which formed three large modules (**Fig. 4d**). The first module contained 540 genes associated with *CCR7*$^+$ T_N cells, including *SELL*, *CCR7*, *CD27* and *CD28* (**Fig. 4d**). The second module contained 453 genes, many of which were associated with an ongoing transcriptional program, such as *RPL21*, *RPS2* and *RPLP1*. The highest activity of this transcriptional gene expression profile coincided with the transition of cells with a *CCR7*$^+$ T_N phenotype into cells with a *SELL*$^-$ T_M phenotype (**Fig. 4c,d**). The third module contained 383 genes (**Fig. 4d**), 106 of which were associated with cellular activation and regulation of the immune system (**Supplementary Fig. 5a**), while 133 encoded proteins known

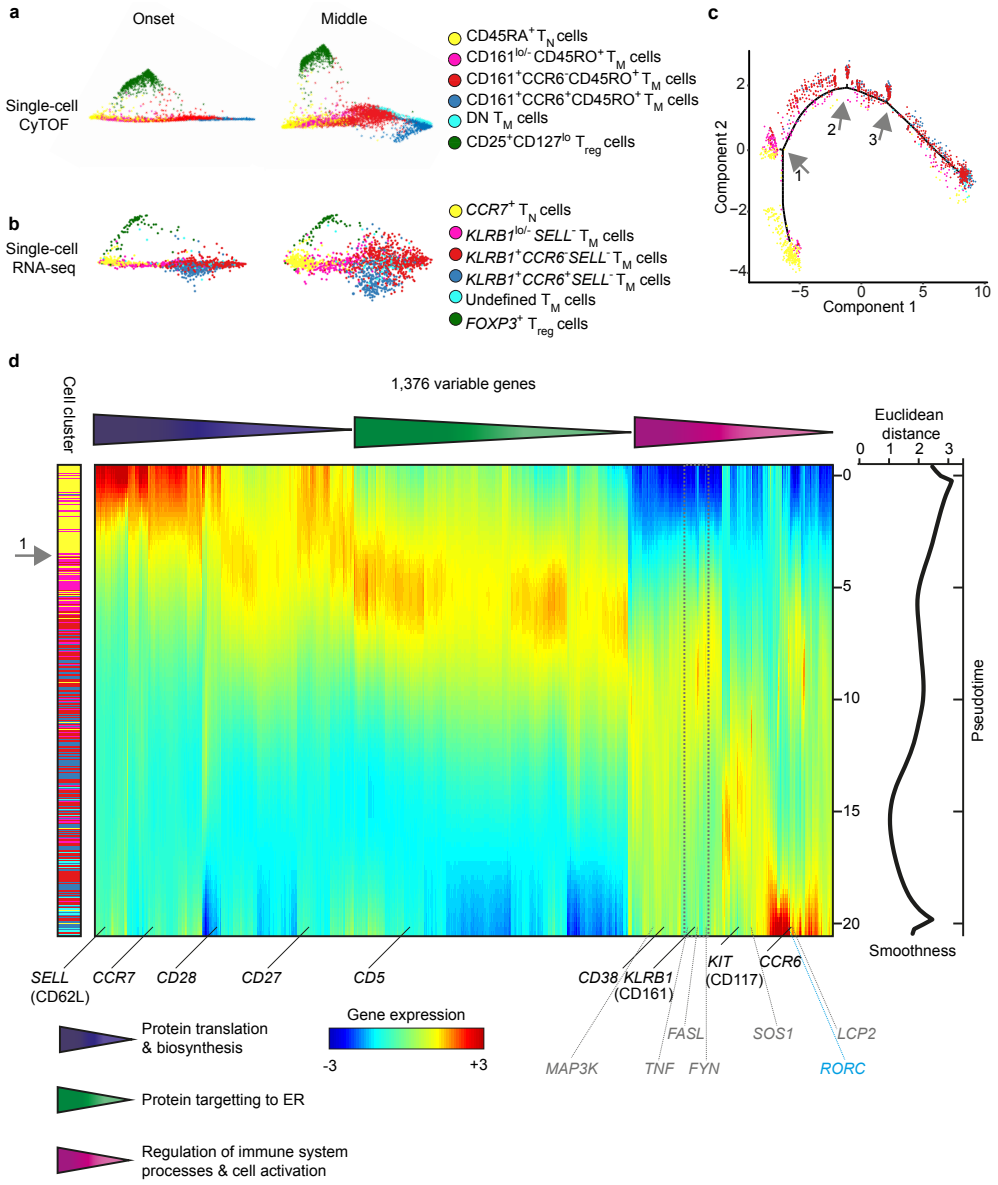


Fig. 4. Single-cell trajectory analysis of fetal intestinal CD4⁺ T cells. a–b, t-SNE embeddings of all fetal intestinal CD4⁺ T cells analyzed by (**a**) mass cytometry (n =10,436) and (**b**) single-cell RNA-seq (n = 1,743) at the onset and at the middle of the t-SNE computation. Colors indicate different cell clusters. **c**, A single-cell trajectory from the RNA-seq data (excluding T_{reg} cells and proliferating cells) recovered by pseudotime analysis. Colors indicate different cell clusters as shown in (**b**). Grey arrows indicate three small branches. **d**, Three kinetic modules of pseudotime-dependent genes (n = 1,376) depicted in a log-variance-stabilized expression heatmap, indicating gene-enriched biological processes. Genes confirmed by mass cytometry and flow cytometry are denoted by black labels, and genes involved in TCR signalling are denoted by gray labels. The dashed grey box indicates the coordinated expression profile of TNF, FASL, and FYN. Euclidean distance values comparing gene expression profiles for each ordered pair of neighboring cells along the pseudotime trajectory are shown in the graph (right).

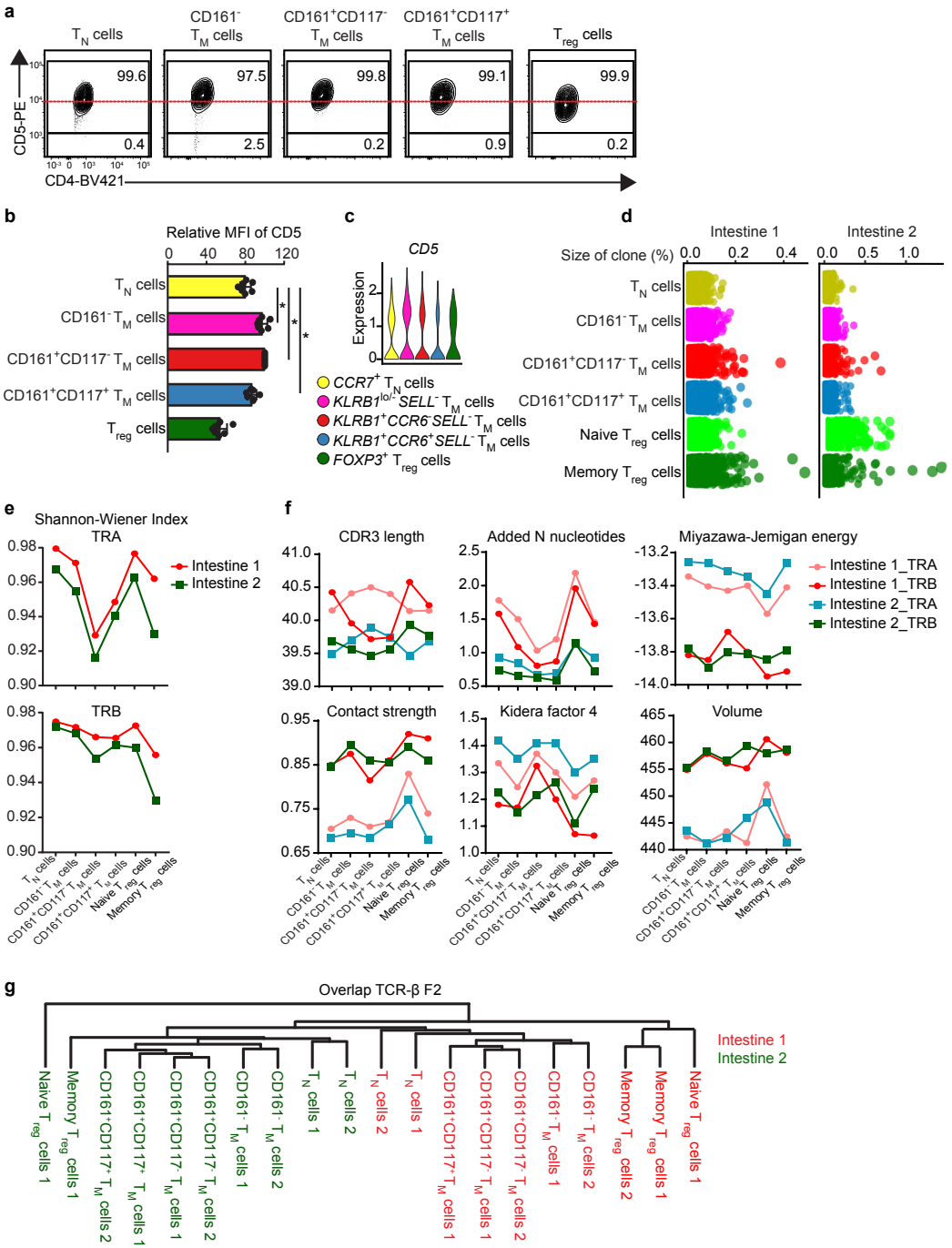
to interact physically with each other (**Supplementary Fig. 5b**). In addition, 23 genes in module 3 could be assigned to cytokine or chemokine receptor pathways, including *CCL20* and its receptor *CCR6*, the interferon receptor *IFNGR1*, TNF family members and IL-1 and IL-17 receptors (**Supplementary Fig. 5b**). Several signaling cascades were also represented in module 3, including the MAPK, TNF, IL-17 and TCR signaling pathways (*FYN*, *LCP2*, *SOS1*, *MAP3K8* kinase, *FASL* and *TNF*) (**Fig. 4d**). The T_H17-associated gene *RORC* was expressed in module 3 (**Fig. 4d**). In addition, the dynamic expression profiles of *FYN*, *FASL* and *TNF* clustered tightly with *KLRB1* (CD161) (**Fig. 4d**) at the point in the pseudotime trajectory where *CCR7*$^+$ T_N cells were aligned next to *SELL*$^-$ T_M cells (**Fig. 4d**). Finally, we quantified the smoothness of cell-to-cell transitioning based on gene expression changes along the trajectory, which showed that the pseudotime trajectory was most uncertain at the beginning and toward the end, but quite robust in the middle, where *CCR7*$^+$ T_N cells were aligned next to *SELL*$^-$ T_M cells (**Fig. 4d**).

In sum, these results identified temporal patterns of gene expression along the single-cell trajectory that is compatible with the transition of cells displaying a T_N phenotype into cells with a T_M phenotype.

TCR analysis reveals clonal expansion of fetal CD4$^+$ T cells

Surface expression of CD5 correlates with TCR avidity[23–26]. Because *CD5* gene expression was upregulated in T_M cells compared to T_N cells, we quantified CD5 expression on all identified fetal intestinal CD4$^+$ T cell subsets using flow cytometry and observed that all the CD4$^+$ T cell subsets expressed CD5 (**Fig. 5a**), but that the median fluorescence intensity (MFI) was higher in CD161$^-$, CD161$^+$CD117$^-$ and CD161$^+$CD117$^+$ T_M cells and lower in CD25$^+$CD127^{lo} T_{reg} cells and CD45RA$^+$ T_N cells (**Fig. 5a–c**), suggesting that cells with a T_M phenotype express TCR with a higher avidity compared to T_N cells.

Fig. 5. CD5 expression analysis and high-throughput TCR-sequencing of fetal intestinal CD4$^+$ T cells. a–b, CD5 expression on the indicated CD4$^+$ T cell clusters. (**a**) The biaxial plots depict one representative experiment, and (**b**) the bar graphs depict the median fluorescence intensity (MFI) of CD5 expression for each cluster relative to the corresponding CD161$^+$CD117$^-$ T_M subpopulation in each fetal intestine (n = 7). Data represent six independent experiments. Error bars indicate mean ± SEM. *p < 0.05, Two-tailed Wilcoxon matched-pairs signed-ranks test. **c**, Expression (log-normalized) of *CD5* gene transcripts in the indicated cell clusters, presented as violin plots. **d**, Dot plots showing the percentage of TCR cDNA molecules per unique TCRβ sequence in each cluster from each fetal intestine. Data are from two independent samples. A single duplicate is shown for samples with technical replicates. **e**, Dot plots showing the normalized Shannon-Wiener index for TCRα (TRA) and TCRβ (TRB) sequences in each cluster from each fetal intestine. Data are from two independent samples. **f**, Dot plots showing averaged TCR repertoire characteristics weighted per clonotype for each cluster. Data are from two independent samples. **g**, Dendrogram showing weighted clonal overlaps for TCRβ nucleotide sequences among clusters, analyzed using the F2 similarity metric in VDJtools. Colors indicate different fetal intestines.



Next, we evaluated the TCR clonotypic architecture of flow-sorted fetal intestinal CD45RA⁺ T_N, CD45RO⁺ T_M and CD25⁺CD127^{lo} T_{reg} subpopulations. Analysis of the

TCRβ rearrangements in a single fetal intestine indicated limited overlap among the distinct subpopulations, most of which were highly polyclonal (not shown). Distinct clonotypes were expanded among CD45RO$^+$ T_M cells compared to CD45RA$^+$ T_N cells (**Supplementary Fig. 6a**). We then used a quantitative high-throughput approach for deep sequencing of TCRα and TCRβ rearrangements in all identified fetal intestinal CD4$^+$ T cell subsets isolated from two additional fetal intestines (**Supplementary Table 2**). Post-analysis of the obtained repertoires was conducted using VDJtools[27]. As expected, all T_M subpopulations showed greater clonality compared to the T_N subpopulation (**Fig. 5d,e**). The averaged characteristics of CDR3 length, added N-nucleotides and physicochemical characteristics of the 5 amino acid residues located in the middle of the CDR3 loop, which are most likely to contact the peptide-MHC complex[28], also differed among all subpopulations (**Fig. 5f**). The latter analysis included the averaged statistical potential of the CDR3 loop with respect to epitope interactions, comprising the estimated "energy" of the interaction with a random epitope[29], the "strength" of the interaction (derivative of "energy", VDJtools[27]), hydrophobicity (Kidera factor 4)[30,31] and "volume" (values from http://www.imgt.org/IMGTeducation/Aide-memoire/_UK/aminoacids/IMGTclasses.html). These analyses provided no evidence for clonal expansion of CD4$^+$ T cells as a function of intrinsically high TCR-avidities for self-derived peptide-MHC complexes (**Fig. 5d-f**), suggesting indirectly that antigen-specific interactions triggered clonal selection of CD4$^+$ T cells from the T_N cell pool. Analysis of V-J segment use (Jensen-Shannon divergence; Supplementary Fig. 6b) and overlaps among repertoires in terms of the weighted proportion of shared TCRβ clonotypes revealed tightly clustered technical replicates and clearly distinguished all subpopulations of CD4$^+$ T cells (Fig. 5g). At the same time, the CD161$^-$, CD161$^+$CD117$^-$ and CD161$^+$CD117$^+$ T_M cells clustered similarly in each fetus, with minimum overlap with the CD45RA$^+$ T_N and CD25$^+$CD127^{lo} T_{reg} cells (**Fig. 5g**). Analysis of the clonal overlap of amino acid CDR3 repertoires between the same populations in the two fetal intestines revealed that the CD161$^-$, CD161$^+$CD117$^-$ and CD161$^+$CD117$^+$ T_M populations displayed much stronger overlap than the CD45RA$^+$ T_N and CD25$^+$CD127^{lo} T_{reg} CD4$^+$ T cells (**Supplementary Fig. 6c**), which could be explained by TCR selection due to exposure to similar foreign antigens. Finally, although the majority of the TCR repertoire was specific for each population, up to 20% of the T cell clones were shared between the CD45RA$^+$ T_N and the three CD45RO$^+$ T_M cell populations (**Supplementary Fig. 6d**), suggesting a potential clonal relationship between CD45RA$^+$ T_N and CD45RO$^+$ T_M cells. These results indicate that avidity-based, clonotype-specific expansion of the T_N pool is associated with T_M formation and confirmed the close relationship between CD161$^-$, CD161$^+$CD117$^-$ and CD161$^+$CD117$^+$ T_M cells.

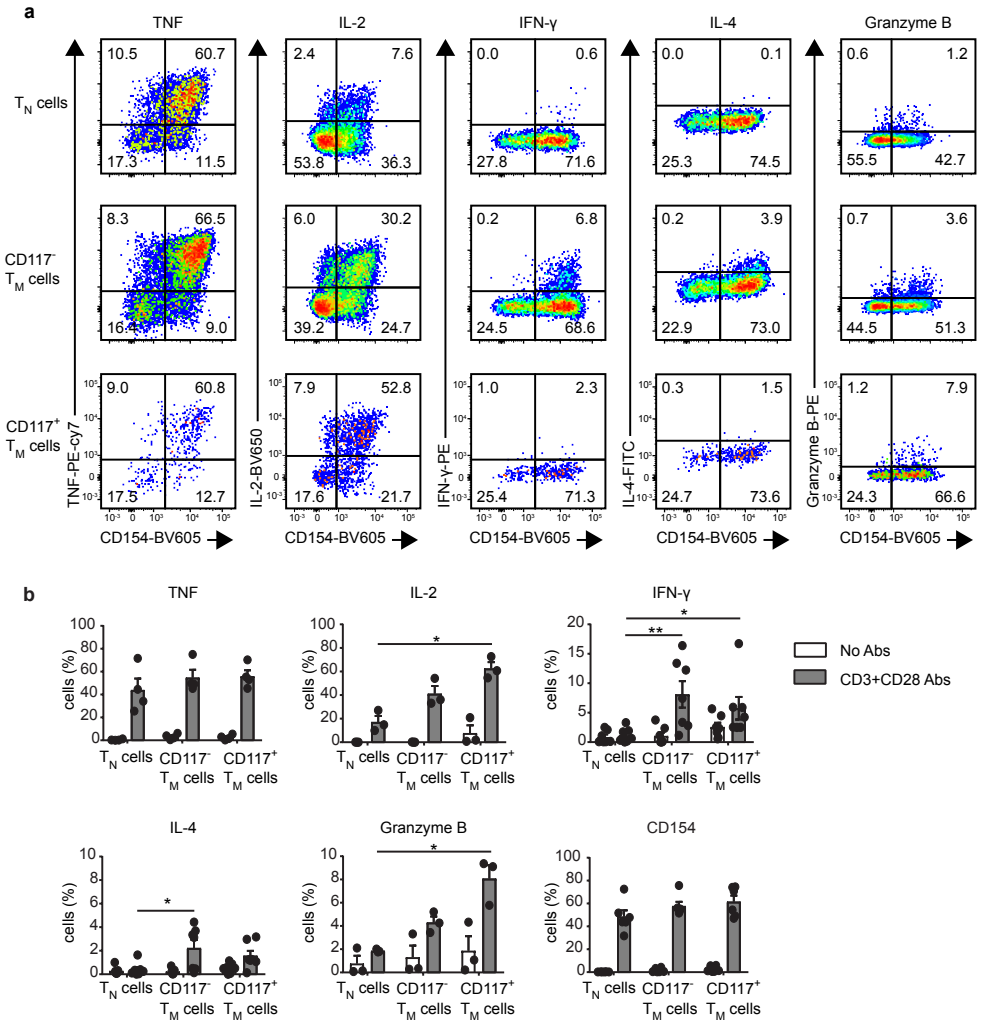


Fig. 6. Functional profiling of fetal intestinal CD4+ T cells. a–b, Purified fetal intestinal CD4+ T cells were treated with a control antibody or stimulated with anti-CD3 and anti-CD28 for 4 h. Intracellular expression of TNF, IL-2, IFN-γ, IL-4, granzyme B, and CD154 was determined for each subpopulation by flow cytometry. (**a**) The biaxial plots show data from one representative experiment after stimulation with anti-CD3 and anti-CD28, and (**b**) the bar charts show data for each subpopulation from each fetal intestine (TNF: n = 4 samples in two independent experiments; IL-2 and granzyme B: n = 3 samples in two independent experiments; IFN-γ, IL-4 and CD154: n = 7 samples in four independent experiments). Error bars indicate mean ± SEM. *p < 0.05, **p <0.01, Kruskal-Wallis test with Dunn's test for multiple comparisons.

Fetal CD4+ T_M cells secrete proinflammatory cytokines

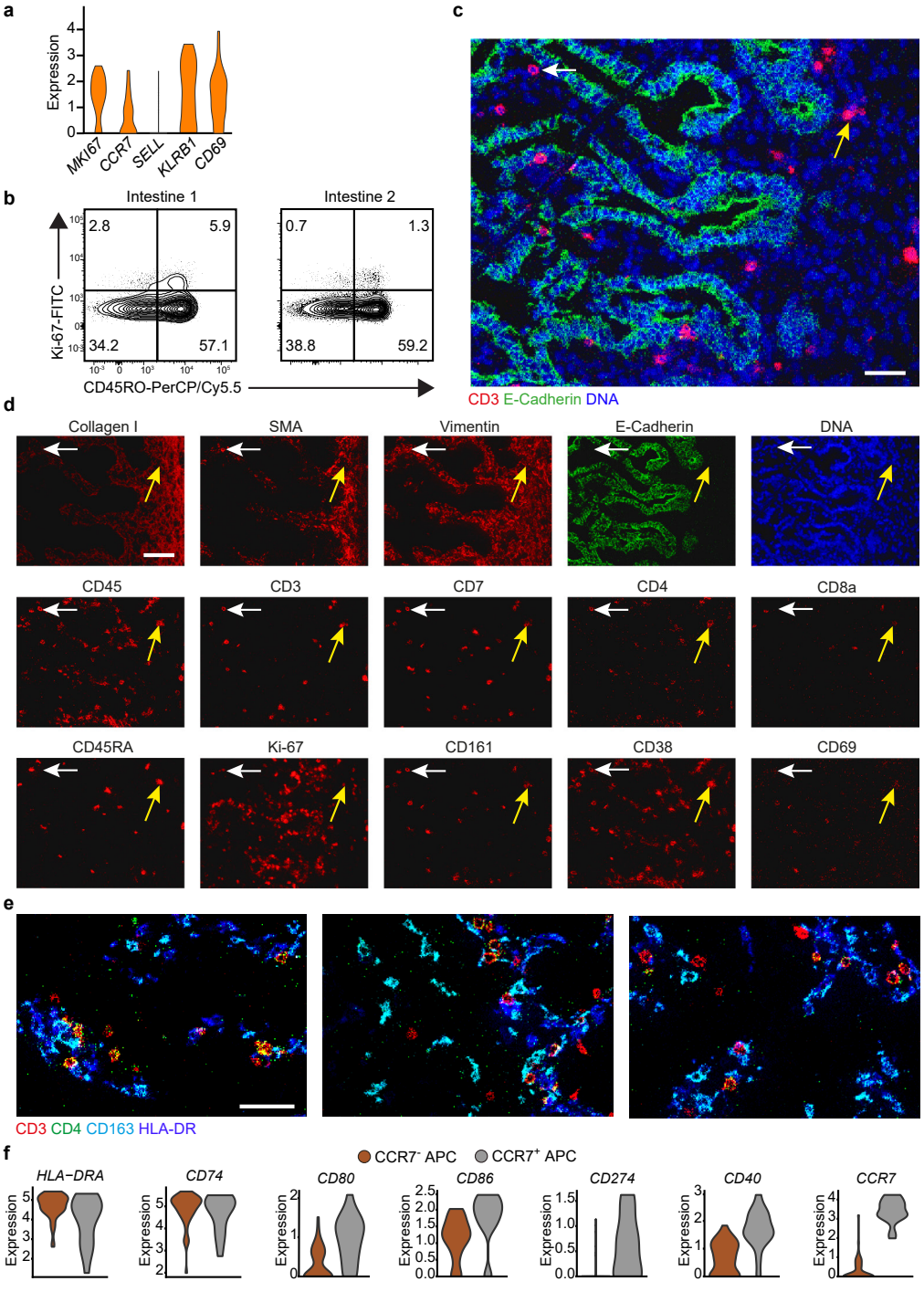
To determine the functional profiles of fetal intestinal CD4+ T cells, we flow-sorted

$CD3^+CD4^+$ T cells and measured expression of TNF, IL-2, IFN-γ, IL-4, granzyme B and IL-17A in $CD45RA^+$ T_N cells and $CD117^-$ and $CD117^+$ T_M cells after cross-linking CD3 and CD28. The activation marker CD154 (CD40L) was upregulated on all cells analyzed (**Fig. 6a,b**), indicating efficient stimulation. All three subpopulations secreted large amounts of TNF (**Fig. 6a,b**), but $CD117^-$ T_M cells and $CD117^+$ T_M cells displayed the highest MFIs (**Supplementary Fig. 7a**). Moreover, IL-2, IFN-γ, IL-4 and granzyme B were more commonly expressed in $CD117^-$ T_M and $CD117^+$ T_M cells relative to $CD45RA^+$ T_N cells (**Fig. 6a,b**). The majority of cytokine-producing $CD4^+$ T cells did not express Ki-67 (**Supplementary Fig. 7b**). Importantly, higher frequencies of $IL\text{-}2^+IFN\text{-}\gamma^+$ cells were detected in the $CD117^-$ T_M and $CD117^+$ T_M cells compared with the $CD45RA^+$ T_N population (**Supplementary Fig. 7c**), suggesting greater polyfunctionality. Although the T_H17-associated *RORC* gene was expressed by 1.3% of $KLRB1^+CCR6^+SELL^-$ T_M cells (**Fig. 4d**), IL-17A production was undetectable by all T_M cells. Thus, fetal intestinal $CD117^-$ T_M and $CD117^+$ T_M cells deployed multiple effector functions reminiscent of classical $CD4^+$ T_M cells in response to TCR-mediated signal transduction and costimulation via CD28.

Fetal $CD4^+$ T cells are co-localized with antigen presenting cells

The single-cell RNA-seq analysis revealed a $MKI67^+$ cluster of proliferating cells, together with high expression of the T_M cell-associated markers *KLRB1* (CD161) and *CD69* and low levels of the T_N cell-associated markers *CCR7* and *SELL* (CD62L) (**Fig. 7a**). Flow cytometry of fetal intestinal $CD4^+$ T cells indicated the presence of $Ki\text{-}67^+$ cells, predominantly within the $CD45RO^+$ compartment (**Fig. 7b**). To assess the spatial distribution of $CD4^+$ T cells *in situ*, we employed a panel of 15 antibodies (**Supplementary Table 3**) combined with a DNA stain to perform imaging-mass cytometry on tissue sections of four human fetal intestinal samples. Stains for collagen I and smooth muscle actin were used to visualize the extracellular matrix of the basement membrane, and the epithelium and lamina propria

Fig. 7. Characterization and spatial localization of fetal intestinal $CD4^+$ T cells and APCs. a, Expression (log-normalized) of the indicated genes in proliferating fetal intestinal $CD4^+$ T cells, presented as violin plot. **b**, Biaxial plots showing expression of Ki-67 vs. CD45RO in the fetal intestinal $CD4^+$ T cell compartment analyzed by flow cytometry. Data represent two independent experiments. **c**, Representative mass cytometry image of a fetal intestine showing the overlay of CD3 (red), Ecadherin (green), and DNA (blue). Scale bar, 100 µm. **d**, Representative mass cytometry images of fetal intestines showing expression of the indicated stromal markers, immune markers, Ki-67 and DNA by the cells identified in (**c**). Scale bar, 200 µm (applies similarly to all the images in (**d**). Yellow arrows: $CD4^+CD45RA^+$ T_N cells; white arrows: $CD4^+CD45RA^-$ T_M cells. **e**, Representative mass cytometry images of a fetal intestine showing the overlay of CD3 (red), CD4(green), CD163 (cyan), and HLA-DR (blue). Scale bar, 50 µm. Colors and scale bars are similar in all three panels. Data in (**b-d**) represent four independent samples in four independent experiments. **f**, Expression (log-normalized) of the indicated genes in two clusters of APCs, presented as violin plots. $CCR7^-$ APCs (n = 49), $CCR7^+$ APCs (n = 17).


a
Expression
MKI67
CCR7
SELL
KLRB1
CD69
b
Intestine 1
Intestine 2
2.8
5.9
34.2
57.1
0.7
1.3
38.8
59.2
Ki-67-FITC
CD45RO-PerCP/Cy5.5
c
CD3 E-Cadherin DNA
d
Collagen I
SMA
Vimentin
E-Cadherin
DNA
CD45
CD3
CD7
CD4
CD8a
CD45RA
Ki-67
CD161
CD38
CD69
e
CD3 CD4 CD163 HLA-DR
f
CCR7⁻ APC
CCR7⁺ APC
HLA−DRA
CD74
CD80
CD86
CD274
CD40
CCR7
Expression

were distinguished as vimentin$^-$E-cadherin$^+$ and vimentin$^+$E-cadherin$^-$, respectively (**Fig. 7c,d**). Most CD4$^+$ T cells localized to the lamina propria (**Fig. 7c,d**). Differential expression of CD45RA further allowed discrimination of CD45RA$^+$ T_N (**Fig. 7c,d**) from CD45RA$^-$ T_M cells in the lamina propria (**Fig. 7c,d**). In addition, all CD4$^+$ T cells expressed CD38, whereas only some CD4$^+$ T cells expressed CD69 (**Fig. 7d**). Using a second panel comprising 10 antibodies (**Supplementary Table 3**), we found that CD4$^+$ T cells frequently colocalized with CD163$^+$HLA-DR$^+$ APCs (**Fig. 7e**). Moreover, the single-cell RNA-seq analysis of fetal intestinal cells revealed two cluster of cells displaying high expression of gene transcripts encoding HLA-DR, CD74 (HLA-class II invariant chain), inhibitory molecule PD-L1 (CD274), CD80 and CD86, typically found in APCs. Moreover, these APCs expressed gene transcripts encoding CD40, consistent with an activated phenotype (**Fig. 7f**), whereas stimulated fetal intestinal CD4$^+$ T cells expressed CD40L (CD154) (**Fig. 6b**). In addition, 25.8% of APCs had high expression of CCR7, potentially enabling migration to the mesenteric lymph nodes (**Fig. 7f**). Collectively, these results indicated the existence of CD4$^+$ T_M cells in the fetal intestine, many of which colocalized in the lamina propria with activated CD163$^+$HLA-DR$^+$ APCs.

Discussion

Here we used mass cytometry and single-cell RNA-seq to characterize CD4$^+$ T cells in the human fetal intestine. Mass cytometry revealed three major populations of fetal intestinal CD4$^+$ T cells (T_N, T_M and T_{reg} cells), that could be further distinguished into eight distinct cells clusters that displayed additional heterogeneity. These cell clusters were present in seven human fetal intestines, suggesting a physiologically robust immune composition. Single-cell RNA-seq revealed the presence of seven CD4$^+$ T cell subpopulations, five of which displayed phenotypic overlap with the mass cytometry-defined CD4$^+$ T cell subpopulations. We used computational tools to construct putative CD4$^+$ T cell differentiation trajectories. Using adapted t-SNE[32], we obtained remarkably similar trajectories for the mass cytometry and RNA-seq data. We identified three distinct gene expression modules along the differentiation trajectory that correspond to an increase in gene translation and subsequent activation of immune related genes. In addition, high-throughput TCR sequencing indicated clonal expansions within the CD4$^+$ T_M cell pool, consistent with the evidence for cell proliferation within the CD45RO$^+$ T_M pool that was obtained at both the mRNA and protein level. Moreover, CD4$^+$ T_M cells secreted higher amounts of pro-inflammatory cytokines upon TCR triggering compared to CD4$^+$ T_N cells. Finally, fetal intestinal CD4$^+$ T_M cells displayed a tissue-resident profile and

were frequently found to colocalize with APCs in the lamina propria. Together, this suggested that clonotype-specific transcriptional programs regulated by antigen encounter underpinned the formation of $CD4^+$ T_M cells in the fetal intestine.

T cells in umbilical cord and peripheral blood obtained of infants aged 2 months were reported to display a typical $CD45RA^+$ T_N phenotype[3]. The observation herein that a large pool of $CD45RO^+$ cells with a tissue-resident profile populated the fetal intestine suggests the compartmentalization of the immune system early in life. In conjunction with the earlier finding that clonally expanded T cells were present in the fetal intestine, but virtually absent in other fetal organs[4], our results further suggest that memory formation was driven by local exposure to foreign antigens. The observation that there is a substantial overlap in the amino acid CDR3 repertoires of the memory $CD4^+$ T cells compartment in the two fetuses analyzed may indicate exposure to similar foreign antigens.

Approximately 50% of all fetal intestinal $CD4^+$ T cells were $CD161^+$ and transcriptionally distinct from their $CD161^-$ counterparts, consistent with a recent study[33]. The kinetics of *KLRB1* (CD161) expression was preceded by increased expression CD5 and coincided with increased expression of several TCR signaling genes, including *FYN*, *FASL* and *TNF*, suggesting a coordinated program of transcription. Of note, CD161 was identified as a costimulatory molecule in the context of TCR stimulation[33].

Although the mass cytometry and RNA-seq data were largely compatible, there were exceptions. For example, coexpression of *CCR6* and *KIT* among $KLRB1^+CCR6^+SELL^-$ T_M cells was not reflected in the gene expression profiles. Conversely, expression of *ITGAE* (CD103) mRNA was not reflected by protein expression. These anomalies were likely attributable to discordant gene transcription and protein expression[34] and may also relate to differences in sensitivity of the employed techniques.

The presence of a large population of T_N cells in the fetal intestine is in stark contrast to the predominace of T_M cells in the adult intestine. As the T_N cells expressed relatively high amounts of CD31, which demarcates recent thymic emigrants, our results indicate direct migration of recent thymic emigrants into the intestine[35,36]. We propose that antigen-specific priming of T_N cells takes place in the mesenteric lymph nodes followed by migration of the resulting T_M cells to the lamina propria leading to a progressive loss of T_N cells. Similarly memory formation is taking place in the $CD8^+$ T cell compartment (not shown).

Distinct subpopulations of fetal intestinal T_{reg} cells were distinguished by several markers, including high expression of CD25 and Foxp3, and a lack of CD127. In

line with previous results[37], approximately 50% of these cells expressed CD45RO, while the remainder expressed CD45RA. The CD45RA$^+$ T_{reg} cells expressed TCRs with longer CDR3β loops, higher numbers of added N-nucleotides and distinct physicochemical characteristics, suggesting higher affinities for self-antigens compared to CD45RO$^+$ T_{reg} cells[38]. The presence of oligoclonal T cell expansions in fetuses with autoimmune conditions associated with a genetic absence of T_{reg} cells indicate a key role for these cells in immune suppression *in utero*[39].

In conclusion, our study revealed a putative differentiation trajectory in the fetal intestinal CD4$^+$ T cell compartment, consistent with the formation of T_M cells *in utero*, presumably as a consequence of exposure to foreign antigens. These could include non-inherited maternal HLA-molecules[40] and pathogen-derived ligands, which could be derived from amniotic fluid[8,9]. We propose that immune priming in the fetal intestine prepares the infant for the massive influx of bacteria that occurs immediately after birth, with anamnestic responses *in situ* facilitated by the colocalization of CD4$^+$ T_M cells with APCs.

Accessions

Gene Expression Omnibus

GSE122846

Acknowledgements

We thank the Center for Contraception, Abortion and Sexuality (Leiden and The Hague) for collection and provision of fetal material, K. Lodder, T. van Herwaarden, M. Bialecka and F. Wang for dissection of fetal tissues, and S. L. Kloet for assistance with single-cell RNA sequencing. This research was supported by Leiden University Medical Center (N.L., V.v.U., N.G., F.K.), The Netherlands Organization for Scientific Research (NWO Applied Technical Sciences grant 12721 and ZonMW grant 91112008) (T.A., T.H., J.E., B.P.F.L.), the Russian Science Foundation (grant 16-15-00149, (S.A.K., E.S.E., M.I., O.V.B, D.M.C.), the Wellcome Trust (grant 100326/Z/12/Z) (K.L., J.E.M., D.A.P), the European Commission under an MSCA-ITN award (grant no. 675743/ISPIC) (T.A.), and China Scholarship Council (N.L. and N.G.).

Author contributions

N.L., V.v.U. and F.K. conceived the study and wrote the manuscript. N.L. performed most of the experiments with help from V.v.U. and N.G.. N.L. performed most of the data analyses with help from V.v.U., T.A. and B.P.F.L.. S.A.K., K.L., J.E.M., E.S.E., M.I., D.M.C., O.V.B. and D.A.P. performed TCR repertoire analyses. D.A.P. revised the manuscript. N.F.C.C.d.M. helped with imaging-mass cytometry experiments. T.H., V.v.U., J.E., and B.P.F.L. developed Cytosplore. S.C.d.S.L. provided human fetal tissues. All authors discussed the results and helped prepare the final manuscript.

Competing interests

The authors declare no competing financial interests.

References

1. Janeway, C. A., Jr. Approaching the asymptote? Evolution and revolution in immunology. *Cold Spring Harb. Symp. Quant. Biol.* **54 Pt 1,** 1–13 (1989).
2. Fan, X. & Rudensky, A. Y. Hallmarks of Tissue-Resident Lymphocytes. *Cell* **164,** 1198–1211 (2016).
3. Thome, J. J. C. *et al.* Early-life compartmentalization of human T cell differentiation and regulatory function in mucosal and lymphoid tissues. *Nat. Med.* **22,** 72–77 (2015).
4. Bunders, M. J. *et al.* Memory CD4(+)CCR5(+) T cells are abundantly present in the gut of newborn infants to facilitate mother-to-child transmission of HIV-1. *Blood* **120,** 4383–4390 (2012).
5. Zhang, X. *et al.* CD4 T cells with effector memory phenotype and function develop in the sterile environment of the fetus. *Sci. Transl. Med.* **6,** 238ra72 (2014).
6. Stout, M. J. *et al.* Identification of intracellular bacteria in the basal plate of the human placenta in term and preterm gestations. *Am. J. Obstet. Gynecol.* **208,** 226.e1–7 (2013).
7. Aagaard, K. *et al.* The placenta harbors a unique microbiome. *Sci. Transl. Med.* **6,** 237ra65 (2014).
8. Collado, M. C., Rautava, S., Aakko, J., Isolauri, E. & Salminen, S. Human gut colonisation may be initiated in utero by distinct microbial communities in the placenta and amniotic fluid. *Sci. Rep.* **6,** 23129 (2016).
9. DiGiulio, D. B. Diversity of microbes in amniotic fluid. *Semin. Fetal Neonatal Med.* **17,** 2–11 (2012).
10. Ardissone, A. N. *et al.* Meconium microbiome analysis identifies bacteria correlated with premature birth. *PLoS One* **9,** e90784 (2014).
11. Hornef, M. & Penders, J. Does a prenatal bacterial microbiota exist? *Mucosal Immunol.* **10,** 598–601 (2017).
12. Li, N. *et al.* Mass cytometry reveals innate lymphoid cell differentiation pathways in the human fetal intestine. *J. Exp. Med.* **215,** 1383-1396 (2018).
13. van Unen, V. *et al.* Visual analysis of mass cytometry data by hierarchical stochastic neighbour embedding reveals rare cell types. *Nat. Commun.* **8,** 1740 (2017).
14. Maaten, L. van der & Hinton, G. Visualizing Data using t-SNE. *J. Mach. Learn. Res.* **9,** 2579–2605 (2008).
15. Höllt, T. *et al.* Cytosplore: Interactive Immune Cell Phenotyping for Large Single-Cell Datasets. *Comput. Graph. Forum* **35,** 171–180 (2016).
16. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36,** 411-420 (2018).
17. Kimmig, S. *et al.* Two subsets of naive T helper cells with distinct T cell receptor excision circle content in human adult peripheral blood. *J. Exp. Med.* **195,** 789–794 (2002).
18. Angerer, P. *et al.* destiny – diffusion maps for large-scale single-cell data in R. *Bioinformatics* **32,** 1241-1243 (2015).
19. Samusik, N., Good, Z., Spitzer, M. H., Davis, K. L. & Nolan, G. P. Automated mapping of phenotype space with single-cell data. *Nat. Methods* **13,** 493–496 (2016).
20. Jolliffe, I. Principal Component Analysis. in *International Encyclopedia of Statistical Science* 1094–1096 (2011).
21. Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32,** 381–386 (2014).
22. Qiu, X. *et al.* Single-cell mRNA quantification and differential analysis with Census. *Nat. Methods* **14,** 309–315 (2017).
23. Azzam, H. S. *et al.* CD5 Expression Is Developmentally Regulated By T Cell Receptor (TCR) Signals and TCR Avidity. *J. Exp. Med.* **188,** 2301–2311 (1998).
24. Mandl, J. N., Monteiro, J. P., Vrisekoop, N. & Germain, R. N. T Cell-Positive Selection Uses Self-Ligand Binding Strength to Optimize Repertoire Recognition of Foreign Antigens. *Immunity* **38,** 263–274 (2013).
25. Persaud, S. P., Parker, C. R., Lo, W.-L., Scott Weber, K. & Allen, P. M. Intrinsic CD4 T cell sensitivity

and response to a pathogen are set and sustained by avidity for thymic and peripheral complexes of self peptide and MHC. *Nat. Immunol.* **15,** 266–274 (2014).

26. Fulton, R. B. *et al.* The TCR's sensitivity to self peptide-MHC dictates the ability of naive CD8(+) T cells to respond to foreign antigens. *Nat. Immunol.* **16,** 107–117 (2015).

27. Shugay, M. *et al.* VDJtools: Unifying Post-analysis of T Cell Receptor Repertoires. *PLoS Comput. Biol.* **11,** e1004503 (2015).

28. Egorov, E. S. *et al.* The Changing Landscape of Naive T Cell Receptor Repertoire With Human Aging. *Front. Immunol.* **9,** 1618 (2018).

29. Miyazawa, S. & Jernigan, R. L. Residue-residue potentials with a favorable contact pair term and an unfavorable high packing density term, for simulation and threading. *J. Mol. Biol.* **256,** 623–644 (1996).

30. Kidera, A., Konishi, Y., Oka, M., Ooi, T. & Scheraga, H. A. Statistical analysis of the physical properties of the 20 naturally occurring amino acids. *J. Protein Chem.* **4,** 23–55 (1985).

31. Rackovsky, S. Global characteristics of protein sequences and their implications. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 8623–8626 (2010).

32. Pezzotti, N. *et al.* Approximated and User Steerable tSNE for Progressive Visual Analytics. *IEEE Trans. Vis. Comput. Graph.* **23,** 1739–1752 (2017).

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

34. Edfors, F. *et al.* Gene-specific correlation of RNA and protein levels in human cells and tissues. *Mol. Syst. Biol.* **12,** 883 (2016).

35. McFarland, R. D., Douek, D. C., Koup, R. A. & Picker, L. J. Identification of a human recent thymic emigrant phenotype. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 4215–4220 (2000).

36. Staton, T. L. *et al.* CD8+ recent thymic emigrants home to and efficiently repopulate the small intestine epithelium. *Nat. Immunol.* **7,** 482–488 (2006).

37. Michaelsson, J., Mold, J. E., McCune, J. M. & Nixon, D. F. Regulation of T Cell Responses in the Developing Human Fetus. *The Journal of Immunology* **176,** 5741–5748 (2006).

38. Feng, Y. *et al.* A mechanism for expansion of regulatory T-cell repertoire and its role in self-tolerance. *Nature* **528,** 132–136 (2015).

39. Allenspach, E. J. *et al.* Absence of functional fetal regulatory T cells in humans causes in utero organ-specific autoimmunity. *J. Allergy Clin. Immunol.* **140,** 616–619.e7 (2017).

40. Gomez de Agüero, M. *et al.* The maternal microbiota drives early postnatal innate immune development. *Science* **351,** 1296–1302 (2016).

Methods

Sample processing and cell isolation

Fetal tissues were obtained from elective abortions with informed consent. The gestational age ranged from 14 to 22 weeks. Intestines were separated from mesentery, cut into small pieces, embedded in optimal cutting temperature compound, and snap-frozen in isopentane. The remaining intestines were used for single-cell isolation as described previously[12]. Briefly, fetal intestines were cleared of meconium, cut into fine pieces, treated with 1 mM dithiothreitol (Fluka) for 2 x 10 min (replacing buffer) at room temperature (rT), and then incubated with 1 mM ethylenediaminetetraacetic acid (Merck) for 2 x 1 h (replacing buffer) at 37 °C under rotation to separate the epithelium from the lamina propria. To obtain single-cell suspensions from the lamina propria, the intestines were rinsed with Hank's balanced salt solution (Thermo Fisher Scientific), incubated with 10 U/mL collagenase IV (Worthington) and 200 µg/mL DNAse I grade II (Roche Diagnostics) overnight at 37 °C, and filtered through a 70 µm nylon mesh. Isolated cells were then further purified with a Percoll gradient (GE Healthcare). Fetal liver and spleen tissues were cut into small pieces and filtered through a 70 µm nylon cell strainer and the immune cells were isolated with Ficoll-Paque™ density gradient. All the isolated cells were stored in liquid nitrogen. Study approval was granted by the Medical Ethics Commission of Leiden University Medical Centre (protocol P08.087). All experiments were conducted in accordance with local ethical guidelines and the principles of the Declaration of Helsinki.

Cell suspension-mass cytometry

Antibodies used for cell suspension-mass cytometry are listed in **Supplementary Table 1**. Purified antibodies lacking carrier protein were conjugated with metal reporters by using a MaxPar X8 Antibody Labeling Kit (Fluidigm). Procedures for antibody staining and data acquisition were described previously[41]. Briefly, cells from fetal intestines were incubated with 5 µM Cell-ID Intercalator-103Rh (Fluidigm) for 15 min at rT and then stained with a cocktail of metal-conjugated antibodies for 45 min at rT. After washing, cells were incubated with 125 nM Cell-ID Intercalator-Ir (Fluidigm) in MaxPar Fix and Perm Buffer (Fluidigm) overnight at 4 °C. Data were acquired using a CyTOF 2™ mass cytometer (Fluidigm) and normalized using EQ Four Element Calibration Beads with the reference EQ Passport P13H2302 (Fluidigm).

Imaging-mass cytometry

Antibodies used for imaging-mass cytometry are listed in **Supplementary Table 3.** Purified antibodies lacking carrier protein were conjugated with metal reporters by using a MaxPar X8 Antibody Labeling Kit (Fluidigm). Snap-frozen human fetal intestinal biopsies were sectioned at a thickness of 5 µm and fixed by incubating with 1% paraformaldehyde for 5 min at rT followed by 100% methanol for 5 min at –20 °C. After fixation, tissue sections were washed in Dulbecco's phosphate-buffered saline (Thermo Fisher Scientific) containing 0.5% bovine serum albumin and 0.05% Tween, rehydrated in additive-free Dulbecco's phosphate-buffered saline, washed again, and blocked with Superblock Solution (Thermo Fisher Scientific). Tissue sections were then stained with a cocktail of metal-conjugated antibodies overnight at 4 °C, washed, and incubated with 125 nM Cell-ID Intercalator-Ir for 30 min at rT. After a further wash, tissue sections were dipped in Milli-Q water (Merck Millipore) for 1 min and dried for 20 min at rT. Data were acquired using a Hyperion™ imaging-mass cytometer (Fluidigm) at a resolution of 1 µm, with settings aligned to company guidelines. The ablation frequency was 200 Hz, and the energy was 6 dB. Regions of interest were acquired at a size of 1 by 1 mm^2. All data were stored as MCD files and txt files.

Single-cell RNA-sequencing

Single, live, $CD8a^-TCR\gamma\delta^-CD4^+$ cells from the intestines of fetus #6 were sorted using a FACSAria III flow cytometer (BD Biosciences). Post-sort purity was 96.5%. Single-cell RNA-sequencing was performed as described previously[42]. Briefly, cells combined with oil, reagents, and beads were loaded on a Chromium Single Cell Controller (10x Genomics). Lysis and barcoded reverse transcription of polyadenylated mRNA from single cells were performed inside each gel bead emulsion. Next-generation sequencing libraries were prepared in a single bulk reaction, and transcripts were sequenced using a HiSeq4000 System (Illumina).

Integrated data analysis

For cell suspension-mass cytometry, data from single, live, $CD45^+$ cells, gated individually using Cytobank as shown in **Supplementary Fig. 1a**, were sample-tagged and hyperbolic-arcsinh-transformed with a cofactor of 5 using Cytosplore^{+HSNE} software[13]. The major immune lineages shown in **Supplementary Fig. 1b** were then identified at the overview level by performing a 3-level HSNE analysis carried out with default parameters (perplexity: 30; iterations: 1,000). All t-SNE plots and Gaussian Mean-Shift clustering-derived cell clusters were generated in Cytosplore[15]. Hierarchical clustering of the phenotype heatmap was created with Euclidean correction and average linkage clustering in Cytosplore^{+HSNE}. Violin plots for cytometry data were generated in R. Diffusion map plots for mass cytometry

data were generated using the "density" package in R. Single-cell force-directed layouts for mass cytometry data were generated using "VorteX" software[19]. For imaging-mass cytometry, all images were generated using MCD Viewer software v1.0.560 (Fluidigm). For single-cell RNA-seq, single-cell transcriptome sequencing data were processed using the single-cell RNA-seq package "Seurat" in R[16]. The Seurat object was generated by following the criteria that each gene was expressed by at least 3 cells and that at least 200 genes were expressed per cell. Data were further filtered based on the parameters: (i) unique gene count per cell >200 and <2,000; and (ii) mitochondrial percentage of all genes <0.05. After log-normalization, a PCA-reduction analysis (pcs.compute = 20) was performed based on the 2,174 variable genes across single cells. Next, graph-based clustering detection and a t-SNE algorithm were applied to the top 13 PCA-dimensions. The resolution for cluster detection was 0.8. Heatmaps, PCA plots, diffusion map plots, and violin plots of the RNA-seq data were generated using the "Seurat" package. The t-SNE plots for RNA-seq data shown in **Fig. 4b** were generated in Cytosplore^{+HSNE}. Only genes with local standardization (>0.5) across all cells were taken into account. Bar graphs and dot plots (showing mean and SD) were generated in Prism (GraphPad). The pseudotime analysis shown in **Fig. 4c,d** was performed using the Monocle 2 toolkit in R as described previously[22], excluding unrelated T_{reg} cells. Briefly, the single-cell trajectory was inferred using the dpFeature unsupervised procedure to identify variable genes, and the dimensions were reduced using t-SNE on the top high-loading principal components. The top 1,000 significant genes were selected as the ordering genes and reduced with the DDRTree method for the single-cell graph shown in **Fig. 4c**. Variable genes were selected at a significant false discovery rate of <10%, clustered by pseudo-temporal expression patterns, and visualized in a heatmap in **Fig. 4d**. Gene list enrichment analysis was performed using ToppGene[43], gene interaction network analysis was performed using the BioGrid interaction database[44], and gene pathway analysis was performed using the Kyoto Encyclopedia of Genes and Genomes[45].

Flow cytometry

For surface staining, cells were incubated with fluorochrome-conjugated antibodies and human Fc block (BioLegend) for 30–45 min at 4 °C. For intracellular cytokine/CD154 staining, cells were stimulated with CD3/CD28-specific (2.5 µg/mL each, BioLegend) or control antibodies (5 µg/mL, BioLegend) for 4 h at 37 °C. Brefeldin A (10 µg/mL, Sigma) was added for the final 3 h. Cells were then fixed/permeabilized using Fixation Buffer and Intracellular Staining Perm Wash Buffer (BioLegend). For intracellular Foxp3/Ki-67 staining, cells were prepared using a

Foxp3 Staining Buffer Set (eBioscience). Electronic compensation was performed using individually stained CompBeads (BD Biosciences). Cells were acquired using an LSR II cytometer (BD Biosciences) or sorted using a FACSAria III flow cytometer (BD Biosciences) as shown in Supplementary Fig. 3a. Data were analyzed with FlowJo software v10 (Tree Star Inc.). The antibodies used in this study are listed in **Supplementary Table 4**.

TCR repertoire analysis

CD4+ T cell subsets were sorted according to the gating strategy shown in **Supplementary Fig. 3a**. For conventional sequencing, a total of 5,000 cells per subset was sorted directly into RNAlater (Applied Biosystems) using a FACSAria III flow cytometer (BD Biosciences). All expressed TCRβ rearrangements were amplified using a template-switch anchored RT-PCR, sequenced, and analyzed as described previously[46]. Gene use was determined according to the ImMunoGeneTics (IMGT) nomenclature[47].

For high-throughput sequencing, an average of 6,700 ± 2,000 cells per subset was sorted directly into RLT buffer (Qiagen) using a FACSAria III flow cytometer (BD Biosciences). Four volumes of TRIzol (Invitrogen) were then added to the RLT cell lysate. RNA was extracted according to the TRIzol Reagent User Guide. Unique molecular identifier (UMI)-labelled 5'RACE TCRα and TCRβ cDNA libraries were prepared using a Human TCR Profiling Kit (MiLaboratory LLC). All extracted RNA was used for cDNA synthesis, and all synthesized cDNA was used for PCR amplification. Libraries were prepared in parallel using the same number of PCR cycles and sequenced in parallel using a 150 + 150 bp MiSeq System (Illumina). This approach generated a total of 11,310,000 TCRα and TCRβ sequencing reads (250,000 ± 150,000 reads per library), from which 625,000 unique unique UMI-labelled TCR cDNA molecules (13,500 ± 7,000 molecules per library) were extracted using MIGEC[48] and MiXCR[49] software with a threshold of at least 2 sequencing reads per UMI. Each library contained an average of 3,500 ± 1,300 functional (in-frame, without stop-codons) CDR3 nucleotide sequences. Averaged TCR repertoire characteristics weighted by clonotype size were analyzed using VDJtools software[27]. Gene use was determined according to the ImMunoGeneTics (IMGT) nomenclature[47].

Statistics

Results are shown as mean ± SEM. The statistics test used were two-tailed Wilcoxon matched-pairs signed-ranks test and Kruskal-Wallis test with Dunn's test for multiple group comparisons, as appropriate (after normality test). $P \leq 0.05$

was considered to be statistically significant. All statistics were analysed using GraphPad Prism 7 software.

Reporting Summary

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

Data availability

Mass cytometry data are available via Flow Repository (https://flowrepository.org/id/FR-FCM-ZYRD). scRNA-seq data are available via Gene Expression Omnibus accession code GSE122846. The remaining data that support the findings of this study are available from the corresponding author upon reasonable request.

41. van Unen, V. *et al.* Mass Cytometry of the Human Mucosal Immune System Identifies Tissue- and Disease-Associated Immune Subsets. *Immunity* **44,** 1227–1239 (2016).

42. Zheng, G. X. Y. *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat. Commun.* **8,** 14049 (2017).

43. Chen, J., Bardes, E. E., Aronow, B. J. & Jegga, A. G. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res.* **37,** W305–W311 (2009).

44. Bean, D. M. *et al.* esyN: network building, sharing and publishing. *PLoS One* **9,** e106035 (2014).

45. Ogata, H. *et al.* KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **27,** 29–34 (1999).

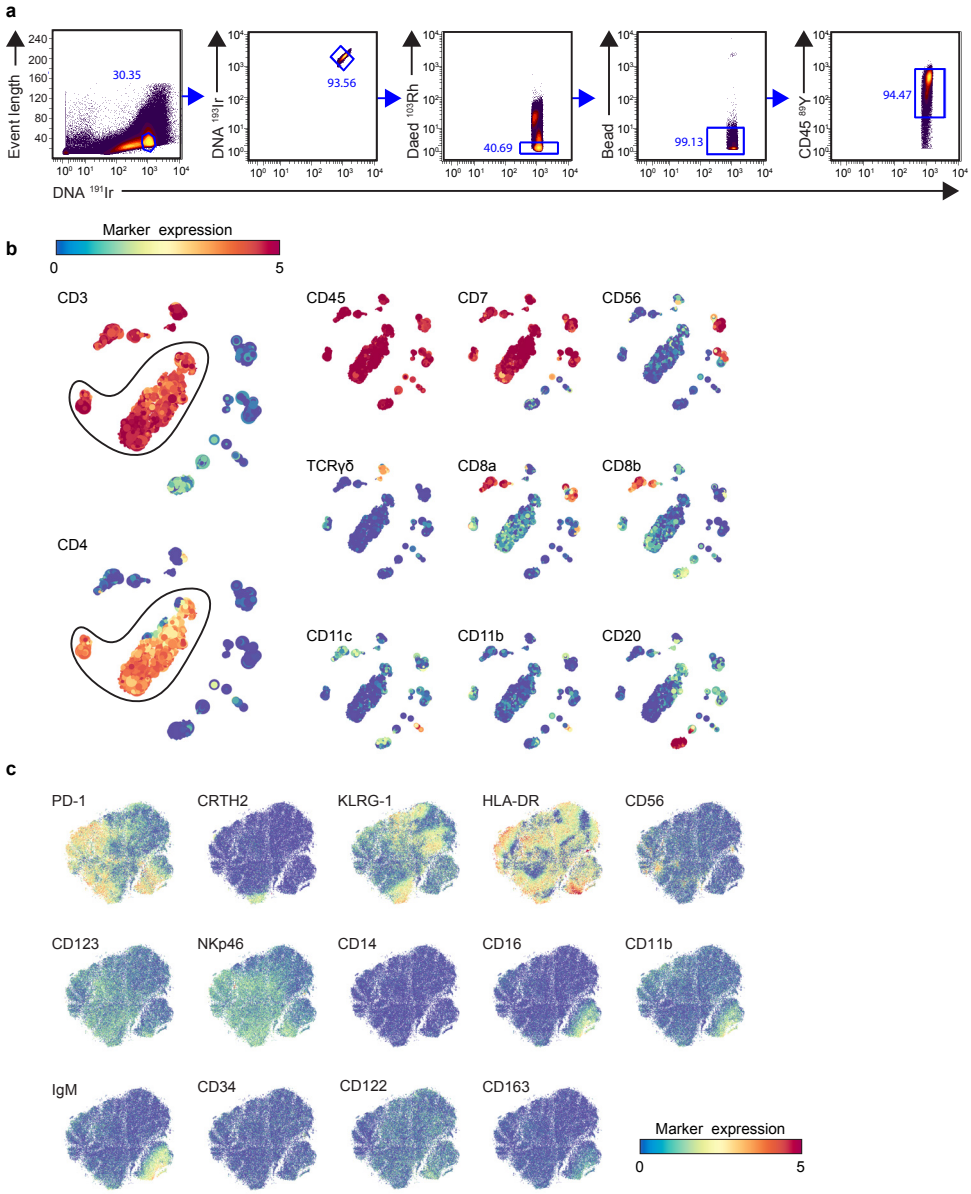
46. Quigley, M. F., Almeida, J. R., Price, D. A. & Douek, D. C. Unbiased molecular analysis of T cell receptor expression using template-switch anchored RT-PCR. *Curr. Protoc. Immunol.* **Chapter 10,** Unit10.33 (2011).

47. Lefranc, M.-P. *et al.* IMGT unique numbering for immunoglobulin and T cell receptor constant domains and Ig superfamily C-like domains. *Dev. Comp. Immunol.* **29,** 185–203 (2005).

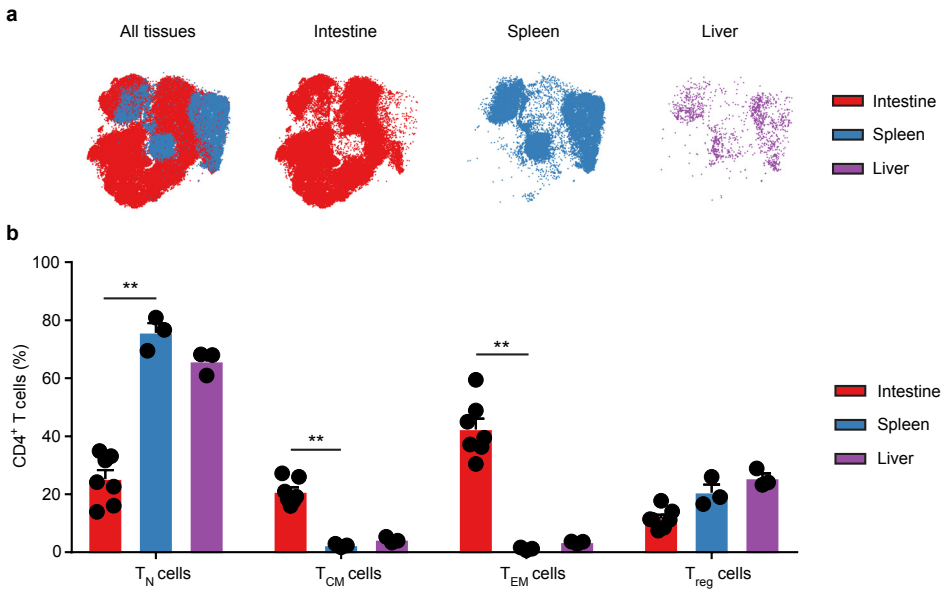
48. Shugay, M. *et al.* Towards error-free profiling of immune repertoires. *Nat. Methods* **11,** 653–655 (2014).

49. Bolotin, D. A. *et al.* MiXCR: software for comprehensive adaptive immunity profiling. *Nat. Methods* **12,** 380–381 (2015).

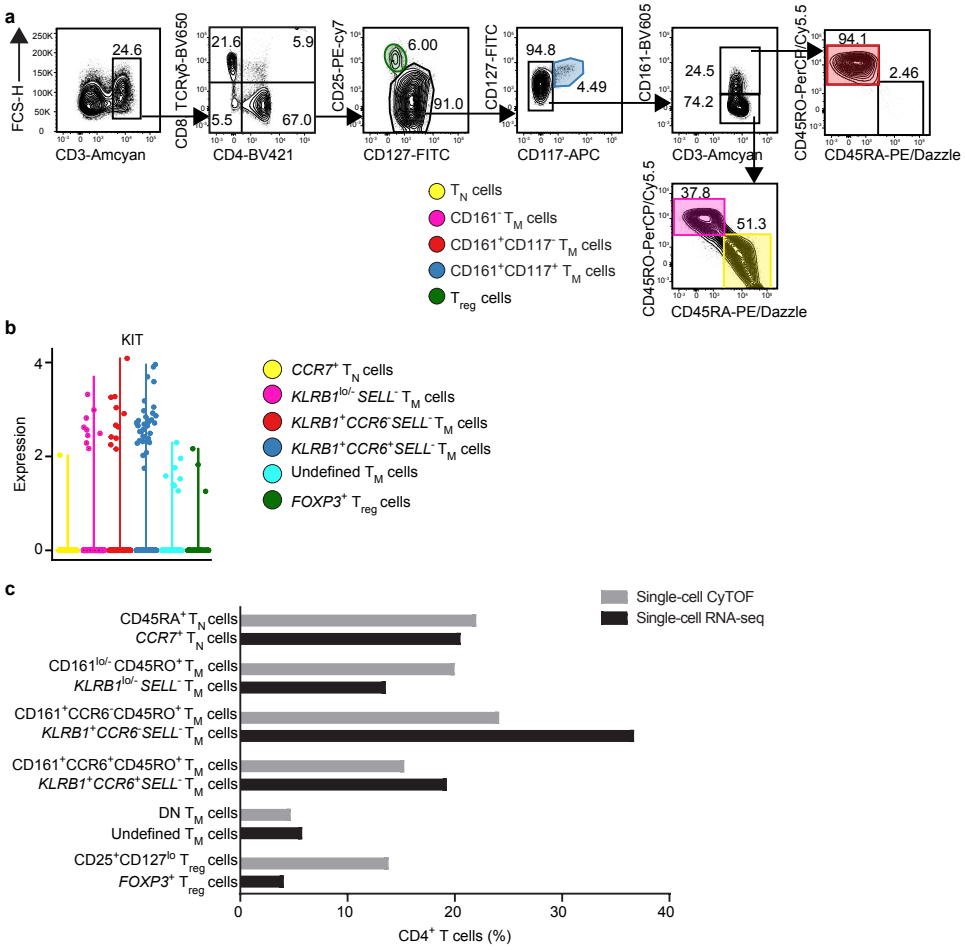
Supplemental information



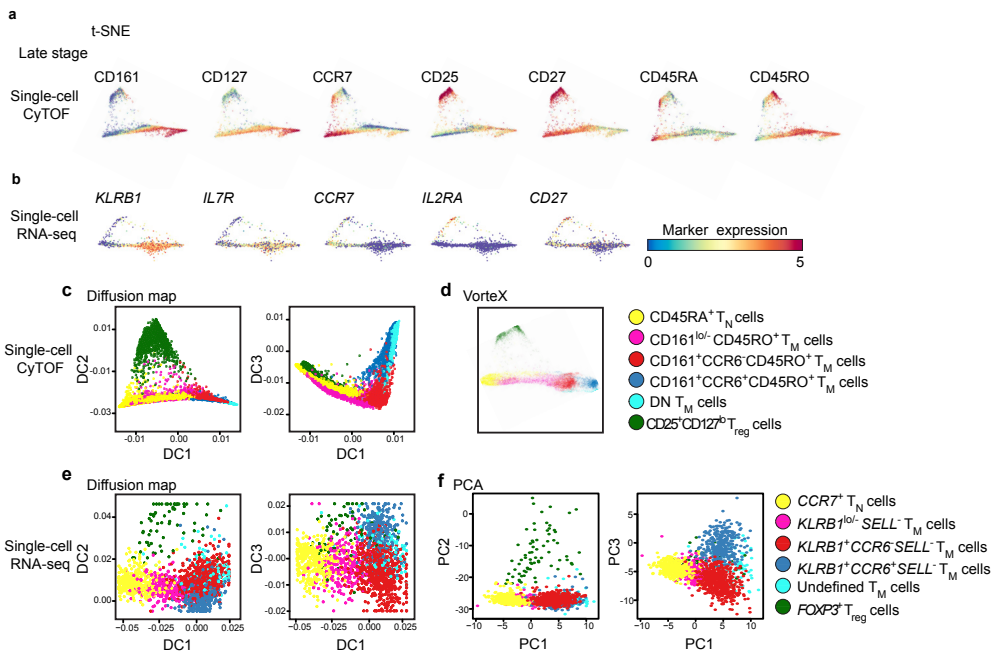
Supplementary Fig. 1. Mass cytometric analysis of fetal intestinal immune cells. (a) Biaxial plots from one fetal intestine showing the gating strategy for single, live, CD45⁺ cells analyzed by mass cytometry (n = 7). **(b)** First-level HSNE embedding of CD45⁺ immune cells (n = 224,286) derived from fetal intestines (n = 7). Each dot represents a landmark. The size of each landmark is proportional to the number of cells that the landmark represents. Colors indicate ArcSinh5-transformed expression values. **(c)** t-SNE embedding showing all CD4⁺ T cells (n = 110,332) derived from fetal intestines (n = 7). Colors indicate the ArcSinh5-transformed expression values.



Supplementary Fig. 2. Mass cytometric analysis of the fetal CD4⁺ T cell compartment across tissues. (**a**) t-SNE embedding of all $CD4^+$ T cells (n = 9.7 x 10^4) derived from fetal intestines (n = 6.6 x 10^4 cells from 7 samples), fetal livers (n = 1,530 cells from 3 samples), and fetal spleens (n = 3.0 x 10^4 cells from 3 samples). Colors indicate tissue origin. Downsampling was performed for samples with more than 10,000 $CD4^+$ T cells. (**b**) Bar graphs showing the quantification of T_{reg} cells and T_N cells, T_{CM} cells, and T_{EM} cells among fetal $CD4^+$ T cells across tissues. T_N cells: $CD45RA^+CCR7^+$; T_{CM} cells: $CD45RA^-CCR7^+$; T_{EM} cells: $CD45RA^-CCR7^-$. Fetal liver and spleen, n = 3 independent samples; fetal intestine, n = 7 independent samples. Error bars indicate mean ± SEM. **p <0.01, Kruskal-Wallis test with Dunn's test for multiple comparisons.



Supplementary Fig. 3. Identification of CD4⁺ T cell clusters in fetal intestines. (**a**) Representative biaxial plots showing the gating strategy for T_N, CD161⁻ T_M, CD161⁺CD117⁻ T_M, CD161⁺CD117⁺ T_M and T_{reg} cell clusters derived from one fetal intestine analyzed by flow cytometry for expression of CD3, CD4, CD8a, CD25, CD45RA, CD45RO, CD117, CD127, CD161 and TCRγδ (n = 10). (**b**) Expression (log-normalized) of *KIT* as determined by single-cell RNA-seq analysis, presented as violin plots. Colors indicate different cell clusters. (**c**) Bar plot depicting cell frequencies for clusters identified by mass cytometry and RNA-seq.



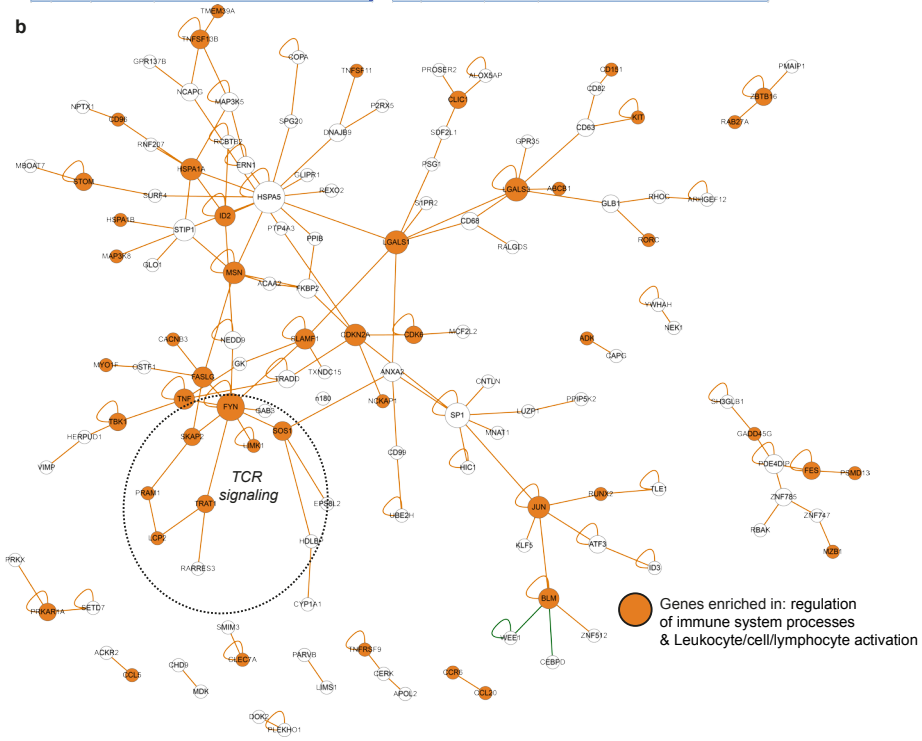
Supplementary Fig. 4. Single-cell trajectories of fetal intestinal CD4⁺ T cells. (**a-b**) t-SNE embeddings of all fetal intestinal CD4⁺ T cells analyzed in (**a**) the mass cytometry dataset (n = 10,436 cells, 35 proteins) and (**b**) the RNA-seq dataset (n = 1,743 cells, 300 variable genes) at the middle of the t-SNE computation. Colors indicate marker expression. (**c**) Diffusion map and (**d**) VorteX analysis of all fetal intestinal CD4⁺ T cells analyzed in mass cytometry dataset (n = 10,436 cells, 35 proteins). Colors indicate different cell clusters. (**e**) Diffusion map and (**f**) PCA analysis of all fetal intestinal CD4⁺ T cells analyzed in the RNA-seq dataset (n = 1,743 cells, 2,174 variable genes). Colors indicate different cell clusters.

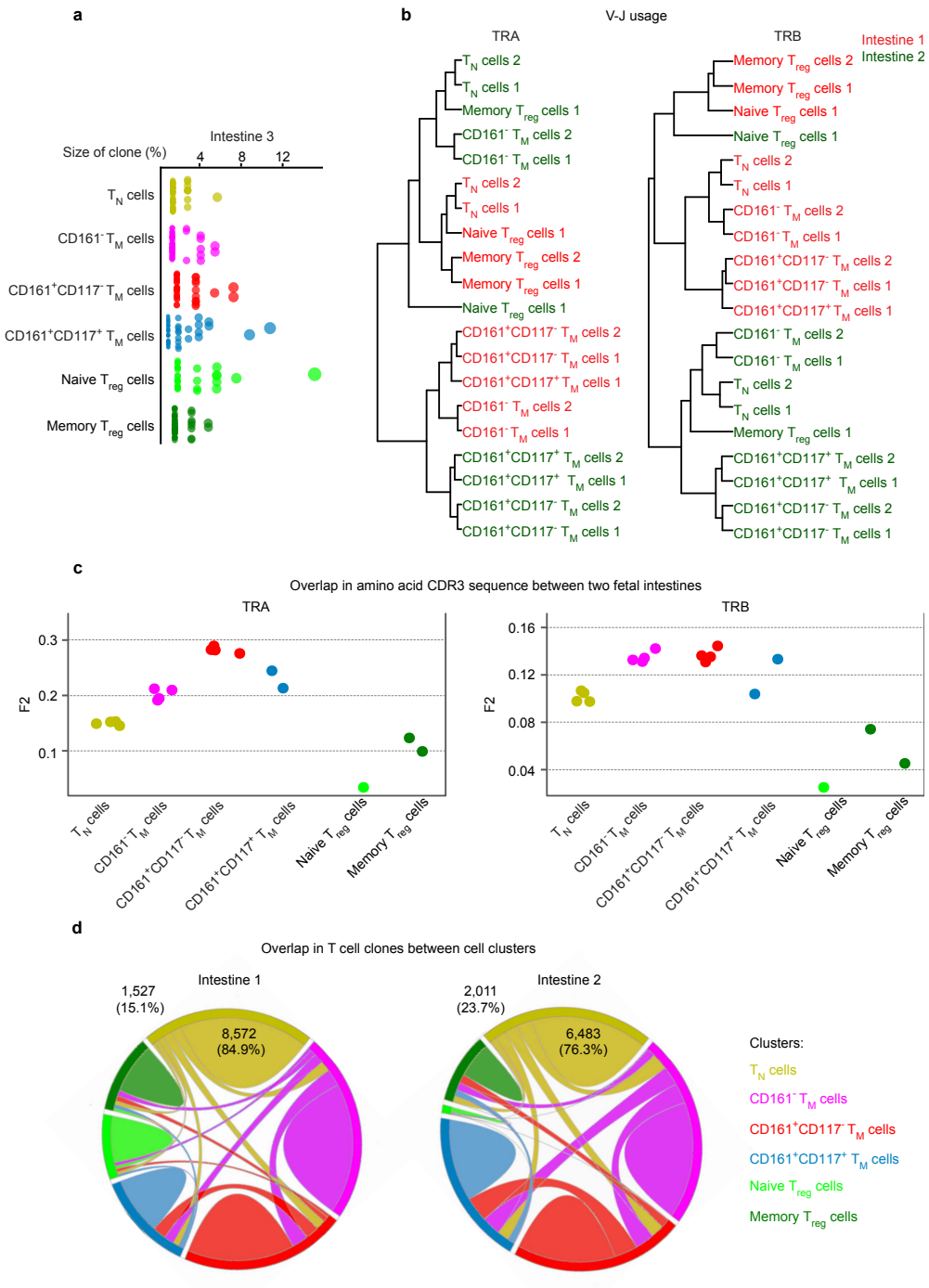
Supplementary Fig. 5. Immune-related genes in the pseudotime-dependent gene module. (**a**) Table listing 106 immune-related genes kinetically enriched in the pseudotime-dependent third gene module (**Fig. 4d**). (**b**) Interaction network of the pseudotime-dependent third gene module using Esyn. Orange lines indicate literature-described physical interactions of the gene-encoded proteins. Orange nodes indicate genes enriched in immunological processes.

a

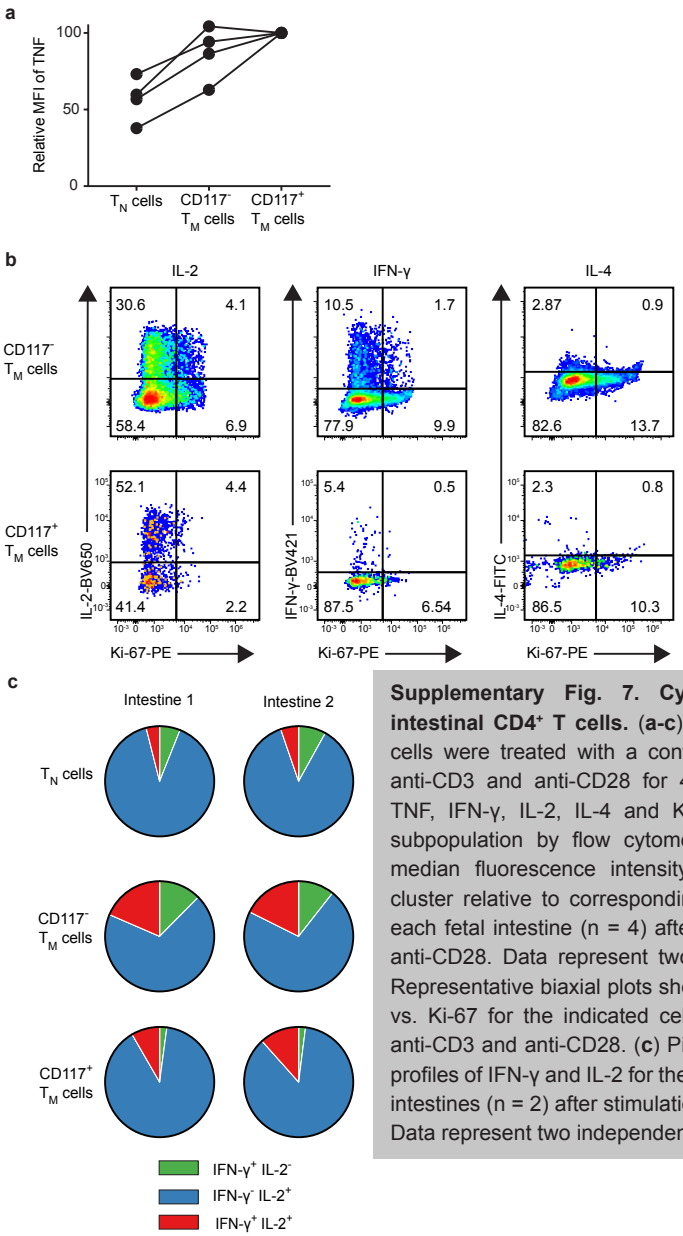
Nr	Gene Symbol	Gene Name	Original Symbol
1	5243	ABCB1	ATP binding cassette subfamily B member 1
2	101	ADAM8	ADAM metallopeptidase domain 8
3	132	ADK	adenosine kinase
4	8943	AP3D1	adaptor related protein complex 3 delta 1 subunit
5	51676	ASB2	ankyrin repeat and SOCS box containing 2
6	641	BLM	Bloom syndrome RecQ like helicase
7	684	BST2	bone marrow stromal cell antigen 2
8	760	CA2	carbonic anhydrase 2
9	784	CACNB3	calcium voltage-gated channel auxiliary subunit beta 3
10	6364	CCL20	C-C motif chemokine ligand 20
11	6352	CCL5	C-C motif chemokine ligand 5
12	1235	CCR6	C-C motif chemokine receptor 6
13	10803	CCR9	C-C motif chemokine receptor 9
14	977	CD151	CD151 molecule (Raph blood group)
15	952	CD38	CD38 molecule
16	966	CD59	CD59 molecule (CD59 blood group)
17	10225	CD96	CD96 molecule
18	1021	CDK6	cyclin dependent kinase 6
19	1029	CDKN2A	cyclin dependent kinase inhibitor 2A
20	160364	CLEC12A	C-type lectin domain family 12 member A
21	64581	CLEC7A	C-type lectin domain containing 7A
22	1192	CLIC1	chloride intracellular channel 1
23	10695	CNPY3	canopy FGF signaling regulator 3
24	10675	CSPG5	chondroitin sulfate proteoglycan 5
25	9646	CTR9	CTR9 homolog, Paf1/RNA polymerase II complex component
26	2833	CXCR3	C-X-C motif chemokine receptor 3
27	23604	DAPK2	death associated protein kinase 2
28	1803	DPP4	dipeptidyl peptidase 4
29	2242	FES	FES proto-oncogene, tyrosine kinase
30	10211	FLOT1	flotillin 1
31	2534	FYN	FYN proto-oncogene, Src family tyrosine kinase
32	10912	GADD45G	grow th arrest and DNA damage inducible gamma
33	10634	GAS2L1	grow th arrest specific 2 like 1
34	2769	GNA15	G protein subunit alpha 15
35	8111	GPR68	G protein-coupled receptor 68
36	3303	HSPA1A	heat shock protein family A (Hsp70) member 1A
37	3304	HSPA1B	heat shock protein family A (Hsp70) member 1B
38	3398	ID2	inhibitor of DNA binding 2
39	3459	IFNGR1	interferon gamma receptor 1
40	3601	IL15RA	interleukin 15 receptor subunit alpha
41	8809	IL18R1	interleukin 18 receptor 1
42	182	JAG1	jagged 1
43	120425	JAML	junction adhesion molecule like
44	3725	JUN	Jun proto-oncogene, AP-1 transcription factor subunit
45	3815	KIT	KIT proto-oncogene receptor tyrosine kinase
46	3820	KLRB1	killer cell lectin like receptor B1
47	3937	LCP2	lymphocyte cytosolic protein 2
48	3956	LGALS1	galectin 1
49	3958	LGALS3	galectin 3
50	3976	LIF	LIF, interleukin 6 family cytokine
51	3984	LIMK1	LIM domain kinase 1
52	7940	LST1	leukocyte specific transcript 1
53	1326	MAP3K8	mitogen-activated protein kinase kinase kinase 8
54	4478	MSN	moesin
55	4542	MYO1F	myosin IF
56	64005	MYO1G	myosin IG
57	51237	MZB1	marginal zone B and B1 cell specific protein
58	4681	NBL1	neuroblastoma 1, DAN family BMP antagonist
59	10787	NCKAP1	NCK associated protein 1
60	114548	NLRP3	NLR family pyrin domain containing 3
61	114770	PGLYRP2	peptidoglycan recognition protein 2
62	118788	PIK3AP1	phosphoinositide-3-kinase adaptor protein 1
63	84106	PRAM1	PML-RARA regulated adaptor molecule 1
64	5573	PRKAR1A	protein kinase cAMP-dependent type I regulatory subunit alpha
65	5719	PSMD13	proteasome 26S subunit, non-ATPase 13
66	26191	PTPN22	protein tyrosine phosphatase, non-receptor type 22
67	5814	PURB	purine rich element binding protein B
68	5873	RAB27A	RAB27A, member RAS oncogene family
69	388	RHOB	ras homolog family member B
70	6095	RORA	RAR related orphan receptor A
71	6097	RORC	RAR related orphan receptor C
72	860	RUNX2	runt related transcription factor 2
73	8935	SKAP2	src kinase associated phosphoprotein 2
74	84174	SLA2	Src like adaptor 2
75	6504	SLAMF1	signaling lymphocytic activation molecule family member 1
76	23657	SLC7A11	solute carrier family 7 member 11
77	6654	SOS1	SOS Ras/Rac guanine nucleotide exchange factor 1
78	2040	STOM	stomatin
79	8464	SUPT3H	SPT3 homolog, SAGA and STAGA complex component
80	29110	TBK1	TANK binding kinase 1
81	55254	TMEM39A	transmembrane protein 39A
82	7124	TNF	tumor necrosis factor
83	8784	TNFRSF18	TNF receptor superfamily member 18
84	8600	TNFSF11	TNF superfamily member 11
85	10673	TNFSF13B	TNF superfamily member 13b
86	8740	TNFSF14	TNF superfamily member 14
87	50852	TRAT1	T cell receptor associated transmembrane adaptor 1
88	28526	TRDC	T cell receptor delta constant
89	53347	UBASH3A	ubiquitin associated and SH3 domain containing A
90	6375	XCL1	X-C motif chemokine ligand 1
91	7704	ZBTB16	zinc finger and BTB domain containing 16
92	80149	ZC3H12A	zinc finger CCCH-type containing 12A
93	3604	TNFRSF9	TNF receptor superfamily member 9
94	5167	ENPP1	ectonucleotide pyrophosphatase/phosphodiesterase 1
95	1075	CTSC	cathepsin C
96	8807	IL18RAP	interleukin 18 receptor accessory protein
97	4217	MAP3K5	mitogen-activated protein kinase kinase kinase 5
98	115361	GBP3	guanylate binding protein 4
99	6846	XCL2	X-C motif chemokine ligand 2
100	9934	P2RY14	purinergic receptor P2Y14
101	27350	APOBEC3C	apolipoprotein B mRNA editing enzyme catalytic subunit 3C
102	6967	TRGC2	T cell receptor gamma constant 2
103	4938	OAS1	2'-5'-oligoadenylate synthetase 1
104	8530	CST7	cystatin F
105	356	FASLG	Fas ligand
106	976	ADGRE5	adhesion G protein-coupled receptor E5

b





Supplementary Fig. 6. TCR sequencing of distinct CD4⁺ T cell populations. (**a**) Dot plot showing the percentage of reads per unique TCRβ sequence in each cluster. Number of unique TCRβ sequences per cluster: T_N cells, 59; CD161⁻ T_M cells, 54; CD161⁺CD117⁻ T_M cells, 37; CD161⁺CD117⁺ T_M cells, 51;



Supplementary Fig. 7. Cytokine production by fetal intestinal CD4⁺ T cells. (**a-c**) Purified fetal intestinal CD4⁺ T cells were treated with a control antibody or stimulated with anti-CD3 and anti-CD28 for 4 h. Intracellular expression of TNF, IFN-γ, IL-2, IL-4 and Ki-67 was determined for each subpopulation by flow cytometry. (**a**) Dot plot showing the median fluorescence intensity (MFI) of TNF for each cell cluster relative to corresponding CD117⁺ T_M subpopulation in each fetal intestine (n = 4) after stimulation with anti-CD3 and anti-CD28. Data represent two independent experiments. (**b**) Representative biaxial plots showing coexpression of cytokines vs. Ki-67 for the indicated cell clusters after stimulation with anti-CD3 and anti-CD28. (**c**) Pie charts depicting coexpression profiles of IFN-γ and IL-2 for the indicated cell clusters from fetal intestines (n = 2) after stimulation with anti-CD3 and anti-CD28. Data represent two independent experiments.

Naive T_{reg} cells, 29; Memory T_{reg} cells, 50. (**b**) Hierarchical clustering of CD4⁺ T subpopulations based on individual V-J rearrangements in theTCRα (TRA) and TCRβ (TRB) datasets. Colors indicate different fetal intestines. (**c**) Dot plot showing weighted clonal overlaps for TCRβ amino acid sequences among clusters from two different fetal intestines analyzed using the F2 similarity metric in VDJtools. (**d**) Chord diagram showing clonal overlaps among different cell clusters. Numbers in the yellow circle indicate the total numbers and frequencies (parentheses) of private TCRβ clonotypes among T_N cells. Numbers outside the plot border indicate the total numbers and frequencies (parentheses) of unique TCRβ clonotypes shared among T_N cells and other CD4⁺ T cell subpopulations. (**a**) Conventional sequencing, (**b-d**) high-throughput sequencing.

	Antigen	Tag	Clone	Supplier	Cat.	Final dilution
1	CD127	^{165}Ho	AO19D5	Flui	3165008B	1/800
2	CCR6	^{141}Pr	G034E3	Flui	3141003A	1/200
3	CD8a	^{146}Nd	RPA-T8	Flui	3146001B	1/200
4	CD11c	^{162}Dy	Bu15	Flui	3162005B	1/200
5	CD38	^{172}Yb	HIT2	Flui	3172007B	1/200
6	CD45	^{89}Y	HI30	Flui	3089003B	1/100
7	CD117	^{143}Nd	104D2	Flui	3143001B	1/100
8	CD4	^{145}Nd	RPA-T4	Flui	3145001B	1/100
9	CD16	^{148}Nd	3G8	Flui	3148004B	1/100
10	CD25	^{149}Sm	2A3	Flui	3149010B	1/100
11	CD123	^{151}Eu	6H6	Flui	3151001B	1/100
12	CD7	^{153}Eu	CD7-6B7	Flui	3153014B	1/100
13	CD163	^{154}Sm	GHI/61	Flui	3154007B	1/100
14	CCR7	^{159}Tb	G043H7	Flui	3159003A	1/100
15	CD14	^{160}Gd	M5E2	Flui	3160001B	1/100
16	CD161	^{164}Dy	HP-3G10	Flui	3164009B	1/100
17	CD27	^{167}Er	O323	Flui	3167002B	1/100
18	CD45RA	^{169}Tm	HI100	Flui	3169008B	1/100
19	CD3	^{170}Er	UCHT1	Flui	3170001B	1/100
20	PD-1	^{175}Lu	EH 12.2H7	Flui	3175008B	1/100
21	CD56	^{176}Yb	NCAM16.2	Flui	3176008B	1/100
22	CD11b	^{144}Nd	ICRF44	Flui	3144001B	1/100
23	TCRγδ	^{152}Sm	11F2	Flui	3152008B	1/50
24	HLA-DR	^{168}Er	L243	BioL	307651	1/200
25	CD20	^{163}Dy	2H7	BioL	302343	1/200
26	CD34	^{142}Nd	HIB19	BioL	343531	1/100
27	IgM	^{150}Nd	MHM88	BioL	314527	1/100
28	CD103	^{155}Gd	Ber-ACT8	BioL	350202	1/100
29	CRTH2	^{156}Gd	BM16	BioL	350102	1/100
30	CD28	^{171}Yb	CD28.2	BioL	302902	1/100
31	CD45RO	^{173}Yb	UCHL1	BioL	304239	1/100
32	CD122	^{158}Gd	TU27	BioL	339002	1/50
33	KLRG-1	^{161}Dy	REA261	MACS	120-014-229	1/50
34	CD8b	^{166}Er	SIDI8BEE	ebio	14-5273	1/50
35	NKp46	^{174}Yb	9E2	BioL	331902	1/40
Fluidigm (Flui), eBioscience (eBio) and Biolegend (BioL).						

Supplementary Table 1 CyTOF antibody panel

Metadata						Basic_Statistics				
sample_id	fetus	subset	chain	replicate	cells sorted	Total_reads	cDNA_molecules_UMI	Reads_per_UMI_threshold	cDNA_molecules_UMI_after_filtering	Clonotypes
Intestine 1_CD117T_TRA_1	Intestine 1	CD161+CD117+ Tm	TRA	1	7140	175103	45659	2	11859	3724
Intestine 2_CD117T_TRA_1	Intestine 2	CD161+CD117+ Tm	TRA	1	7507	556612	98407	2	18135	3970
Intestine 2_CD117T_TRA_2	Intestine 2	CD161+CD117+ Tm	TRA	2	3124	281838	50527	2	7599	2026
Intestine 1_memonyCD161neg_TRA_2	Intestine 1	CD161- Tm	TRA	2	7888	333917	66317	2	14426	4915
Intestine 1_memonyCD161neg_TRA_1	Intestine 1	CD161- Tm	TRA	1	7731	254378	54054	2	13125	4713
Intestine 2_memonyCD161neg_TRA_2	Intestine 2	CD161- Tm	TRA	2	7714	554393	115584	2	16868	4541
Intestine 2_memonyCD161neg_TRA_1	Intestine 2	CD161- Tm	TRA	1	7617	197813	44832	2	6826	3073
Intestine 1_memonyCD161pos_TRA_2	Intestine 1	CD161+CD117- Tm	TRA	2	7600	528867	103916	2	13493	4093
Intestine 1_memonyCD161pos_TRA_1	Intestine 1	CD161+CD117- Tm	TRA	1	7556	219828	49602	2	10631	3671
Intestine 2_memonyCD161pos_TRA_2	Intestine 2	CD161+CD117- Tm	TRA	2	7481	741432	175332	2	21492	3714
Intestine 2_memonyCD161pos_TRA_1	Intestine 2	CD161+CD117- Tm	TRA	1	7581	360210	91069	2	17586	3600
Intestine 1_memonyTreg_TRA_2	Intestine 1	Memory Tregs	TRA	2	2343	111156	23571	2	3957	1442
Intestine 1_memonyTreg_TRA_1	Intestine 1	Memory Tregs	TRA	1	7536	119133	40023	2	8774	2757
Intestine 2_memonyTreg_TRA_1	Intestine 2	Memory Tregs	TRA	1	7217	353446	93680	2	15382	2599
Intestine 1_NaiveCD161neg_TRA_1	Intestine 1	Tn	TRA	1	7697	246744	57601	2	9071	4182
Intestine 1_NaiveCD161neg_TRA_2	Intestine 1	Tn	TRA	2	7654	153032	42957	2	9550	4278
Intestine 2_NaiveCD161neg_TRA_2	Intestine 2	Tn	TRA	2	7672	298840	69436	2	11419	4289
Intestine 2_NaiveCD161neg_TRA_1	Intestine 2	Tn	TRA	1	7542	418501	82985	2	8439	3216
Intestine 1_nTreg_TRA_1	Intestine 1	Naive Tregs	TRA	1	7140	66778	22677	2	5103	2916
Intestine 2_nTreg_TRA_1	Intestine 2	Naive Tregs	TRA	1	716	31684	8182	2	1267	531
Intestine 1_CD117T_TRB_1	Intestine 1	CD161+CD117+ Tm	TRB	1	7140	219407	53164	2	18482	4555
Intestine 2_CD117T_TRB_1	Intestine 2	CD161+CD117+ Tm	TRB	1	7507	350183	83370	2	25446	4521
Intestine 2_CD117T_TRB_2	Intestine 2	CD161+CD117+ Tm	TRB	2	3124	147361	37373	2	10450	2241
Intestine 1_memonyCD161neg_TRB_2	Intestine 1	CD161- Tm	TRB	2	7888	209909	60182	2	20581	5313
Intestine 1_memonyCD161neg_TRB_1	Intestine 1	CD161- Tm	TRB	1	7731	294147	74522	2	21375	5517
Intestine 2_memonyCD161neg_TRB_2	Intestine 2	CD161- Tm	TRB	2	7714	208020	56757	2	24319	5202
Intestine 2_memonyCD161neg_TRB_1	Intestine 2	CD161- Tm	TRB	1	7617	117422	37026	2	13320	4200
Intestine 1_memonyCD161pos_TRB_2	Intestine 1	CD161+CD117- Tm	TRB	2	7600	260491	64692	2	19719	4885
Intestine 1_memonyCD161pos_TRB_1	Intestine 1	CD161+CD117- Tm	TRB	1	7556	161120	46178	2	16619	4452
Intestine 2_memonyCD161pos_TRB_2	Intestine 2	CD161+CD117- Tm	TRB	2	7481	479970	129609	2	33397	4352
Intestine 2_memonyCD161pos_TRB_1	Intestine 2	CD161+CD117- Tm	TRB	1	7581	274782	79895	2	26100	4292
Intestine 1_memonyTreg_TRB_2	Intestine 1	Memory Tregs	TRB	2	2343	69426	16249	2	5620	1444
Intestine 1_memonyTreg_TRB_1	Intestine 1	Memory Tregs	TRB	1	7536	141814	38439	2	13827	3133
Intestine 2_memonyTreg_TRB_1	Intestine 2	Memory Tregs	TRB	1	7217	237380	72338	2	22032	2870
Intestine 1_NaiveCD161neg_TRB_1	Intestine 1	Tn	TRB	1	7697	391216	77080	2	16174	5306
Intestine 1_NaiveCD161neg_TRB_2	Intestine 1	Tn	TRB	2	7654	128811	42777	2	15246	5040
Intestine 2_NaiveCD161neg_TRB_2	Intestine 2	Tn	TRB	2	7672	139563	53435	2	15449	4928
Intestine 2_NaiveCD161neg_TRB_1	Intestine 2	Tn	TRB	1	7542	226971	59369	2	15403	4017
Intestine 1_nTreg_TRB_1	Intestine 1	Naive Tregs	TRB	1	7140	70846	30168	2	9278	3812
Intestine 2_nTreg_TRB_1	Intestine 2	Naive Tregs	TRB	1	716	19528	7090	2	2010	585

Supplementary Table 2 Information of TCR repertoire analysis with NGS

	Antigen	Tag	Clone	Supplier	Cat.	Final dilution	Panel
1	CD45	^{89}Y	HI30	Flui	3089003B	1/50	1 and 2
2	CD3	^{170}Er	UCHT1	Flui	3170001B	1/100	1 and 2
3	CD7	^{153}Eu	CD7-6B7	Flui	3153014B	1/100	1 and 2
4	CD4	^{145}Nd	RPA-T4	Flui	3145001B	1/50	1 and 2
5	CD38	^{172}Yb	HIT2	Flui	3172007B	1/50	1 and 2
6	CD8a	^{146}Nd	RPA-T8	Flui	3146001B	1/50	1 and 2
7	CD45RA	^{169}Tm	HI100	Flui	3169008B	1/100	1 and 2
8	Ki-67	^{166}Er	D3B5	CST	CST9129BF	1/200	1
9	CD161	^{164}Dy	HP-3G10	Flui	3164009B	1/50	1 and 2
10	CD69	^{144}Nd	FN50	Flui	3149010B	1/50	1
11	CD163	^{154}Sm	GHI/61	Flui	3154007B	1/100	1 and 2
12	HLA-DR	^{168}Er	L243	BioL	307651	1/300	2
13	Collagen I	^{147}Sm	polyclonal	Millipore	AB758	1/100	1
14	Vimentin	^{175}Lu	D21H3	CST	CST5741BF	1/200	1
15	SMA	^{148}Nd	1A4	CST	CST5685BF	1/200	1
16	E-Cadherin	^{150}Nd	2.40E+11	CST	CST3195BF	1/50	1
	Fluidigm (Flui), Cell Signaling Technology (CST) and Biolegend (BioL)						

Supplementary Table 3 Imaging-mass cytometry antibody panel.

	Antigen	clone	Flourchrome	Supplier
1	CD3	SK7	Amcyan	BD
2	CD4	RPA-T4	BV421	BioL
3	CD5	L17F12	PE	BD
4	CD8a	RPA-T8	BV650	BD
5	TCRγδ	11F2	BV650	BD
6	CD25	M-A251	PE-cy7	BD
7	CD31	WM59	PE	BD
8	CD45RO	UCHL1	PerCP/Cy5.5	BioL
9	CD45RA	HI100	PE/Dazzle	BioL
10	CD62L	SK11	PE	BD
11	CD69	FN50	PE	BD
12	CD127	A019D5	PE	BioL
13	CD127	A019D5	FITC	BioL
14	CD127	A019D5	PE-cy7	BioL
15	CD117	YB5.B8	APC	BD
16	CD154	24-31	BV605	BioL
17	CD161	DX12	BV605	BD
18	CXCR3	1C6	PE	BD
19	CCR4	1G1	PE	BD
20	DNAM-1	DX11	PE	BD
21	TNF	MAb11	PE-cy7	ebio
22	IFN-γ	4S.B3	PE	BioL
23	IFN-γ	4S.B3	BV421	BioL
24	IL-4	MP4-25D2	FITC	BD
25	IL-17A	BL168	BV421	BioL
26	IL-2	MQ1-17H12	BV650	BioL
27	Ki-67	20Raj1	FITC	eBio
28	Ki-67	20Raj1	PE	eBio
29	Granzyme B	GB11	PE	eBio
30	Foxp3	PCH101	PE	eBio
	eBioscience (eBio), and Biolegend (BioL).			

Supplementary Table 4 Flow cytometry antibody panel.