Universiteit
Leiden
The Netherlands

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

Li, N.

Citation

Li, N. (2019, October 8). *Development of the human fetal immune system: novel insights from high-dimensional single-cell technologies*. Retrieved from https://hdl.handle.net/1887/78475

Version:	Publisher's Version
License:	Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden
Downloaded from:	https://hdl.handle.net/1887/78475

Note: To cite this publication please use the final published version (if applicable).

Cover Page

Universiteit Leiden

Leiden University Repository

The following handle holds various files of this Leiden University dissertation:
http://hdl.handle.net/1887/78475

Author: Li, N.
Title: Development of the human fetal immune system: novel insights from high-dimensional single-cell technologies
Issue Date: 2019-10-08

Development of the human fetal immune system

Novel insights from high-dimensional single-cell technologies

Na Li
李娜

© Na Li, Leiden, the Netherlands
Development of the human fetal immune system: Novel insights from high-dimensional single-cell technologies.

All rights reserved. No part of this thesis may be reproduced or transmitted in any form, by any means, electronic or mechanical without prior written permission of the author, or where appropriate, of the publisher of the articles.

The work presented in this thesis was performed at the Department of Immunohematology and Blood Transfusion of the Leiden University Medical Center in the Netherlands.

ISBN: 978-94-6323-812-0

Printed by: Gildeprint Drukkerijen

Financial support for the publication of this thesis was kindly provided by Fluidigm Corporation.

Development of the human fetal immune system

Novel insights from high-dimensional single-cell technologies

Proefschrift

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,

op gezag van Rector Magnificus prof. mr. C.J.J.M. Stolker,
volgens besluit van het College voor Promoties

te verdedigen op dinsdag 8 oktober 2019
klokke 10.00 uur

door

Na Li

geboren te Tangshan, China
in 1987

Promotor:	Prof. Dr. F. Koning
Co-promotor:	Prof. Dr. Ir. B.P.F.Lelieveldt
Leden Promotiecommisie:	Prof. Dr. F. J.T. Staal
	Prof. Dr. M Yazdanbakhsh
	Prof. Dr. E. Nieuwenhuis (UMC Utrecht)
	Dr. J.N Samsom (Erasmus Medical Center)

For my family

Table of Contents

Chapter 1 General introduction 8

Chapter 2 Mass Cytometry Reveals Innate Lymphoid Cell Differentiation Pathways in the Human Fetal Intestine 22
Journal of Experimental Medicine 215:5, 1383-1396 (2018)

Chapter 3 Memory $CD4^+$ T cells are generated in the human fetal intestine 54
Nature Immunology 20:3, 301-312 (2019)

Chapter 4 Early-life compartmentalization of immune cells in human fetal tissues revealed by high-dimensional mass cytometry 94
Accepted by ***Frontiers in immunology*** (2019)

Chapter 5 Mass Cytometry of the Human Mucosal Immune System Reveals Tissue- and Disease-Associated Immune Subsets 124
Immunity 44:5, 1227-1239 (2016)

Chapter 6 Summarising discussion 158

Chapter 7 Appendices 176
English summary
Nederlandse samenvatting
List of publications
Curriculum vitae
Acknowledgements

Chapter 1

General introduction

The immune system provides protection against pathogens such as bacteria, viruses, parasites and fungi. It comprises a cellular and humoral compartment. The former is a complicated network of immune cells in lymphoid organs, tissues and the circulation and is classically divided into the innate and adaptive immune compartment. Innate immunity provides a crucial first line of defense mediated by a swift, general and non-specific response to the invader. The innate compartment consists of various types of phagocytic granulocytes, myeloid cells, such as macrophages and dendritic cells (DCs), and innate lymphoid cells (ILCs) including cytolytic natural killer (NK) cells and $CD127^+$ ILCs. Simultaneously, the adaptive immune system will mount a more tailored and specific response through antigen-specific receptors-expressing B and T lymphocytes, the latter including helper $CD4^+$ T cells and cytotoxic $CD8^+$ T cells. Ultimately, the combined innate and adaptive will lead to suppression or elimination of the invading pathogen and the generation of immunological memory that provides long-lasting immunity to the pathogen. Typically, adaptive immune responses will be initiated in the lymph nodes by dendritic cells that have moved into the lymph node upon encounter with and internalization of pathogens in peripheral tissues or fluids, upon which the activated immune cells will leave the lymph node and attack the invading pathogen in the body. Lymphoid organs include primary organs (thymus and bone marrow) and secondary tissues such as spleen and mucosal-associated lymphoid tissue, in which the immune process takes place.

The immune system develops in the fetus during pregnancy. The developing fetus expresses the polymorphic major histocompatibility complex (MHC) molecules, half of which are inherited from the mother and half from the father. This implies that the paternal MHC antigens are potential allogeneic targets for the maternal immune system. Similarly, the non-inherited maternal MHC antigens can potentially be recognized by the fetal immune system. For a successful pregnancy it is thus crucial that both the fetal and maternal immune system remain tolerant to the semi-allogeneic environment, but it is at present unclear how this is achieved. On the other hand, the fetus has to be prepared for the massive colonization of mucosal and epithelial surfaces with commensals and exposure to pathogens directly upon birth. In this respect, it is intriguing that in utero the fetus is thought to be protected from exposure to foreign antigens. In agreement, the vast majority of cord blood T cells exhibit a naive phenotype. Yet, it is unclear how such a naive immune compartment would be able to adequately deal with the rapid microbial colonization upon birth. Thus, the developing immune system should be poised for environmental attack while remaining tolerant to the semi-allogeneic environment, a seemingly conflicting and impossible task.

Hematopoiesis

In human fetus, the first occurrence of hematopoietic stem cells (HSCs) is in the dorsal aorta of the aorta-gonad-mesonephros (AGM) region at week 5 of gestation and before their appearance in the yolk sac[1]. The HSCs then migrate into the fetal liver, which becomes the main hematopoietic site at the 5-6 weeks of gestation. Hematopoiesis in the fetal liver reaches the peak in the second trimester and terminates soon upon birth. In the second trimester, the HSCs are mainly produced by the bone marrow, which remains the primary organ of hematopoiesis throughout whole life. The HSCs from fetal liver are the first to colonize the thymus and spleen[2, 3, 4].

Development of the fetal immune system

T cells

The early prothymocytes can be detected in the fetal liver at week 7 of gestation and are characterized by the expression of CD45, CD7 and cytoplasmic CD3 but lacking expression of membrane CD3, TCR β chain and TdT (terminal deoxynucleotidyl transferase) a DNA polymerase, which can insert random N-nucleotides into the DJ region of the T-cell receptor (TCR), contributing to the diversity of TCR repertoire[5]. T cells first appear in the perithymic mesenchyme at around week 7 of gestation and become mature between week 7 to 12 of gestation as indicated by the expression of P80, a marker for mature thymocytes[6]. These cells begin to migrate to the periphery at approximately week 14 of gestation[7]. The expression of TdT enzyme is not readily detectable before week 19 of gestation in the thymus, however, TdT mRNA can be amplified in the thymus as early as week 8 of gestation at a minimal quantity[8].

Compared with a predominant memory T cells ($CD45RO^+$) in the adult tissue, the fetal cord blood T cells mainly display a naive phenotype ($CD45RA^+$) at birth[9]. However, a recent study has identified a minor population of effector memory T cells in cord blood, which expresses the tissue resident and activation marker CD69 and displays both Th1- and Th2-type functional profiles in the absence of known infections during pregnancy[10]. In addition, Byrne et al. have provided evidence for the existence of abundant $CD45RO^+$ memory T cells (including both $CD4^+$ and $CD8^+$ T populations) in the human fetal spleen from 14-22 weeks of gestation, which were able to proliferate in response to IL-2[11]. Between week 16 and 18 of gestation the frequencies of these cells can reach levels comparable to those in the adult spleen but they are dramatically decreased in the neonatal spleen[11]. In

addition, memory T cells have also been identified in the human fetal skin[12] and intestine[13]. Although scarce, T cells are first detectable in the epithelial and lamina propria of the human fetal intestine from week 12 of gestation onwards, and the number of T cells increases along with the gestational age[14]. Bunders et al. demonstrated that in the human fetal mucosa, there is a dominance of $CD45RO^+$ memory T cells with Th1- and Th17-phenotype. These cells express the activation marker HLA-DR and are polyclonal and highly susceptible to HIV infection[13]. The observations of memory T cells in fetal tissues suggest the occurrence of complete T cell maturation in utero but it is unclear how this is induced or regulated. Interestingly, almost 50% of $CD8^+$ T cells in the fetal intestinal lamina propria express CD8aa homodimers, as opposed to CD8αβ heterodimers, commonly found on peripheral $CD8^+$ T cells, which indicates that these CD8aa cells may be thymus-independent and develop in the mucosal tissue niche[15]. As it has been well documented in human studies that from 18 to 24 weeks of gestation, the fetal mesenteric lymph nodes (mLNs) are dominated by $CD45RA^+$ T cells, which can proliferate upon addition of exogenous of IL-2 but not in response to PHA or CD3 specific monoclonal antibodies[16], fetal intestinal memory T cells exhibit unique properties. Finally, mucosal associated invariant T (MAIT) cells have also been identified in the human fetal intestine at the second trimester but their functional properties remain to be further explored[17].

Thus, many issues regarding the nature and development of T cells in the fetal intestine remain unresolved.

Regulatory T cells

Regulatory T cells (Tregs), characterized as $CD3^+CD4^+CD25^{high}FoxP3^+$, play a crucial role in modulating the immune system to maintain immune tolerance and prevent autoimmunity. They are classified into two types: "natural" Tregs, which differentiate from T cell precursors in the thymus[18] and "induced" Tregs, which are generated from the conventional $Foxp3^-CD4^+$ T cells outsize the thymus[19]. It has been reported that Tregs are more abundant in the developing fetus compared with newborns and adult[9, 20] and are present in the human fetal thymus, spleen, lymph nodes (LNs) during the second trimester[20, 21] as well as in cord blood[20]. Moreover, Mold et al. reported that after stimulation with foreign APCs in vitro, human fetal $CD4^+$ T cells derived from fetal LNs and spleens preferentially gave rise to Tregs, which were specific for the non-inherited maternal antigens (NIMAs)[22]. Moreover, HSCs derived from fetal liver and fetal bone marrow were found to generate a significant higher frequency of $Foxp3^+$ Tregs compared to adult derived bone marrow HSCs, indicating that fetal HSCs are prone to Treg differentiation[23].

Finally, the absence of Foxp3$^+$ Tregs in utero has been linked with diseases like type 1 diabetes and dermatitis[24], which further highlights the value to study the Tregs in fetal tissues in detail.

Innate lymphoid cells

Innate lymphoid cells (ILCs) lack the expression of antigen-specific receptors and the major immune lineage markers but otherwise they are the innate counterparts of T cells. NK cells, group 1 ILCs (ILC1s), group 2 ILCs (ILC2s) and group 3 ILCs (ILC3s) mirror CD8$^+$ cytotoxic T cells, CD4$^+$ Th1, Th2 and Th17 cells, respectively, with respect to their function and development[25]. It has been reported that fetal ILCs are found in the human liver, lymphoid tissue, lung and intestine[26, 27, 28]. In the fetal liver, NKp44$^-$ ILC3s are abundant and appear as early as week 6 of gestation whereas the other subsets are detectable only after week 15[26]. Compared to fetal liver, NKp44$^+$ ILC3s are enriched in the human fetal intestine, where there are hardly any ILC1s, suggesting that the differentiation of ILC1s may be related to microbiota colonization[29]. Lymphoid tissue inducer (LTi) cells, a subtype of ILC3s expressing Neuropilin-1[27], play a crucial role in second lymphoid organs formation during the fetal development, and have been identified in the fetal LNs and spleen in the first and second trimester[27, 30, 31]. In addition, Mjösberg et al. has provided evidence for the existence of ILC2s in the mucosal tissue such as fetal intestine and lung during the second trimester[32]. Thus, tissue specific localization of particular subsets is already evident in the developing fetus. In recent years, substantial plasticity and heterogeneity within ILC compartment have been described in adults[25, 33] but this is still unexplored in the human fetus.

The knowledge on ILC development is mainly based on studies in mice, where the common lymphoid progenitors (CLPs) give rise to the common innate lymphoid progenitors (CILPs), which further differentiate into NK cell precursors (NKP) and common helper innate lymphoid progenitors (CHILPs). CHILPs give rise to lymphoid tissue inducer progenitors and innate lymphoid cell precursors (ILCPs), which generate LTi and ILC1s, ILC2s, and ILC3s, respectively[25]. In humans, multipotent CLP-like CD34$^+$ progenitor and CD34$^+$CD117$^+$ CILPs have been identified in human lymphoid tissues, such as tonsil[34, 35]. Additionally, a subset of CD34$^-$CD127$^+$CD117$^+$ ILCPs has been described in different tissues, including fetal liver[36]. Recently, Chen et al. demonstrated that human CD56$^+$ ILCPs can give rise to NK cells and ILC3s whereas CD56$^-$ ILCPs generate ILC2s, pointing towards differences in the developmental pathways of ILC in human and mice[37]. Finally, as only the NKP and ILC3 precursor have been identified in humans[38], several other precursors and intermediates of the ILC developmental pathways still need to be

identified.

B cells and dendritic cells

B cells appear in the fetal liver as early as week 8 of gestation[39] and in the spleen by week 13[40] and B cells with a diverse B cell receptor repertoire are detectable in cord blood already at week 12 of gestation[41]. The presence of DCs has been demonstrated in different fetal tissues, such as spleen, skin and intestine, where the DCs suppress the T-cell immunity by promoting the induction of Tregs and inhibiting the production of tumor-necrosis factor (TNF)[42]. In addition, follicular DCs co-localized with B cells in the fetal spleen, indicating antigen presentation to B cells[43]. Although different types of subsets in both innate and adaptive immunity have been identified in utero, most of the knowledge is based on studies with cord blood collected at birth due to scarcity of human fetal tissues. In addition, the recent discoveries of effector T cells and mature DCs in the developing fetus suggest that the fetal immune system is more complex than previously appreciated. Therefore, further investigations into the development of early-life immunity with a global and detailed view across fetal tissues such as the intestine are needed.

Intestinal mucosal immune system

The intestine harbors the largest immune compartment in our body and is a main interface for direct interaction between the immune system and a large array of environmental antigens. It is continuously exposed to both antigens derived from harmless dietary constituents and potential harmful (commensal) bacteria, separated from each other by only a single layer of epithelial cells. For this, an elaborate intestinal immune system has evolved to distinguish the harmful from beneficial antigens. T cells are scattered throughout the epithelium and the underlying lamina propria harbors both adaptive and innate immune cells as well as a variety of antigen presenting cells. In addition, there are gut-associated lymphoid tissues, such as mesenteric lymph nodes and Peyer's patches[44], the latter of which are first detectable in the fetal intestine between week 16-19 of gestation[45]. While there is extensive literature on the organization and composition of the immune system in the intestine in children and adults, relatively little is known about this issue in the developing fetal intestine. In the current study, we have approached this issue by making use of recently introduced single cell technologies, including (imaging) mass cytometry and single-cell RNA-sequencing.

Mass cytometry

Mass cytometry or cytometry by time-of-flight (CyTOF) is a new platform that couples flow cytometry with mass spectrometry. This technique uses rare earth metals as reporters instead of fluorophores, which removes the limitation on the number of antibodies that can simultaneously be used due to the spectral overlap of the fluorochromes, and eliminates the problem of autofluorescence. Through mass cytometry up to 42 markers can now be used simultaneously at single-cell resolution with minimal overlap between channels[46, 47]. Theoretically, with the availability of additional metal reporters, over 100 markers can be measured simultaneously in the future. The schematic overview of the mass cytometry workflow is described in **Figure 1**. Therefore, mass cytometry now provides the opportunity to study the diversity and heterogeneity of the immune system with an unprecedent high-resolution. Although mass cytometry has many advantages over traditional flow cytometry, there are still some challenges with mass cytometry[46]. First, the sensitivity of the metal reporters used in mass cytometry is lower than the brightest fluorophores in flow cytometry. Second, the acquisition speed can reach thousands of events per second with around 95% recovery in flow cytometry, compared to only 300 events per second with around 60% recovery (Helios) in mass cytometry. Third, as the cells are destroyed during the data acquisition, mass cytometry cannot be used to sort cells for further functional analysis. Finally, in mass cytometry the classical forward and size scatter parameters are lacking.

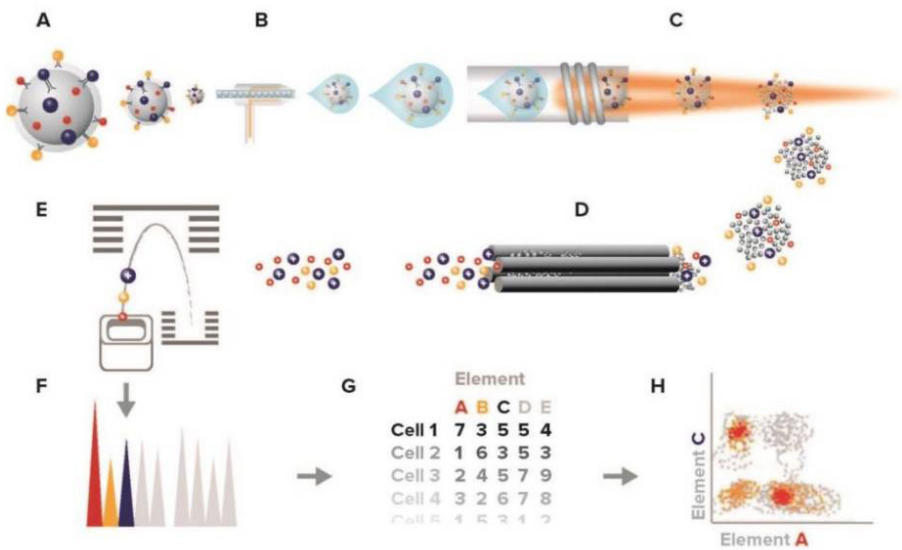


Figure 1.Workflow of CyTOF. A single-cell suspension is labeled with metal-conjugated antibodies (A) and injected into the nebulizer (B). After aerosolizing into a single-cell droplet, it is directed into the inductively coupled plasma (ICP) torch, in which it is vaporized, atomized and ionized (C). The low mass ions are removed by the high pass optic (D), resulting in an ion cloud enriched with heavy metal isotypes, which then enters the time of flight (TOF) chamber (E). The ions are then separated based on the their mass as they accelerate to the detector, where the intensity of each metal for each individual cells will be measured (F). Finally, data is generated in the FSC format (G) and analyzed using third-party software (H). Adapted from Fluidigm.

Data analysis in mass cytometry

In flow cytometry, cell populations are manually identified by using sequential gating based on bimodal expression patterns within one- or two-dimensional plots. With the substantial increase in the number of parameters in mass cytometry, the complexity of the data interpretation has challenged this manual "gating strategy" and inspired the development of new computational tools. t-distributed Stochastic Neighbor Embedding (t-SNE) is a non-linear dimensionality reduction algorithm and currently widely used for the analysis of mass cytometry data. t-SNE takes all the marker expression patterns into account simultaneously and projects the high dimensional data onto two dimensional plots[48]. Nevertheless, conventional t-SNE has several limitations like a restriction of the size of the dataset and long computing time. Hierarchical Stochastic Neighbor Embedding (HSNE), an in-house developed software, removes the scalability limitation of the conventional t-SNE analysis and allows the rapid exploration of millions of cells simultaneously. Furthermore, the implementation of HSNE in Cytosplore[49], named Cytosplore^{+HSNE} [50], allows the user to analyze the entire dataset at several levels of hierarchy, going from a global overview of the data structure to full single cell resolution.

Imaging-mass cytometry

Recently, imaging-mass cytometry has been developed as a new technology to gain spatial information in situ, which couples a laser ablation system with a mass cytometer. In this way, it enables the visualization of dozens of markers on the same tissue section simultaneously with a sub-cellular resolution of 1 μm[51]. A schematic overview of the imaging-mass cytometry workflow is described in **Figure 2**. In addition, the traditional imaging analysis can be performed at the cell level by applying cell segmentation, cell neighborhood analysis and distance measurement. Up till now, imaging-mass cytometry has been mainly applied for biomarker identification in pharmaceutical research[52] and analysis of the tumor environment in cancer research[51, 53, 54]. However, imaging-mass cytometry can be of great value in many other studies as well as it offers an opportunity to gain unprecedented insight into the organization of the immune system in situ.

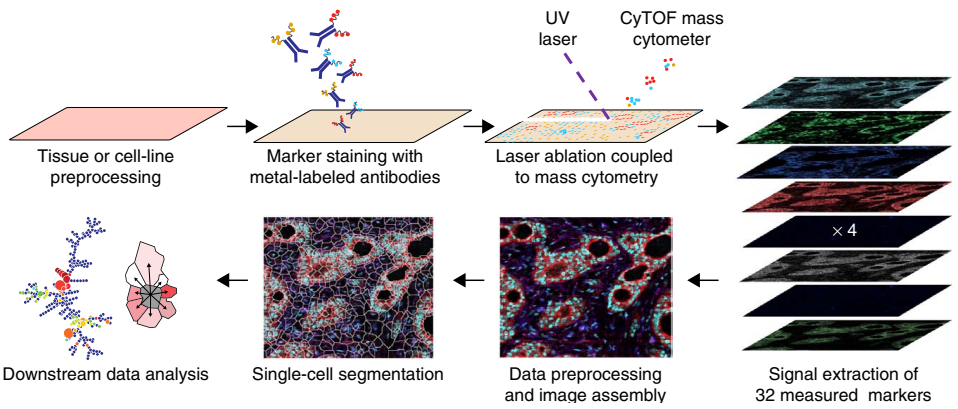


Figure 2. Work flow of imaging-mass cytometry. Adapted from Giesen et al. (2014)

Single-cell RNA-sequencing

As a complementary approach to mass cytometry, single-cell RNA sequencing dissects the gene expression profiles on tens of thousands of individual cells at the single-cell level. In recent years, a varieties of single-cell RNA-sequencing techniques have been developed such as microfluidic-, plate- and droplet-based approaches[55]. As each of them has its own advantages and drawbacks[56], the users can choose the best approach based on their research goals. As such, single-cell RNA-sequencing provides an opportunity to determine the cell transcriptomic heterogeneity, allows the discovery of rare cell populations, can be used to determine cellular differentiation trajectories and to reveal regulatory networks among cells at the gene level[57].

Outline of the thesis

In recent years, several studies have highlighted the uniqueness of the human immune system in early life. Due to the scarceness of human fetal tissues and technical limitations, a system-wide and detailed phenotypical characterization of the composition and development of the human fetal immune system was lacking. In this thesis, I describe my work to gain further insight in the composition and development of the human fetal immune system using an array of advanced high-throughput technologies.

The immune system comprises both the innate and adaptive immune compartments. With respect to the innate compartment, I was the first to apply mass cytometry to delineate the innate lymphoid cells (ILCs) in the human fetal intestine (**Chapter 2**). This revealed extensive heterogeneity of ILCs. Beside

the known ILC subsets, a previously unrecognized subset was identified, named int-ILCs, with a cell surface phenotype that combined CD127$^+$ ILC and NK cell properties. In line with this, visualization of the t-SNE computation demonstrated that this cell positioned itself in between the CD127$^+$ ILC and NK clusters, giving rise to the hypothesis that these cells might be precursors that could give rise to both NK cells and CD127$^+$ ILCs. In functional in vitro studies I demonstrated that this was indeed the case. This study was one of the first to generate a data-driven hypothesis by exploring the heterogeneity of the immune system using mass cytometry, followed by functional validation.

With respect to the adaptive immune compartment, I used an array of advanced single-cell techniques, including high-throughput mass cytometry and imaging-mass cytometry, single-cell RNA-sequencing and TCR sequencing, to characterize the CD4$^+$ T cell compartment in the human fetal intestine (**Chapter 3**). By combining the acquired datasets with advanced computational analysis tools and functional analysis this revealed that memory-like CD4$^+$ T cells with pro-inflammatory properties were already generated in the developing human fetal intestinal immune system, indicative of in utero exposure to foreign antigens.

To further extend the understating of the fetal immune system, I applied mass cytometry and imaging-mass cytometry to both the innate and adaptive immune compartments in the fetal intestine, spleen and liver (**Chapter 4**). This revealed an early-life immune compartmentalization in these different fetal tissues.

Also, I was involved in a study in which we analyzed the composition of the mucosal immune system of patients with inflammatory intestinal diseases and controls using mass cytometry (**Chapter 5**). This first in its kind study revealed previously unrecognized heterogeneity in the mucosal immune system and tissue- and disease-specific immune subsets.

Finally, in **chapter 6**, I discuss the major findings described in **Chapter 2-5** and their implications for shaping the prenatal immune system. In addition, potential directions for future research are discussed.

References

1. Ivanovs, A. et al. Highly potent human hematopoietic stem cells first emerge in the intraembryonic aorta-gonad-mesonephros region. J Exp Med 208, 2417-2427 (2011).
2. Christensen, R.D. Hematopoiesis in the fetus and neonate. Pediatr Res 26, 531-535 (1989).
3. Migliaccio, G. et al. Human embryonic hemopoiesis. Kinetics of progenitors and precursors underlying the yolk sac----liver transition. J Clin Invest 78, 51-60 (1986).
4. Zhang, X., Zhivaki, D. & Lo-Man, R. Unique aspects of the perinatal immune system. Nat Rev Immunol 17, 495-507 (2017).
5. Haynes, B.F., Denning, S.M., Singer, K.H. & Kurtzberg, J. Ontogeny of T-cell precursors: a model for the initial stages of human T-cell development. Immunol Today 10, 87-91 (1989).
6. Lobach, D.F., Hensley, L.L., Ho, W. & Haynes, B.F. Human T-Cell Antigen Expression during the Early Stages of Fetal Thymic Maturation. Journal of Immunology 135, 1752-1759 (1985).
7. Haynes, B.F. The Human Thymic Microenvironment. Advances in Immunology 36, 87-142 (1984).
8. George, J.F., Jr. & Schroeder, H.W., Jr. Developmental regulation of D beta reading frame and junctional diversity in T cell receptor-beta transcripts from human thymus. J Immunol 148, 1230-1239 (1992).
9. Thome, J.J. et al. Early-life compartmentalization of human T cell differentiation and regulatory function in mucosal and lymphoid tissues. Nat Med 22, 72-77 (2016).
10. 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, 238ra272 (2014).
11. Byrne, J.A., Stankovic, A.K. & Cooper, M.D. A novel subpopulation of primed T cells in the human fetus. J Immunol 152, 3098-3106 (1994).
12. Schuster, C. et al. Phenotypic characterization of leukocytes in prenatal human dermis. J Invest Dermatol 132, 2581-2592 (2012).
13. 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).
14. Spencer, J., MacDonald, T.T., Finn, T. & Isaacson, P.G. The development of gut associated lymphoid tissue in the terminal ileum of fetal human intestine. Clin Exp Immunol 64, 536-543 (1986).
15. Latthe, M., Terry, L. & MacDonald, T.T. High frequency of CD8 alpha alpha homodimer-bearing T cells in human fetal intestine. Eur J Immunol 24, 1703-1705 (1994).
16. von Hoegen, P., Sarin, S. & Krowka, J.F. Deficiency in T cell responses of human fetal lymph node cells: a lack of accessory cells. Immunol Cell Biol 73, 353-361 (1995).
17. Leeansyah, E., Loh, L., Nixon, D.F. & Sandberg, J.K. Acquisition of innate-like microbial reactivity in mucosal tissues during human fetal MAIT-cell development. Nat Commun 5, 3143 (2014).
18. Geiger, T.L. & Tauro, S. Nature and nurture in Foxp3(+) regulatory T cell development, stability, and function. Hum Immunol 73, 232-239 (2012).
19. Chen, W. et al. Conversion of peripheral CD4+CD25- naive T cells to CD4+CD25+ regulatory T cells by TGF-beta induction of transcription factor Foxp3. J Exp Med 198, 1875-1886 (2003).
20. Michaelsson, J., Mold, J.E., McCune, J.M. & Nixon, D.F. Regulation of T cell responses in the developing human fetus. J Immunol 176, 5741-5748 (2006).
21. Cupedo, T., Nagasawa, M., Weijer, K., Blom, B. & Spits, H. Development and activation of regulatory T cells in the human fetus. Eur J Immunol 35, 383-390 (2005).
22. Mold, J.E. et al. Maternal alloantigens promote the development of tolerogenic fetal regulatory T cells in utero. Science 322, 1562-1565 (2008).
23. Mold, J.E. et al. Fetal and adult hematopoietic stem cells give rise to distinct T cell lineages in humans. Science 330, 1695-1699 (2010).
24. Bennett, C.L. et al. The immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX) is caused by mutations of FOXP3. Nature Genetics 27, 20-21 (2001).
25. Vivier, E. et al. Innate Lymphoid Cells: 10 Years On. Cell 174, 1054-1066 (2018).

26. Forkel, M. et al. Composition and functionality of the intrahepatic innate lymphoid cell-compartment in human nonfibrotic and fibrotic livers. Eur J Immunol 47, 1280-1294 (2017).

27. Shikhagaie, M.M. et al. Neuropilin-1 Is Expressed on Lymphoid Tissue Residing LTi-like Group 3 Innate Lymphoid Cells and Associated with Ectopic Lymphoid Aggregates. Cell Reports 18, 1761-1773 (2017).

28. Miller, D., Motomura, K., Garcia-Flores, V., Romero, R. & Gomez-Lopez, N. Innate Lymphoid Cells in the Maternal and Fetal Compartments. Front Immunol 9, 2396 (2018).

29. Hazenberg, M.D. & Spits, H. Human innate lymphoid cells. Blood 124, 700-709 (2014).

30. Hoorweg, K. et al. A Stromal Cell Niche for Human and Mouse Type 3 Innate Lymphoid Cells. J Immunol 195, 4257-4263 (2015).

31. Cupedo, T. et al. Human fetal lymphoid tissue-inducer cells are interleukin 17-producing precursors to RORC+ CD127+ natural killer-like cells. Nat Immunol 10, 66-74 (2009).

32. Mjosberg, J.M. et al. Human IL-25- and IL-33-responsive type 2 innate lymphoid cells are defined by expression of CRTH2 and CD161. Nat Immunol 12, 1055-1062 (2011).

33. Bjorklund, A.K. et al. The heterogeneity of human CD127(+) innate lymphoid cells revealed by single-cell RNA sequencing. Nat Immunol 17, 451-460 (2016).

34. Freud, A.G. et al. Evidence for discrete stages of human natural killer cell differentiation in vivo. J Exp Med 203, 1033-1043 (2006).

35. Scoville, S.D. et al. A Progenitor Cell Expressing Transcription Factor RORgammat Generates All Human Innate Lymphoid Cell Subsets. Immunity 44, 1140-1150 (2016).

36. Lim, A.I. et al. Systemic Human ILC Precursors Provide a Substrate for Tissue ILC Differentiation. Cell 168, 1086-1100 e1010 (2017).

37. Chen, L. et al. CD56 Expression Marks Human Group 2 Innate Lymphoid Cell Divergence from a Shared NK Cell and Group 3 Innate Lymphoid Cell Developmental Pathway. Immunity 49, 464-476 e464 (2018).

38. Juelke, K. & Romagnani, C. Differentiation of human innate lymphoid cells (ILCs). Curr Opin Immunol 38, 75-85 (2016).

39. Thilaganathan, B., Nicolaides, K.H., Mansur, C.A., Levinsky, R.J. & Morgan, G. Fetal B lymphocyte subpopulations in normal pregnancies. Fetal Diagn Ther 8, 15-21 (1993).

40. Namikawa, R. et al. Ontogenic development of T and B cells and non-lymphoid cells in the white pulp of human spleen. Immunology 57, 61-69 (1986).

41. Rechavi, E. et al. Timely and spatially regulated maturation of B and T cell repertoire during human fetal development. Sci Transl Med 7, 276ra225 (2015).

42. McGovern, N. et al. Human fetal dendritic cells promote prenatal T-cell immune suppression through arginase-2. Nature 546, 662-+ (2017).

43. Timens, W., Rozeboom, T. & Poppema, S. Fetal and neonatal development of human spleen: an immunohistological study. Immunology 60, 603-609 (1987).

44. Mowat, A.M. & Agace, W.W. Regional specialization within the intestinal immune system. Nat Rev Immunol 14, 667-685 (2014).

45. Howie, D. et al. Extrathymic T cell differentiation in the human intestine early in life. J Immunol 161, 5862-5872 (1998).

46. Stern, L. et al. Mass Cytometry for the Assessment of Immune Reconstitution After Hematopoietic Stem Cell Transplantation. Front Immunol 9, 1672 (2018).

47. Bendall, S.C., Nolan, G.P., Roederer, M. & Chattopadhyay, P.K. A deep profiler's guide to cytometry. Trends Immunol 33, 323-332 (2012).

48. van der Maaten, L. & Hinton, G. Visualizing Data using t-SNE. Journal of Machine Learning Research 9, 2579-2605 (2008).

49. Hollt, T. et al. Cytosplore: Interactive Immune Cell Phenotyping for Large Single-Cell Datasets. Computer Graphics Forum 35, 171-180 (2016).

50. 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).

51. Giesen, C. et al. Highly multiplexed imaging of tumor tissues with subcellular resolution by

mass cytometry. Nature Methods 11, 417-+ (2014).

52. Chang, Q. et al. Biodistribution of cisplatin revealed by imaging mass cytometry identifies extensive collagen binding in tumor and normal tissues. Sci Rep 6, 36641 (2016).

53. Singh, M. et al. Highly Multiplexed Imaging Mass Cytometry Allows Visualization of Tumor and Immune Cell Interactions of the Tumor Microenvironment in FFPE Tissue Sections. Blood 130 (2017).

54. Schulz, D. et al. Simultaneous Multiplexed Imaging of mRNA and Proteins with Subcellular Resolution in Breast Cancer Tissue Samples by Mass Cytometry. Cell Syst 6, 25-36 e25 (2018).

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

56. Chen, G., Ning, B. & Shi, T. Single-Cell RNA-Seq Technologies and Related Computational Data Analysis. Front Genet 10, 317 (2019).

57. Hwang, B., Lee, J.H. & Bang, D. Single-cell RNA sequencing technologies and bioinformatics pipelines. Exp Mol Med 50, 96 (2018).

JEM
Journal of Experimental Medicine
VOL 215 • NO 5 • MAY 2018
WWW.JEM.ORG
Rockefeller University Press

Chapter 2

Mass Cytometry Reveals Innate Lymphoid Cell Differentiation Pathways in the Human Fetal Intestine

Na Li,[1,*] Vincent van Unen,[1,*] Thomas Höllt,[2,3] Allan Thompson,[1]
Jeroen van Bergen,[1] Nicola Pezzotti,[2] Elmar Eisemann,[2] Anna Vilanova,[2]
Susana M. Chuva de Sousa Lopes,[4] Boudewijn P.F. Lelieveldt,[5,6] Frits Koning[1]

[1] Department of Immunohematology and Blood Transfusion,
Leiden University Medical Center, Leiden, Netherlands

[2]Computer Graphics and Visualization Group,
Delft University of Technology, Delft, Netherlands

[3]Computational Biology Center,
Leiden University Medical Center, Leiden, Netherlands

[4]Department of Anatomy and Embryology,
Leiden University Medical Center, Leiden, Netherlands

[5]Department of LKEB Radiology,
Leiden University Medical Center, Leiden, Netherlands

[6] Department of Pattern Recognition and Bioinformatics group,
Delft University of Technology, Delft, Netherlands

* Equal contribution

Journal of Experimental Medicine 215:5, 1383-1396 (2018)

ON THE COVER

Li et al. apply mass cytometry to delineate the fetal gut innate lymphoid cell (ILC) population and use a t-SNE-based approach to predict potential differentiation trajectories. This image represents the composition of the ILC compartment in the individual fetal intestines. The image was taken from the original manuscript and modified by the *JEM* editorial office. See page 77

Abstract

Innate lymphoid cells (ILCs) are abundant in mucosal tissues and involved in tissue homeostasis and barrier function. While several ILC subsets have been identified, it is unknown if additional heterogeneity exists and their differentiation pathways remain largely unclear. We applied mass cytometry to analyze ILCs in the human fetal intestine and distinguished 34 distinct clusters through a t-SNE-based analysis. A lineage (Lin)$^-$CD7$^+$CD127$^-$CD45RO$^+$CD56$^+$ population clustered between the CD127$^+$ ILC and natural killer (NK) cell subsets, and expressed diverse levels of Eomes, T-bet, GATA3 and RORγt. By visualizing the dynamics of the t-SNE computation, we identified smooth phenotypic transitions from cells within the Lin$^-$CD7$^+$CD127$^-$CD45RO$^+$CD56$^+$ cluster to both the NK cells and CD127$^+$ ILCs, revealing potential differentiation trajectories. In functional differentiation assays the Lin$^-$CD7$^+$CD127$^-$CD45RO$^+$CD56$^+$CD8a$^-$ cells could develop into CD45RA$^+$ NK cells and CD127$^+$RORγt$^+$ ILC3-like cells. Thus, we identified a previously unknown intermediate innate subset that can differentiate into ILC3 and NK cells.

Introduction

Innate lymphoid cells (ILCs) lack expression of T cell receptors but otherwise are a functional counterpart of cytotoxic and helper T cell subsets. Helper ILCs are classified into 3 groups: ILC1, ILC2 and ILC3 [1]. ILC1s are mainly characterized as Lineage (Lin)$^-$CD161$^+$CD127$^+$CRTH2$^-$CD117$^-$, express the transcription factor T-bet and produce T helper 1 (T_H1) cell-associated cytokines. ILC2s are Lin$^-$CD161$^+$CD127$^+$CRTH2$^+$, express GATA3, and produce T helper 2 (T_H2) cell-associated cytokines. ILC3s, including fetal lymphoid tissue-inducer (LTi) cells, are Lin$^-$CD161$^+$CD127$^+$CRTH2$^-$CD117$^+$, RORγt$^+$, and secrete T_H17/T_H22 helper T cell-associated cytokines [1,2]. A fraction of human ILC3s expresses natural cytotoxicity receptors such as NKp44, NKp46 and NKp30, and neural cell adhesion molecule CD56, similar to natural killer (NK) cells [3,4]. NK cells are a cytotoxic subset of ILCs that express the transcription factor T-bet and/or Eomes and produce IFN-γ, granzymes and perforin [1]. Also, ILCs are most abundant and reside in (mucosal) tissues such as the tonsil, lung and intestine, where they can expand locally [5].

Several studies have reported the differentiation pathways of ILCs in a variety of tissues in both mice and humans [6,7]. For example, in murine fetal liver and adult intestine, a CXCR6$^+$RORγt$^+$α4β7$^+$ subset has been identified that can differentiate into ILC3s and NK cells [8]. As this subset was not found in adult bone marrow, it might migrate to the intestine during fetal development. In humans, RORγt$^+$CD34$^+$ progenitor cells were identified in the tonsil and intestine, but these were absent in peripheral blood, umbilical cord blood, bone marrow and thymus [9,10]. Since these progenitors could differentiate into helper ILCs and NK cells, mucosal organs might be the preferential sites for ILC differentiation. In addition, a CD127$^+$CD117$^+$ ILC precursor (ILCP) has been identified in cord blood, peripheral blood and tissues, including fetal liver, adult lung and tonsil, which can generate all ILC subsets in situ and could represent an intermediate between precursor cells and mature ILCs [11]. Also, previous studies have observed ILC plasticity mainly in mucosal tissues, such as the small intestine [12–15], suggesting that environmental cues may play an important role in cell-fate decision. So far, most of the studies on human ILC differentiation used CD34$^+$ progenitors and mature types of ILCs [6], while the intermediates or transitional stages connecting the CD34$^+$ populations to mature types of ILCs have not been fully identified.

High-dimensional mass cytometry provides an opportunity to analyze the heterogeneity and potential differentiation pathways of human ILCs in an unbiased and data-driven fashion based on the simultaneous measurement of over 30 cellular markers at single-cell resolution [16]. Although the sensitivity of

metal reporters in mass cytometry is not as sensitive as some of the brightest fluorochromes in flow cytometry, the advantage of including many more markers in a single antibody panel offers unique opportunities to evaluate the composition of the immune system with unprecedented resolution. Up to recently, analysis of flow cytometry data was mainly performed with gating strategies based on (primarily) bimodal expression patterns. The incorporation of over 30 markers in mass cytometry antibody panels is not well compatible with such an analysis approach. Instead, t-Distributed Stochastic Neighbor Embedding (t-SNE)-based approaches are currently becoming the standard in the field as they allow the simultaneous analysis of all marker expression profiles in an unbiased fashion. Hierarchical SNE, for example, allows efficient analysis of mass cytometry data sets on tens of millions of cells at the single-cell level [17]. Here, we applied mass cytometry to analyze the ILC compartment in the human fetal intestine and provide evidence for previously unrecognized heterogeneity within this compartment. Moreover, we utilized a t-SNE-based computational approach to predict potential differentiation trajectories in silico, and provide evidence for the existence of a previously unrecognized innate cell subset that can differentiate into both NK cells and ILC3 in vitro.

Results

High-dimensional analysis reveals previously unrecognized heterogeneity in the ILC compartment

We developed a 35 metal isotope-tagged monoclonal antibody panel (**Table S1**) to identify the 6 major immune lineages (B cells, myeloid cells, $CD4^+$, $CD8^+$, γδ T cells, and Lin^-CD7^+ cells; the latter hereafter referred to as ILCs) and heterogeneity within those lineages. For this purpose the panel included lineage markers and markers linked to cell differentiation, activation, trafficking and responsiveness to humoral factors. With this panel, single-cell suspensions prepared from 7 fetal intestines were analyzed individually. Single, live $CD45^+$ cells were discriminated by event length, DNA stainings and CD45 antibody staining (**Fig. S1 A**). All antibodies showed clear discrimination between antibody-positive and -negative cells (**Fig. S1 B**). Similar to our previous study [18], we applied a combined t-SNE [19]-ACCENSE [20] data analysis approach to the 6 major cell lineages (**Fig. S1 C**) which revealed a large degree of heterogeneity within these lineages.

We next focused on the ILC compartment (**Fig. S2 A**) which comprised 20.4% ± 7.8% of the $CD45^+$ cells. We pooled the data from the 7 samples and performed a

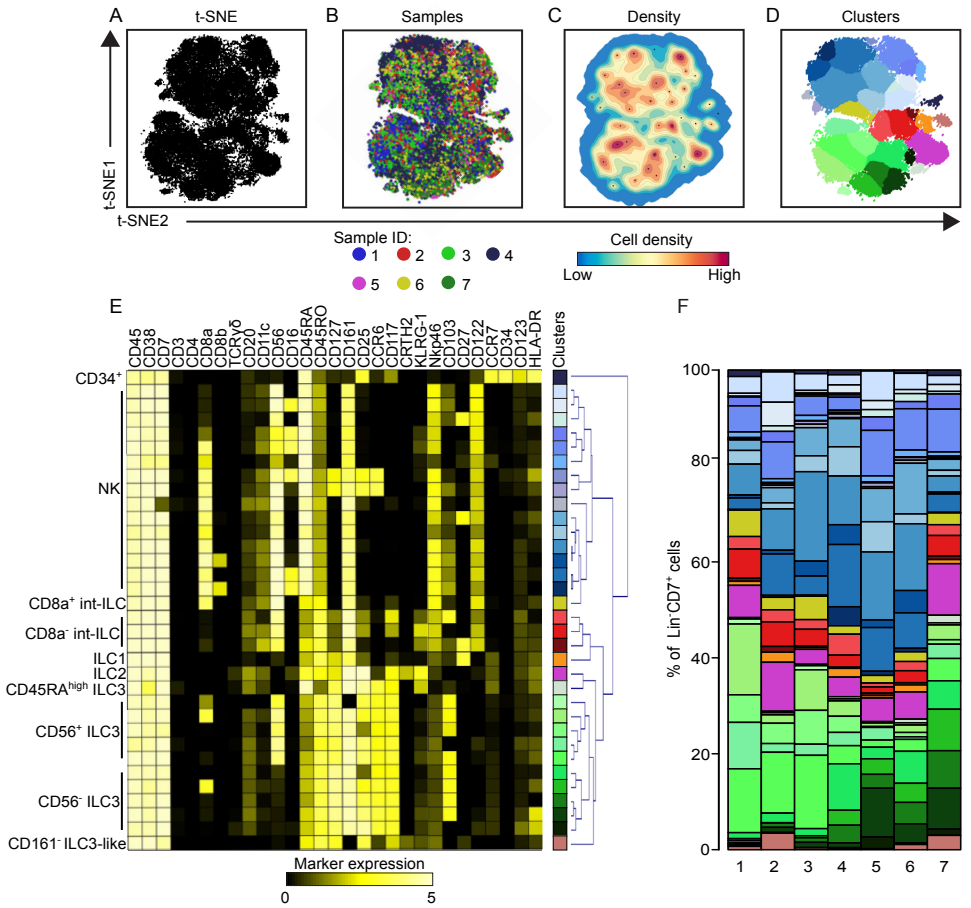
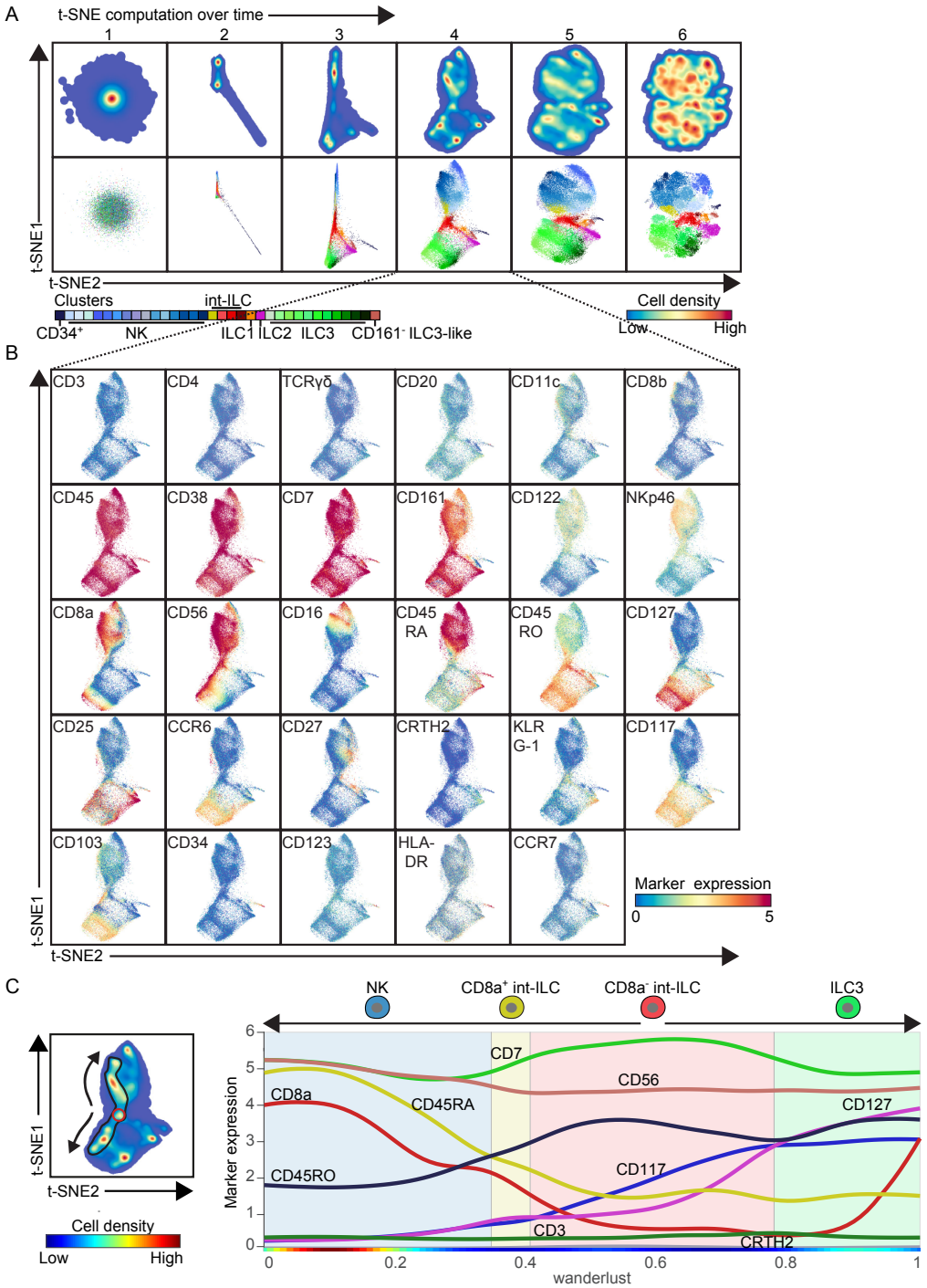


Fig. 1. High-dimensional mass cytometry analysis reveals previously unrecognized heterogeneity in the fetal intestinal ILC compartment (**A**) A t-SNE embedding showing the collective ILCs ($4.5x10^4$ cells) derived from 7 fetal intestines, each black dot represents a single cell. (**B**) A t-SNE embedding as shown in panel A. Colors represent the different samples. (**C**) A density map describing the local probability density of t-SNE-embedded cells where black dots represent centroids of identified clusters using Gaussian Mean Shift clustering. (**D**) A t-SNE plot showing cluster partitions in different colors. (**E**) Heatmap showing the median ArcSinh5-transformed marker expression values (black-to-yellow scale) of the clusters identified in panel A with description of ten annotated categories based on pre-existing literature, and hierarchical clustering thereof. (**F**) Composition of the ILC compartment in the individual fetal intestines (N=7) represented in vertical bars where the size of the colored segments represents the proportion of cells as % of total ILC in the sample. Colors as in panel A.

t-SNE analysis in Cytosplore [21]. This provided a two-dimensional map where cellsare positioned based on the similarity in expression of all marker simultaneously (**Fig. 1 A and B**). Based on the density features of the t-SNE-embedded cells, we identified 34 phenotypically distinct clusters (**Fig. 1, C and D**) using the Gaussian Mean Shift clustering and generated a heatmap showing the distinct marker


A
t-SNE computation over time
1
2
3
4
5
6
t-SNE1
t-SNE2
Clusters
int-ILC
Cell density
CD34+
NK
ILC1
ILC2
ILC3
CD161- ILC3-like
Low
High
B
CD3
CD4
TCRγδ
CD20
CD11c
CD8b
CD45
CD38
CD7
CD161
CD122
NKp46
CD8a
CD56
CD16
CD45 RA
CD45 RO
CD127
CD25
CCR6
CD27
CRTH2
KLR G-1
CD117
CD103
CD34
CD123
HLA-DR
CCR7
Marker expression
0
5
t-SNE1
t-SNE2
C
NK
CD8a+ int-ILC
CD8a- int-ILC
ILC3
t-SNE1
t-SNE2
Cell density
Low
High
Marker expression
CD8a
CD45RO
CD45RA
CD7
CD56
CD127
CD117
CD3
CRTH2
wanderlust
0
0.2
0.4
0.6
0.8
1

expression profiles for each cluster (**Fig. 1 E**). Unbiased hierarchical clustering revealed distinct clusters including a group of $CD34^+$ cells expressing CD45RA and CD117, a larger cluster of several types of NK cells, and a $CD127^+$ ILC cluster with cells expressing markers corresponding to ILC1, ILC2, $CD45RA^{high}$ ILC3, subsets of several types of $CD56^+$ and $CD56^-$ ILC3 [2,22] and a $CD161^-$ ILC3-like population [23]. In addition, several unrecognized cell clusters with a $Lin^-CD7^+CD127^-CD45RO^+CD56^+$ phenotype [referred to as intermediate ILC (int-ILC) hereafter] were identified which clustered between the NK cells and $CD127^+$ ILCs (**Fig. 1, C-E**). While the majority of these int-ILCs (6.6% of ILCs ± 2.3%) were $CD8a^-$, a smaller related population (3.0% ± 1.6%) was $CD8a^+$ (**Fig. 1 F**). Importantly, analysis of the composition of the cluster frequencies in the individual fetal samples demonstrated that even though quantitative differences exist, most of the identified clusters, including the int-ILCs were present in all 7 samples (**Fig. 1 F**).

Together, these data indicate that all known NK and $CD127^+$ ILC cell clusters could be identified simultaneously while evidence for the existence of previously unrecognized clusters was obtained as well.

Visualization of the t-SNE computation dynamics predicts potential differentiation trajectories in the ILC compartment

The cell surface phenotype of int-ILC (i.e. $CD127^-CD45RO^+$) places them in between the $CD127^+$ ILCs and the NK cells (**Fig. 1, C-E**), suggesting potential relationships with both. To investigate this in more detail, we sought to visualize potential relationships between cell populations without prior designation of a user-defined starting cell type in silico. To this end, we exploited the ability of Cytosplore to visualize the evolution of the t-SNE map [21]. Separating the computational modelling into 6 stages revealed how distinct cell clusters were formed, while their high-dimensional similarities were projected onto a two-dimensional map,

Fig. 2. Monitoring t-SNE computation dynamics predicts potential differentiation trajectories of ILCs (**A**) t-SNE embeddings of the collective ILC single-cell data derived from 7 fetal intestines showing density features (upper row) and single cells (bottom row) at 6 stages over the course of the t-SNE computation. Colors represent the local probability density of t-SNE-embedded cells (upper row), or cluster partitions (bottom row), as described in Fig. 1 A. (**B**) t-SNE embeddings at stage 4 of the optimization phase as described in panel A. Colors of the cells represent ArcSinh5-transformed expression values of indicated markers. (**C**) Left panel: t-SNE embedding at stage 4 as in panel A. Colors represent density features and black encirclement indicates the trajectory of cells along the CD56 expression continuum shown in panel B. Right panel: Wanderlust graph (trajectory 0-1.0) of the CD56 positive cells in the left panel showing median ArcSinh5-transformed expression of CD8a, CD45RO, CD45RA, CD7, CD3, CD56, CD117, CD127 and CRTH2 from the $CD8a^-$ int-ILCs (shaded red box) via the $CD8a^+$ int-ILC (shaded yellow box) to NK cells (shaded blue box), and from the $CD8a^-$ int-ILCs to ILC3s (shaded green box). The rainbow color bar indicates relative cell density.

and linked to each other based on marker expression profiles (**Fig. 2 A**). Since the initial positions in the t-SNE map are assigned randomly, at the first stage of the t-SNE computation all cells were unordered around a single density peak. Shortly thereafter the $CD34^+$ lymphoid precursor cells separated from the other cells (stage 2) and the first formation of the NK and $CD127^+$ ILC clusters became apparent (stage 2 and 3). These early events were based on relatively large and highly discriminatory differences in the expression profiles between cell clusters, like the unique combination of CD34 and HLA-DR expression by $CD34^+$ cells. At stage 4 of the t-SNE computation the int-ILC cluster was positioned in the center with several distinct strands of cells forming trajectories towards the NK, $CD27^+$ ILC1, KLRG-1^+ ILC2 and $CD103^+$ ILC3 clusters. In addition, a trajectory between the ILC2 and ILC3 clusters was visible (stage 4). Furthermore, cells from the $CD8a^-$ int-ILC population connected via the $CD8a^+$ int-ILC population with NK cells, further supporting the notion that these two $CD8a^-$ and $CD8a^+$ int-ILC populations are highly related (**Fig. 2 A and Fig. S2 B**). At the final stage of the t-SNE computation the 34 clusters were defined while the connections between the individual clusters were less clear as the t-SNE algorithm eventually assigns cells in between two clusters to either one of the two. Individual marker expression patterns at stage 4 of the t-SNE computation gave insight into the separations of and the connections between clusters (**Fig. 2 B**). Here, the NK cluster was characterized by the co-expression of CD45RA, CD56, CD122 and NKp46 while the $CD127^+$ ILC cluster expressed CD45RO, CD117, CD127, CD25 and to a lesser extent CCR6 and CD103. Similarly, connections between int-ILC to ILC1, ILC2, ILC3 and NK cells were marked by (gradients of) expression of CD27, KLRG-1, CD103 and CD56, respectively. The absence of CD45RA, CD127, and CCR6 in combination with the presence of CD45RO and CD56, and divergent expression of CD117, CD122 and CD25, positioned the int-ILC in between the NK cell and ILC clusters. Interestingly, CD56 expression linked the ILC3 to the $CD8a^-$ int-ILC, and the $CD8a^+$ int-ILC to the NK cells.

We next applied Wanderlust [24] to determine changes in marker expression along this CD56 continuum (**Fig. 2 C**), which demonstrated that the expression of CD127, CD117 and CD45RO gradually decreased while that of CD45RA and CD8a increased moving from $CD8a^-$ int-ILC to NK cells, via $CD8a^+$ int-ILC; and the expression of CD127 and CD117 gradually increased from $CD8a^-$ int-ILC to ILC3. Altogether, these results suggest that these t-SNE-based trajectories may reflect potential differentiation pathways.

Expression of cytokines, transcription factors and CD94 distinguish int-ILCs from mature CD127$^+$ ILCs and NK Cells

To further characterize the int-ILC population, we used the mass cytometry data (**Fig. S3 A**) to design a minimal antibody panel to distinguish the CD127$^-$CD45RO$^+$ int-ILCs from CD45RA$^+$ NK cells, and to identify the mature CD127$^+$ ILC types through differential expression of CD117 and CRTH2 (**Fig. 3 A**). Subsequently, we analyzed the proliferative state and examined the capacity of the subsets to

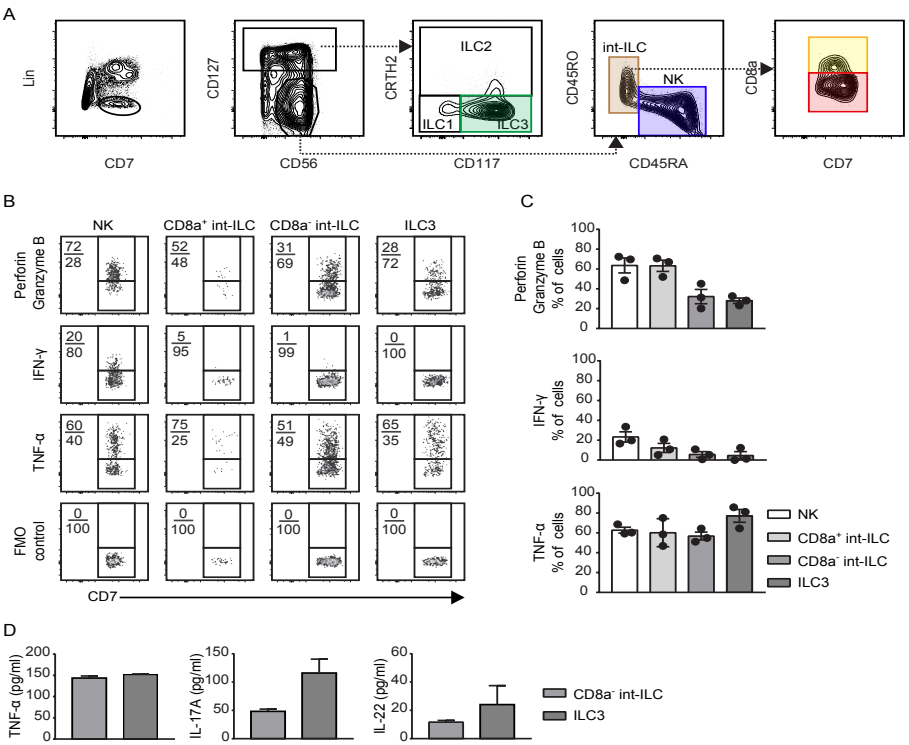


Fig. 3. Cytokine production profiles of fetal intestinal ILCs *ex vivo.* (**A**) Representative biaxial plots depicting the gating strategy for ILC1, ILC2, ILC3, NK and int-ILC subsets derived from a human fetal intestine analyzed by flow cytometry. The antibody cocktail contains lineage (Lin) markers (CD3, CD19, CD11c and CD14), and CD7, CD127, CD56, CRTH2, CD117, CD45RA, CD45RO and CD8a to allow distinction of the ILC subsets. (**B and C**) Expression of cytotoxic molecules (Perforin + Granzyme B) and cytokines (IFN-γ and TNF-α) by the indicated subsets defined in panel A after stimulation with PMA and ionomycin for 6 h. The biaxial plots (**B**) depict one representative experiment and the bar graphs (**C**) depict quantification of data obtained from 3 different human fetal intestines.Three independent experiments). FMO, fluorescence-minus-one control. Error bar shows mean ± SEM. (**D**) Bar plots depict the secretion of TNF-α, IL-17A and IL-22 by CD8a$^-$ int-ILC and ILC3, after stimulation with IL-2, IL-1β and IL-23 for 4 days, using Luminex bead-based assay of an experiment with 3 intestines and duplicate wells. (Two independent experiments). Error bar shows mean ± SD.

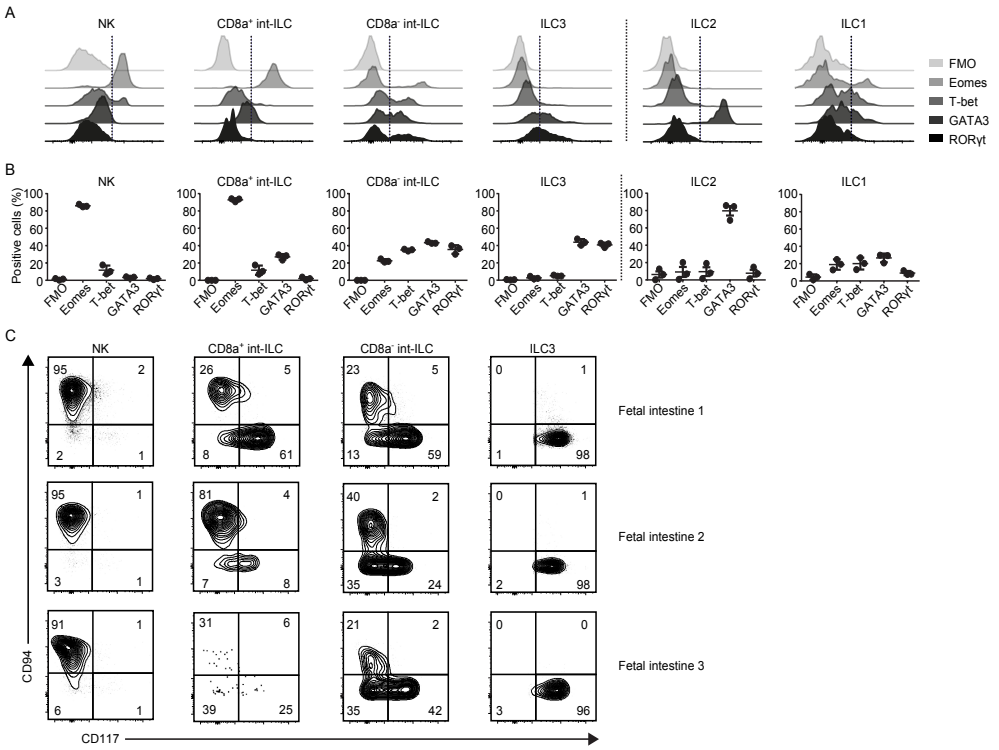


Fig. 4. Transcription factors and CD94 expression profiles of fetal intestinal ILCs *ex vivo*. (**A and B**) Flow cytometric determination of the expression of the transcription factors Eomes, T-bet, GATA3 and RORγt by the indicated ILC subsets as defined in Fig. 3 A. Histograms (**A**) depict the results with one representative human fetal intestine and the graphs (**B**) depict quantification of data obtained from 3 different human fetal intestines. (Three independent experiments). FMO, fluorescence-minus-one control. Error bar shows mean ± SD. (**C**) Biaxial plots showing the expression of CD94 and CD117 by the indicated subsets. Results on 3 human fetal intestines are shown. (Three independent experiments).

produce cytokines and express markers linked to cytolytic potential by flow cytometry. For the former we stained the cells with the proliferation marker Ki-67 ex vivo. The highest percentage of Ki-67 positive cells was present in the CD8a$^+$ int-ILC population (43.9%) while on average 20% of cells in the other subsets were Ki-67 positive (**Fig. S3 B**). Upon stimulation with PMA and ionomycin Perforin/Granzyme B was detectable in all subsets, but more profoundly in the NK cells and CD8a$^+$ int-ILCs compared with the CD8a$^-$ int-ILCs and ILC3s (**Fig. 3, B and C**). Moreover, all subsets expressed high levels of TNF-α, while IFN-γ was detected mainly in the NK cells and CD8a$^+$ int-ILCs but hardly in the CD8a$^-$ int-ILCs and ILC3s (**Fig. 3, B and C**). ILC2s expressed IL-4, IL-5 and IL-13, but ILC1s very little IFN-γ (not shown). In contrast, IL-4, IL-5 and IL-13 was undetectable in any of the other subsets (not shown) while IL-17A and IL-22 expression was higher by

ILC3s than CD8a$^-$ int-ILCs (**Fig. 3 D**).

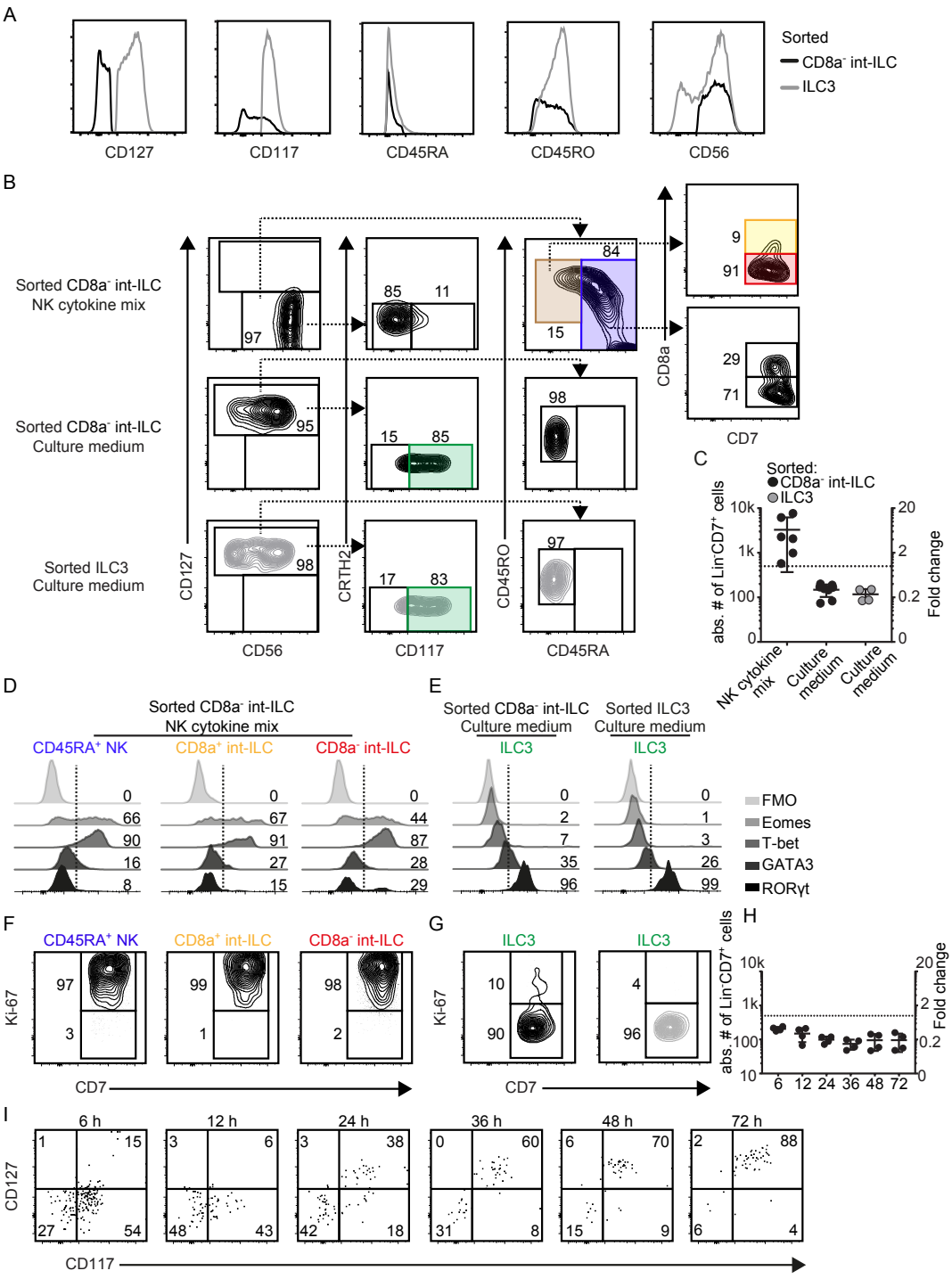
Next we determined the expression of key transcription factors associated with ILC development and phenotype. The expression of ID2, TCF7, AHR, NFIL3, ZBTB16 and TOX did not discriminate between the subsets (**Fig. S3 C**). In line with previous work [25], the ILC2 subset was strongly GATA3 positive and RORγt negative, while ILC3s were GATA3 and RORγt positive (**Fig. 4, A and B and Fig. S3 D**). However, we found only low levels of T-bet expression by ILC1 (**Fig. 4, A and B**). Notably, both mature NK cells and CD8a$^+$ int-ILCs expressed high levels of Eomes (**Fig. 4, A and B and Fig. S3, E and F**). In contrast, the CD8a$^-$ int-ILCs were heterogeneous with respect to the expression of the 4 transcription factors which were all expressed by a proportion of the cells (**Fig. 4, A and B**), an expression profile that does not correspond to those found in mature CD127$^+$ ILCs. Furthermore, multiple lineage transcription factors could be simultaneously expressed by CD8a$^-$ int-ILCs, such as T-bet and GATA3 (26.1% of CD8a$^-$ int-ILCs) (**Fig. S3, G-I**). Finally, the frequency of cells expressing Eomes decreased along the potential differentiation trajectory linking the NK cells to CD8a$^+$ int-ILCs, CD8a$^-$ int-ILCs and ILC3s, while that of RORγt increased (**Fig. 4, A and B**).

To investigate the relationship between the int-ILCs, NK cells and ILC3s further, we evaluated the expression of CD62L, CD57, CD5 and the NK cell-associated C-type lectin receptor CD94. Here, expression of CD62L, CD57 and CD5 was almost lacking and did not discriminate between the subsets (not shown) while CD94 expression was high on NK cells but virtually absent from ILC3s (**Fig. 4 C**), in agreement with previous studies [2,10]. In contrast, only part of the int-ILCs were CD94 positive with a higher expression level of CD94 on CD8a$^+$ int-ILCs compared to CD8a$^-$ int-ILCs (**Fig. 4 C**), a result in line with the lower expression of Eomes in the latter. Furthermore, in contrast with CD94$^-$ cells, CD94$^+$ cells lacked the expression of CD117, similar to mature NK cells (**Fig. 4 C**).

Together, these data indicate that the int-ILC subset is distinct from mature ILCs, where the expression pattern of the cytokines, CD94 and transcription factors link the CD8a$^+$ int-ILCs to NK cells and the CD8a$^-$ int-ILCs to ILC3s.

int-ILC can differentiate into CD45RA$^+$ NK cells

To test the hypothesis that the int-ILC subset may differentiate into CD127$^+$ ILCs and/or NK cells, we first purified the CD8a$^-$ int-ILCs by flow cytometry (**Fig. 5 A**) and performed functional differentiation assays by co-culturing these with OP9 stromal cells expressing the Notch ligand Delta-like 1 (OP9-DL1). After 7 days of culture in medium containing stem cell factor (SCF), IL-7, IL-2 and IL-15


A
CD127
CD117
CD45RA
CD45RO
CD56
Sorted
CD8a- int-ILC
ILC3
B
Sorted CD8a- int-ILC
NK cytokine mix
Sorted CD8a- int-ILC
Culture medium
Sorted ILC3
Culture medium
CD127
CD56
CRTH2
CD117
CD45RO
CD45RA
CD8a
CD7
C
Sorted:
CD8a- int-ILC
ILC3
abs. # of Lin-CD7+ cells
Fold change
NK cytokine mix
Culture medium
Culture medium
D
Sorted CD8a- int-ILC
NK cytokine mix
CD45RA+ NK
CD8a+ int-ILC
CD8a- int-ILC
E
Sorted CD8a- int-ILC
Culture medium
Sorted ILC3
Culture medium
ILC3
FMO
Eomes
T-bet
GATA3
RORγt
F
CD45RA+ NK
CD8a+ int-ILC
CD8a- int-ILC
Ki-67
CD7
G
ILC3
ILC3
Ki-67
CD7
H
abs. # of Lin-CD7+ cells
Fold change
6 12 24 36 48 72
I
6 h
12 h
24 h
36 h
48 h
72 h
CD127
CD117

(hereafter referred to as NK cytokine mix), the majority of the int-ILCs acquired a CD45RA$^+$ phenotype (**Fig. 5 B**) and expanded substantially (**Fig. 5 C**). Also, these cells upregulated CD94 (42% positive) (**Fig. S4 A**) and displayed expression of Eomes and/or T-bet, but no RORγt or GATA3 (**Fig. 5 D**), all similar to mature NK cells. Furthermore, part of these cells expressed CD8a (**Fig. 5 B**), a marker expressed by most fetal NK cells ex vivo (**Fig. 2 B**), where most of the CD8a$^+$ cells displayed the highest expression of Eomes (**Fig. S4 B**). Moreover, a small fraction of generated cells maintained the CD45RA$^-$CD45RO$^+$ int-ILC phenotype and a fraction of them also acquired CD8a (**Fig. 5 B**). In line with the suggested differentiation trajectory (**Fig. 2 C**), the expression of ILC3-associated RORγt decreased from the CD8a$^-$ int-ILCs to CD8a$^+$ int-ILCs to CD45RA$^+$ NK cells while all populations expressed high levels of NK cell-associated Eomes and/or T-bet (**Fig. 5 D**). Also, the cells became uniformly Ki-67 positive (**Fig. 5 F**), consistent with the observed increase in cell numbers (**Fig. 5 C**). As similar results were obtained when purified CD8a$^-$ int-ILCs were co-cultured with OP9 stromal cells without Delta-like 1, Notch signalling appears not to be involved (**Fig. S4 C**). In addition, upon culture on OP9-DL1 purified CD8a$^+$ int-ILCs also acquired the CD45RA$^+$ NK cell phenotype, maintained CD8a expression and expressed high levels of CD94 (84% positive) (**Fig. S4 D**). Finally, purified CD8a$^-$ int-ILCs co-cultured with OP9-DL1

Fig. 5. int-ILC can differentiate into CD45RA$^+$ NK cells and ILC3. (**A**) Representative histograms depicting the expression of CD127, CD117, CD45RA, CD45RO, CD56 by flow cytometry-purified CD8a$^-$ int-ILCs (black line) and ILC3s (grey line) from human fetal intestines. Data are representative of six independent experiments. (**B-G**) Purified CD8a$^-$ int-ILCs and ILC3s were co-cultured in 96 well plates at 500 cells/well with irradiated OP9-DL1 stromal cells for 7 days with culture medium alone or supplemented with SCF, IL-7, IL-2, IL-15 (referred to as NK cytokine mix). Generated cells were analyzed by flow cytometry. Duplicated wells were included for each condition. Representative plots show a single duplicate. (**B**) Representative biaxial plots depict the phenotypes of generated Lin$^-$CD7$^+$ cells based on the gating strategy for ILC1, ILC2, ILC3, NK and int-ILC subsets as shown in Fig. 3 A, for three different combinations of sorted cell populations (int-ILC in black contours, ILC3 in grey contours) and culture conditions as indicated. (Three to five independent experiments). (**C**) Quantification of the generated Lin$^-$CD7$^+$ cells in panel B in absolute cell number (left axis) and fold change (right axis) compared to the number of initially sorted cells (dashed line). (Two to four independent experiments). Error bar shows mean ± SD. (**D and E**) Histograms depict the expression of transcription factors Eomes, T-bet, GATA3 and RORγt by the indicated subsets generated from (**D**) sorted CD8a$^-$ int-ILCs with NK cytokine mix and (E) sorted CD8a$^-$ int-ILCs or ILC3s with culture medium. Numbers indicate the percentage of positive cells. FMO, fluorescence-minus-one control. Combined data on 5 human fetal intestines. (**F and G**) Biaxial plots depict the expression of Ki-67 by indicated subsets generated from the combinations of sorted cell populations (int-ILC in black contours, ILC3 in grey contours) and culture conditions as in panel D and E. Combined data on 5 human fetal intestines. (**H-I**) Purified CD8a$^-$ int-ILCs were co-cultured at 500 cells/well with irradiated OP9-DL1 stromal cells in culture medium and harvested at the time points indicated in hours. Duplicated wells were included in each experiment. (Two independent experiments). (**H**) Quantification of the generated Lin$^-$CD7$^+$ cells in absolute cell number (left axis) and fold change (right axis) compared to the number of initially sorted cells (dashed line). (**I**) Representative biaxial plots show the expression of CD127 and CD117 by the generated Lin$^-$CD7$^+$ cells.

and IL-15 cytokine only similarly expanded (not shown), acquired CD45RA, upregulated CD94 (41% positive), and became in part CD8a positive (**Fig. S4 E**). However, under these conditions approximately 60% of these generated cells remained CD117$^+$ (**Fig. S4 E**), suggesting an incomplete conversion to the mature NK cell phenotype [26]. Together these data indicate that in the presence of NK cytokines, proliferative CD45RA$^+$ NK cells are generated from int-ILC.

CD8a$^-$ int-ILC can differentiate into ILC3

In marked contrast, when purified CD8a$^-$ int-ILCs and ILC3s (**Fig. 5 A**) were individually co-cultured with OP9-DL1 in cytokine-free culture medium, the ILC3s retained their phenotype while the CD8a$^-$ int-ILCs acquired an ILC3 phenotype as they became CD127$^+$CD117$^+$ (**Fig. 5 B**), remained CD45RA$^-$CD45RO$^+$ (**Fig. 5 B**) and CD8a$^-$ (not shown), in the absence of cell expansion and proliferation (**Fig. 5, C and G**). This phenotype was also stable during prolonged culture (**Fig. S4 F**). In addition, these cells homogeneously expressed RORγt, but no Eomes or T-bet (**Fig. 5 E**), suggesting an established ILC3 population [2]. As similar results were observed when we co-cultured the CD8a$^-$ int-ILCs with OP9 stromal cells that lacked Notch ligand Delta-like 1, Notch signalling appears not to be involved (**Fig. S4, G and H**). Unlike CD8a$^-$ int-ILCs and ILC3s, both purified CD45RA$^+$ NK cells and CD8a$^+$ int-ILCs did not survive under these conditions. To exclude that the generation of ILC3 by int-ILCs was due to outgrowth of contaminating ILC3s, we determined cell numbers and the acquisition of CD127 and CD117 at various time points during culture. After 24 and 72 h of culture, 38% and 88% of purified CD8a$^-$ int-ILCs had acquired both CD127 and CD117, respectively, while no increase in cell numbers was observed (**Fig. 5, H and I**). Together with the observation that only a very small proportion of both the purified mature ILC3s and differentiated ILC3s from int-ILCs were Ki-67$^+$ (**Fig. 5 G**), this indicates that it is highly unlikely that selective outgrowth of contaminating ILC3s could explain the appearance of cells with an ILC3 phenotype in the CD8a$^-$ int-ILC/OP9 co-cultures. Thus, these results indicate that the CD8a$^-$ int-ILC population can differentiate into ILC3 in vitro.

Differentiation properties of CD8a$^-$ int-ILC subpopulations

By the differential expression of CD94 and CD117 three distinct CD8a$^-$ int-ILC subpopulations could be distinguished: CD94$^+$CD117$^-$, CD94$^-$CD117$^-$ and CD94$^-$CD117$^+$ (**Fig. 4 C**). We therefore investigated whether these subsets could differentiate into either NK cells or ILC3s in vitro. For this purpose, we first examined the expression of transcription factors (**Fig. 6, A and B**). This revealed that Eomes was primarily present in the CD94$^+$CD117$^-$ subset while RORγt

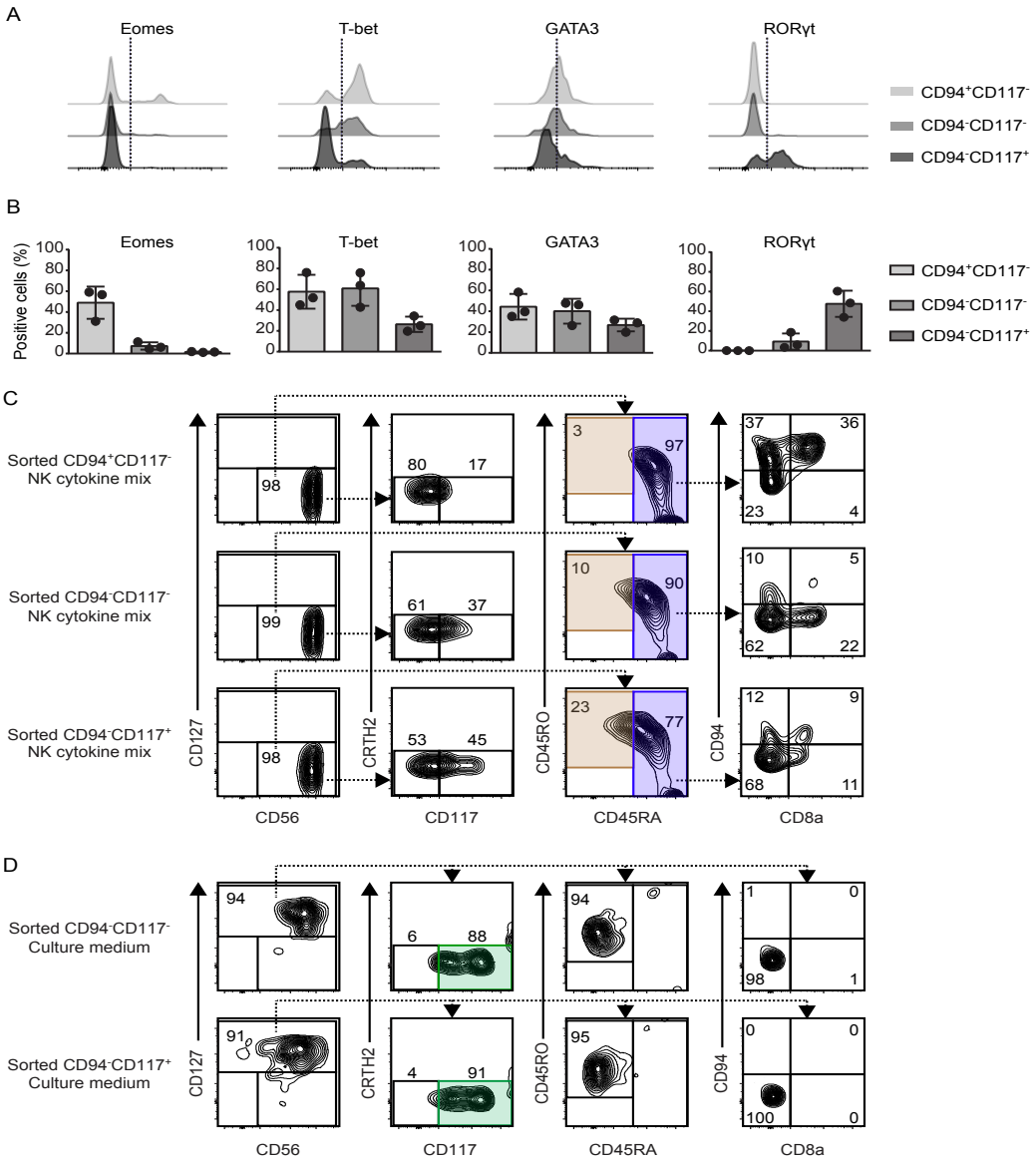


Fig. 6. Distinct differentiation properties of CD8a$^-$ int-ILC subpopulations. (**A and B**) Expression of the transcription factors Eomes, T-bet, GATA3 and RORγt by the indicated subpopulations of CD8a$^-$ int-ILC *ex vivo*. Histograms (**A**) depict the results with one fetal intestine and the graphs (**B**) depict quantification of data obtained from 3 intestines. (Two independent experiments). Error bar shows mean ± SD. (C and D) Purified CD94$^+$CD117$^-$CD8a$^-$ int-ILC, CD94$^-$CD117$^-$CD8a$^-$ int-ILC and CD94$^-$CD117$^+$CD8a$^-$ int-ILC populations were co-cultured in 96 well plates at 500 cells/well with irradiated OP9-DL1 stromal cells for 7 days with (**C**) culture medium supplemented with NK cytokine mix or (**D**) culture medium alone. Generated cells were analyzed by flow cytometry. Representative biaxial plots depict the phenotypes of the generated Lin$^-$CD7$^+$ cells based on the gating strategy for ILC1, ILC2, ILC3, NK and int-ILC subsets as shown in Fig. 3 A, for the five combinations of cell populations and culture conditions indicated. Duplicated wells were included in each condition. Representative plots show a single duplicate. (Three to four independent experiments).

expression was most pronounced in the CD94$^-$CD117$^+$ subset. In addition, all subsets expressed GATA3 and T-bet, where T-bet expression was most pronounced by the CD117$^-$ subsets. Together, these results explain the heterogeneity in the expression of transcription factors by CD8a$^-$ int-ILCs and position the CD94$^-$CD117$^-$ subset in between the CD94$^+$CD117$^-$ and CD94$^-$CD117$^+$ subsets.

Next, we purified CD94$^+$CD117$^-$, CD94$^-$CD117$^-$, and CD94$^-$CD117$^+$ subsets individually and co-cultured them with OP9-DL1 stromal cells with either NK cytokine mix or cytokine-free medium. In the presence of NK cytokine mix (**Fig. 6 C**), virtually all of the CD94$^+$CD117$^-$ cells acquired the CD45RA$^+$ NK cell phenotype, maintained expression of high levels of CD94 (76%), acquired CD8a expression (43%) and expanded substantially (12-fold; not shown). Similarly, most of the CD94$^-$CD117$^-$ cells and the majority of the CD94$^-$CD117$^+$ cells also acquired the CD45RA$^+$ NK cell phenotype, acquired expression of CD94 (17% and 25%, respectively) and CD8a (24% and 16%, respectively), and expanded (9-fold and 7-fold, respectively; not shown). In contrast, in cytokine-free medium (**Fig. 6 D**) both the CD94$^-$CD117$^-$ and CD94$^-$CD117$^+$ int-ILCs acquired an ILC3 phenotype as they became CD127$^+$, remained CD45RO$^+$CD45RA$^-$CD94$^-$CD8a$^-$ and acquired or increased the levels of CD117 expression, respectively (**Fig. 6 D**). Similar to CD8a$^+$ int-ILCs and CD45RA$^+$ NK cells the CD94$^+$CD117$^-$ int-ILCs did not survive under these conditions.

Taken together, in these in vitro experiments all three subpopulations of the CD8a$^-$ int-ILCs can differentiate into NK cells, whereas the CD94$^-$CD117$^-$ and CD94$^-$CD117$^+$ cells, but not the CD94$^+$CD117$^-$ cells can differentiate into ILC3s.

Discussion

Numerous studies have reported substantial heterogeneity in the ILC compartment [2,22]. In our mass cytometry-based approach the NK, ILC1, ILC2 and ILC3 subsets could be readily identified in the human fetal intestine as well as substantial variability within those subsets, results that were highly consistent in several samples analyzed. Based on the marker expression profiles, we identified a large number of distinct NK cell clusters, whose biological significance needs to be investigated in the future studies. In addition, we found a small CD27$^+$ ILC cluster that matches ILC1 criteria [1,27]. However, in contrast to findings from Bernink et al. (2013), but in line with other studies [22,23], only few of these ILC1s expressed T-bet. Furthermore, two additional CD127$^+$ ILC1-like clusters were identified which clustered with the NK cells due to the expression of several NK cell-associated markers, including

CD45RA, CD56, CD8a and NKp46. Consistent with previous reports [2,13,22] the ILC3 compartment was most frequent and heterogeneous, including $CD45RO^+$ ILC3, $CD45RA^+$ ILC3, $HLA\text{-}DR^+$ ILC3 and $CD56^{+/-}$ ILC3 clusters. Further studies will be required to clarify the potential functional significance of observed heterogeneity in the ILC3 lineage. We could not distinguish LTi cells from ILC3s as no specific cell surface marker for human LTi cells was available at the time we performed our analysis. Moreover, while most of the human ILCs described express CD161 [1,2], we also detected a recently described $CD161^-CD117^+$ ILC3-like cluster that clustered with the $CD127^+$ ILCs [23]. In addition, we identified two previously unknown $CD8a^+$ counterparts of ILC3s that warrants further investigation. We also observed a rare $CD34^+CD45RA^+CD117^+$ population that resembles the $CD34^+$ precursors recently described in human tonsils and intestines after birth [9,10].

Finally, we identified a $Lin^-CD7^+CD127^-CD45RO^+CD56^+$ group of cells which by unbiased clustering were positioned between the $CD45RA^+$ NK cells and $CD127^+$ ILCs, and were termed int-ILC. While in previous studies [1,2] such $CD56^+CD127^-$ cells were classified as NK cells, the simultaneous use of CD45RA and CD45RO allowed us to distinguish these $CD45RO^+$ cells from the $CD45RA^+$ NK cells. It is important to note that these int-ILCs display variable expression of several surface markers, including CD8a, CD94, CD117, CD122, CD25, CD27, KLRG-1 and CD103, indicating that they are not a homogenous group of cells. Their unique position in between the NK cells and $CD127^+$ ILCs, however, prompted us to investigate potential relationships between these cell clusters.

In recent years, several studies have explored developmental pathways by applying computational approaches on mass cytometry datasets. Wanderlust, accurately predicted B cell lymphopoiesis [24] and Wishbone was found to recover the T cell developmental pathway [28] at single-cell resolution. However, Wanderlust is not suitable to predict multi-lineage differentiation trajectories. While Wishbone can be used to identify bifurcating developmental trajectories, it needs the designation of a user-defined precursor cell type. As we wished to investigate potential developmental relationships without pre-assumptions we developed a novel computational approach to visualize the evolution of the t-SNE map over the course of the optimization. t-SNE is a non-linear dimensionality reduction algorithm which projects the high-dimensional similarities onto a two-dimensional map based on the concurrent marker expression profiles [19]. Here t-SNE not only employs bimodal distribution patterns but also incorporates gradients of the expression of cellular markers and therefore offers superior resolution. Importantly, the Cytosplore framework [21] can visualize every iteration by making use of the A-tSNE [29]. This allowed us to analyze potential relationships between

the int-ILC, ILC1, ILC2, ILC3 and NK clusters through visualization of gradients along putative differentiation trajectories. Such gradients are clearly visible in our Cytosplore analysis (**Fig. 2**) and contribute to the generation of cell clusters. Our current results indicate that at least some of those gradual changes in marker expression profiles correlate with differentiation pathways of immune subsets. Consistent with the observed plasticity among ILCs [13,15], the analysis revealed a clear trajectory between the ILC2 and ILC3 clusters. This may imply that the CD103$^-$ ILC2 can differentiate into CD103$^+$ ILC3 locally depending on physiological or pathological conditions. Alternatively, the fetal ILC2s may leave the intestine, in line with previous reports that ILC2s can be found in the peripheral blood [30] but are virtually absent in the human intestine after birth [13]. In addition, a trajectory within the CD56$^+$ cell compartment was revealed, where the CD8a$^-$ int-ILC was connected with the CD56$^+$ ILC3 on one side and with the CD8a$^+$ int-ILC and the NK cells on the other. Importantly, while the above analysis was performed on the innate cell population present in the 7 fetal intestines collectively, similar relationships between cell populations were revealed when the innate cell compartment of each fetal sample was analyzed individually, attesting to the robustness of the approach (**Fig. S2 B**).

The putative link between the int-ILC and the mature ILC and NK cells is further strengthened by several other observations. First, both the cytokine production profiles and CD94 expression profile ex vivo link the CD8a$^+$ int-ILCs to NK cells and the CD8a$^-$ int-ILCs to the ILC3s. Second, the expression pattern of the transcription factors by int-ILCs is heterogeneous with features of both NK cells (Eomes and T-bet) and ILC3s (GATA3 and RORγt) where the CD8a$^+$ int-ILCs resemble the NK cells while the CD8a$^-$ int-ILCs are closer to CD127$^+$ ILCs. Finally, in the OP9-based co-culture system, in the presence of NK cytokines purified int-ILC expanded and displayed a CD45RA$^+$Eomes$^+$/T-bet$^+$ NK cell phenotype while in the absence of cytokines the cells did not expand but acquired a stable CD127$^+$CD117$^+$RORγt$^+$ ILC3 phenotype.

Further dissection of the CD8a$^-$ int-ILC compartment demonstrated the existence of three distinct subpopulations based on differential expression of CD94 and CD117: CD94$^+$CD117$^-$, CD94$^-$CD117$^-$ and CD94$^-$CD117$^+$. In our in vitro experiments both the Eomes-expressing CD94$^+$CD117$^-$ and non-Eomes-expressing CD94$^-$CD117$^-$ and CD94$^-$CD117$^+$ subpopulations could differentiate into CD45RA$^+$ NK cells when cultured with NK cytokine mix. Furthermore, the differentiation into CD45RA$^+$ NK cells was most efficient in the case of the CD94$^+$CD117$^-$ cells and least for the CD94$^-$CD117$^+$ cells, compatible with a model where the CD94$^+$CD117$^-$ cells are positioned close to NK cells, the CD94$^-$CD117$^+$ cells most distant and the CD94$^-$

CD117$^-$ cells in between. The acquisition of CD8a by the CD94$^-$CD8a$^-$ int-ILCs in these cultures indicates that the CD8a$^+$ int-ILC may also be an intermediate stage towards NK cell differentiation. In the absence of cytokines CD117$^+$CD127$^+$ ILC3-like cells could be generated from both CD94$^-$CD117$^+$ and CD94$^-$CD117$^-$ cells but not from CD94$^+$CD117$^-$ cells. Together this indicates that CD94$^+$CD117$^-$ cells can exclusively differentiate into NK cells while the other two populations can differentiate into both NK cells and ILC3s, at least under the in vitro conditions employed.

In the absence of cytokines the int-ILC changed into ILC3-like cells without signs of cell expansion, cell division or cell death, arguing that the generation of ILC3 from the int-ILC is not due to selective outgrowth of contaminating ILC3s. In agreement, we did not observe any proliferative response of flow cytometry-purified ILC3 under the same experimental conditions.

In mice, it has been shown that developmental hierarchy of ILCs goes from the common lymphoid progenitor (CLP) to mature ILCs via α4β7-expressing lymphoid progenitor (αLP), early innate lymphoid progenitor (EILP), common helper ILC progenitor (CHILP), and the ILC precursor (ILCP) [7]. Here EILP has been distinguished from other progenitors by the lack of CD127 and ILC lineage-specific transcription factors [31], while the ILCP exhibits co-expression of transcription factors associated with ILC1, ILC2 and ILC3 subsets [32]. Finally, a murine fetal transitional CD127$^+$ ILCP (ftILCP) has been identified in the intestine which expresses varying amount of T-bet, GATA3 and RORγt [33]. In humans, two studies have shown that ILC3s can be generated from RORγt$^+$CD34$^+$ progenitors from tonsils and intestines [9,10] but little is known about the intermediate stages. Interestingly, Scoville et al. (2016) generated both NK cells and the three types of helper ILCs from these CD34$^+$ progenitors in vitro, however, the generated cells did not express CD127 but rather CD161 and intracellular ILC-related cytokine profiles. Also, a NK cell lineage-restricted CD34$^+$ progenitor was identified in the human fetal liver and bone marrow [34]. In contrast, CD127$^-$ int-ILC in our study express CD45RO and variable levels of Eomes, T-bet, GATA3 and RORγt, but no CD34 and CD45RA, markers expressed by most human progenitors. Also, with our mass cytometry approach we have been unable to identify cells with an int-ILC phenotype in the human fetal liver and fetal spleen (not shown), indicating that the int-ILC may specifically reside in mucosal tissues. Together, the cell surface phenotype, tissue distribution and transcription factor profiles of CD127$^-$ int-ILC suggests that these cells are distinct from previously identified ILC progenitors and may be in an intermediate differentiation stage among ILC lineages. Finally, we observed that the ILC3s and NK cells derived from the CD8a$^-$ int-ILC could partly revert their

phenotype upon prolonged culture indicating that the int-ILC may represent an intermediate between two plastic lineages (not shown).

It has been shown that environmental cues including OP9/OP9-DL1 stromal cells and cytokines such as IL-7 and SCF play an important role in driving ILC3 differentiation [9,10,35]. While the addition of SCF and IL-7 did promote significant expansion of the CD8a$^-$ int-ILCs, they did not differentiate into other types of cells (not shown). Instead, the differentiation of CD8a$^-$ int-ILCs toward ILC3s occurred in cytokine-free medium. In mice, Notch signalling for ILC3 development is necessary in adults but not in fetuses [8], while in humans, the differentiation of CD34$^+$ progenitors to ILC3s can occur without Notch signalling [9,10]. Consistent with these observations the generation of ILC3s from the CD8a$^-$ int-ILCs was Notch independent.

In conclusion, we delineated the heterogeneity of ILCs in the human fetal intestine and developed a computational model to predict potential differentiation trajectories based on mass cytometry data. This allowed the identification of a previously unidentified innate cell cluster that harbors cells that can differentiate into NK cells and ILC3-like cells in vitro. This may provide plasticity in the human fetal intestine in response to (external) stimuli.

Material and methods

Human fetal intestine and cell isolation

Human fetal intestines from elective abortions were collected after informed consent. Approval by the medical ethical commission of the LUMC (protocol P08.087) was obtained in accordance with the local ethical guidelines and the declaration of Helsinki. The gestational age ranged from 16 to 22 weeks. Single cell suspensions from fetal intestines were prepared as previously described. Briefly, the mesentery, colon part and meconium were removed from the fetal intestine. The intestines were then cut into small fragments and treated with 1 mM 1,4-Dithiothreitol (Fluka) in 15 mL of HBSS (Sigma-Aldrich) for 2 x 10 min (replacing buffer) at room temperature (rT) to dissolve the mucus and subsequently with 1 mM EDTA (Merck) in 15 mL of HBSS under rotation for 2 x 1 h (replacing buffer) at 37 °C to separate the epithelium from the lamina propria fraction. To obtain single-cell suspensions from the lamina propria, the intestines were rinsed with HBSS and incubated with 15 mL Iscove's Modified Dulbecco's Medium (IMDM; Lonza) supplemented with 10% FCS, 10 U/mL collagenase IV (Worthington), 200

µg/mL DNAse I grade II (Roche Diagnostics), at 37 °C overnight, after which cell suspensions were filtered through a 70 µm nylon cell strainer. Finally, the immune cells were isolated with a Percoll (GE Healthcare) gradient and stored in liquid nitrogen.

Mass cytometry antibody staining and data acquisition

Details on antibodies used are listed in **Table S1**. Conjugation of the purified antibodies with metal reporters was performed with the MaxPar X8 antibody labeling kit (Fluidigm Sciences) according to the manufacturer's instruction. Procedures for mass cytometry antibody staining and data acquisition were carried out as previously described [18]. Briefly, cells from fetal intestinal lamina propria were thawed and incubated with 1 mL 500x diluted 500 µM Cell-ID intercalator-103Rh (Fluidigm Sciences) for 15 min at rT to identify dead cells. Cells were then stained with metal-conjugated antibodies for 45 min at rT. After staining, cells were labeled with 1 mL 1,000x diluted 125 µM Cell-ID intercalator-Ir (Fluidigm Sciences) to stain all cells in MaxPar Fix and Perm Buffer (Fluidigm Sciences) overnight at 4 °C. Finally, cells were acquired by CyTOF 2™ mass cytometer (Fluidigm Sciences). Data were normalized by using EQ Four Element Calibration Beads (Fluidigm Sciences) with the reference EQ passport P13H2302 during the course of each experiment.

Mass cytometry data analysis

The biaxial plots showing antibody staining patterns in **Fig. S1** were generated in Cytobank [36]. Data for single, live $CD45^+$ cells gated from each fetal intestine individually using Cytobank [36] as shown in **Fig. S1 A**, were sample tagged and hyperbolic arcsinh transformed with a cofactor of 5 prior to t-SNE analysis. The 907 clusters shown in **Fig. S1 C** were identified by analyzing the entire immune system ($CD45^+$ cells) using the t-SNE-ACCENSE analysis pipeline as described before [18]. Next, t-SNE was performed for the ILC dataset using A-tSNE [29] in Cytosplore [21]. t-SNE was carried out with default parameters (perplexity: 30; iterations: 1000). All t-SNE plots were generated in Cytosplore. Hierarchical clustering of the heatmap was created with Pearson Correlation and average linkage clustering in MultiExperiment Viewer (http://www.tm4.org). Wanderlust analysis was performed on cells that were selected along the linear CD56 expression continuum at stage 4 of the t-SNE computation with the $CD56^+CD8a^+$ ILC3 cluster as starting point (as this cluster is located at the outer end of the CD56 trajectory), using Cyt in Matlab [24].

Antibodies and flow cytometry

FITC-conjugated anti-CD11c (3.9), PerCP-Cy5.5-conjugated anti-CD45RO (UCHL1), PE/Dazzle 594-conjugated anti-CD45RA (HI100), PE-Cy7-conjugated anti-CD127 (A019D5), Brilliant Violet 605-conjugated anti-CRTH2 (BM16), anti-T-bet (4B10), PE-conjugated anti-T-bet (4B10), anti-IFN-γ (4S.B3), anti-IL-5 (TRFK5), anti-IL-13 (JES10-5A2), anti-IL-17A (BL168) and anti-TNF-α (Mab11) were purchased from Biolegend. The following monoclonal antibodies were purchased from BD: FITC-conjugated anti-CD3 (SK7), anti-CD19 (4G7) and anti-CD14 (MφP9), APC-conjugated anti-CD117 (YB5.B8), APC-R700-conjugated anti-CD56 (NCAM16.2), V450-conjugated anti-CD7 (M-T701), Brilliant Violet 605-conjugated anti-CD94 (HP-3D9), PE-conjugated anti-CD94 (HP-3D9), anti-RORγt (Q21-559), anti-Ki-67 (B56), anti-Perforin (δG9), anti-IL-4 (3010.211). PE-conjugated anti-Eomes (WD1928), anti-GATA3 (TWAJ), anti-Granzyme B (GB11), and eFluor 660-conjugated anti-GATA3 (TWAJ) were purchased from eBioscience. PE-conjugated anti-IL-22 (IC7821P) was purchased from R&D systems. Pacific Orange-conjugated anti-CD8a (3B5) was purchased from Life technologies.

For the cell surface staining, cells were incubated with fluorochrome-conjugated antibodies and human FC block (Biolegend) for 30-45 min at 4 °C. The transcription factor staining was performed by using Foxp3 Staining Buffer Set (eBioscience) according to the manufacturer's instruction. For the intracellular cytokine staining/ cytotoxic molecule, cells were stimulated with 0.1 mg/mL PMA (Sigma-Aldrich) and 1 μg/mL Ionomycin (Sigma-Aldrich) for 6 h at 37 °C and GolgiPlug (BD Biosciences) was added for the final 4 h after which cells were stained by using Fixation Buffer and Intracellular Staining Perm Wash Buffer (Biolegend). Cells were acquired on an LSR II cytometer (BD Biosciences) or sorted on a FACSAria™ III sorter (BD Biosciences) based on the gating strategy as shown in **Fig. 3 A.** Data were analyzed with FlowJo V10 software.

Quantitative Real-Time PCR (RT-PCR)

RNA extraction was performed with the NucleoSpin® RNA XS kit (Macherey-Nagel). cDNA was synthesized with the High Capacity cDNA Reverse Transcription kit (Applied Biosystems). RT-PCR was performed in a StepOnePlus™ Real-Time PCR Systems (Applied Biosystems) with FastStart Universal SYBR Green Master Mix (Roche). ΔCt values were calculated using GAPDH as reference gene. The sequences of RT-PCR primers are as follows: GAPDH, forward primer: 5'-GTCTCCTCTGACTTCAACAGCG-3'; reverse primer, 5'-ACCACCCTGTTGCTGTAGCCAA-3'; AHR, forward primer: 5'-CTTAGGCTCAGCGTCAGTTA-3'; reverse primer, 5'-GTAAGTTCAGGCCTTCTCTG-3'; ID2, forward primer: 5'-TTGTCAGCCTGCATCACCAGAG-3';

reverse primer, 5'-AGCCACACAGTGCTTTGCTGTC-3'; NFIL3, forward primer: 5'-TGGAGAAGACGAGCAACAGGTC-3'; reverse primer, 5'-CTTGTGTGGCAAGGCAGAGGAA-3'; ZBTB16, forward primer: 5'-GAGCTTCCTGATAACGAGGCTG-3'; reverse primer, 5'-AGCCGCAAACTATCCAGGAACC-3'; TOX, forward primer: 5'-AGCATACAGAGCCAGCCTTG-3'; reverse primer, 5'-TGCATGGCAGTTAGGTGAGG-3'; and TCF1, forward primer: 5'-TGCAGCTATACCCAGGCTGG-3'; reverse primer, 5'-CCTCGACCGCCTCTTCTTC-3'.

Cell culture and differentiation assays

OP9-DL1 or OP9 stromal cells were maintained in Minimum Essential Medium α (Lonza) supplemented with 10% FCS. Flow cytometry-purified CD8a$^-$ int-ILCs or CD94$^+$CD117$^-$, CD94$^-$CD117$^-$ and CD94$^-$CD117$^+$ subpopulations (500 cells/well) or CD8a$^+$ int-ILCs (100 cells/well) were co-cultured with irradiated OP9 or OP9-DL1 stromal cells (1,500 RAD, 5,000 cells/well) in a 96 well plate (Corning) and maintained in culture medium (IMDM supplemented with 10% human serum or in culture medium containing 25 ng/mL SCF (Miltenyi Biotec), 25 ng/mL IL-7 (Peprotech), 10 U/mL IL-2 (Novartis) and 10 ng/mL IL-15 (R&D Systems) or only IL-15. The phenotype of generated progeny was determined by flow cytometry.

Cytokine secretion

CD8a$^-$ int-ILCs and ILC3 (2,000 cells/well) were stimulated with 10 U/mL IL-2 (Novartis), 50 ng/mL IL-1β (Peprotech) and 50 ng/mL IL-23 (Peprotech) for 4 days. TNF-a, IL-17A and IL-22 were measured by using Bio-Plex Pro™ human cytokine 17-plex panel kit and Bio-Plex Pro™ human Treg cytokine panel 12-plex kit (Bio-Rad).

Online supplemental material

Fig. S1 shows the mass cytometry-based analysis of the entire immune system in the human fetal intestine. Fig. S2. Top panel: t-SNE plots revealing the major lineage subsets, Bottom panel: t-SNE analysis of the innate cell compartment in the individual fetal samples. Fig. S3 shows Ki-67 and transcription factor expression profiles of fetal intestinal ILCs ex vivo. Fig. S4 is related to **Fig. 5**, and shows that CD8a$^-$ int-ILC can differentiate into CD45RA$^+$ NK Cells and ILC3. Table S1 lists the information of antibodies used in the mass cytometry data.

Author contributions

NL, VvU and FK conceived the study and wrote the manuscript. NL performed most experiments with the help of VvU. TH, NP, VvU, EE, AV and BL developed the Cytosplore application. VvU and NL performed mass cytometry analyses. AT performed the RT-PCR. SCdSL provided human fetal intestines. JvB provided conceptual input. All authors discussed the results and commented on the manuscript.

Acknowledgements

We thank the Center for Contraception, Abortion and Sexuality (CASA) in Leiden and The Hague for their efforts in collecting and providing the fetal material, K. Lodder, T. van Herwaarden, Dr. M. Bialecka, Dr. F. Wang and M. Gomes Fernandes for the dissection of the fetal intestines, E. de Haas for flow cytometric cell sorting and E. van Beelen for Luminex assay. This research was supported by the Leiden University Medical Center, The Netherlands Organization for Scientific Research (ZonMW grant 91112008) and NWO Applied Technical Sciences grant 12721 (VANPIRE). N. Li was supported by the China Scholarship Council. We thank Drs. F.J.T. Staal, K. Schepers and T. Cupedo for critical review of the manuscript.

The authors declare no competing financial interests.

References

1. Spits, H. et al. Innate lymphoid cells--a proposal for uniform nomenclature. Nat. Rev. Immunol. 13, 145–149 (2013).
2. Hazenberg, M. D. & Spits, H. Human innate lymphoid cells. Blood 124, 700–709 (2014).
3. Cella, M. et al. A human natural killer cell subset provides an innate source of IL-22 for mucosal immunity. Nature 457, 722–725 (2009).
4. Cupedo, T. et al. Human fetal lymphoid tissue-inducer cells are interleukin 17-producing precursors to RORC+ CD127+ natural killer-like cells. Nat. Immunol. 10, 66–74 (2009).
5. Gasteiger, G., Fan, X., Dikiy, S., Lee, S. Y. & Rudensky, A. Y. Tissue residency of innate lymphoid cells in lymphoid and nonlymphoid organs. Science 350, 981–985 (2015).
6. Juelke, K. & Romagnani, C. Differentiation of human innate lymphoid cells (ILCs). Curr. Opin. Immunol. 38, 75–85 (2016).
7. Ishizuka, I. E., Constantinides, M. G., Gudjonson, H. & Bendelac, A. The Innate Lymphoid Cell Precursor. Annu. Rev. Immunol. 34, 299–316 (2016).
8. Possot, C. et al. Notch signaling is necessary for adult, but not fetal, development of RORγt(+) innate lymphoid cells. Nat. Immunol. 12, 949–958 (2011).
9. Montaldo, E. et al. Human RORγt CD34 Cells Are Lineage-Specified Progenitors of Group 3 RORγt Innate Lymphoid Cells. Immunity 41, 988–1000 (2014).
10. Scoville, S. D. et al. A Progenitor Cell Expressing Transcription Factor RORγt Generates All Human Innate Lymphoid Cell Subsets. Immunity 44, 1140–1150 (2016).
11. Lim, A. I. et al. Systemic Human ILC Precursors Provide a Substrate for Tissue ILC Differentiation. Cell 168, 1086–1100.e10 (2017).
12. Bernink, J. H. et al. Human type 1 innate lymphoid cells accumulate in inflamed mucosal tissues. Nat. Immunol. 14, 221–229 (2013).
13. Bernink, J. H. et al. Interleukin-12 and -23 Control Plasticity of CD127(+) Group 1 and Group 3 Innate Lymphoid Cells in the Intestinal Lamina Propria. Immunity 43, 146–160 (2015).
14. Lim, A. I. et al. IL-12 drives functional plasticity of human group 2 innate lymphoid cells. J. Exp. Med. 213, 569–583 (2016).
15. Bal, S. M. et al. IL-1β, IL-4 and IL-12 control the fate of group 2 innate lymphoid cells in human airway inflammation in the lungs. Nat. Immunol. 17, 636–645 (2016).
16. Bandura, D. R. et al. Mass cytometry: technique for real time single cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry. Anal. Chem. 81, 6813–6822 (2009).
17. van Unen, V. et al. Visual analysis of mass cytometry data by hierarchical stochastic neighbour embedding reveals rare cell types. Nat. Commun. (2017).
18. 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).
19. Maaten, L. van der & Hinton, G. Visualizing Data using t-SNE. J. Mach. Learn. Res. 9, 2579–2605 (2008).
20. Shekhar, K., Brodin, P., Davis, M. M. & Chakraborty, A. K. Automatic Classification of Cellular Expression by Nonlinear Stochastic Embedding (ACCENSE). Proc. Natl. Acad. Sci. U. S. A. 111, 202–207 (2014).
21. Höllt, T. et al. Cytosplore: Interactive Immune Cell Phenotyping for Large Single-Cell Datasets. Comput. Graph. Forum 35, 171–180 (2016).
22. Björklund, Å. K. et al. The heterogeneity of human CD127(+) innate lymphoid cells revealed by single-cell RNA sequencing. Nat. Immunol. 17, 451–460 (2016).
23. Simoni, Y. et al. Human Innate Lymphoid Cell Subsets Possess Tissue-Type Based Heterogeneity in Phenotype and Frequency. Immunity 46, 148–161 (2017).
24. Bendall, S. C. et al. Single-Cell Trajectory Detection Uncovers Progression and Regulatory Coordination in Human B Cell Development. Cell 157, 714–725 (2014).

25. Artis, D. & Spits, H. The biology of innate lymphoid cells. Nature 517, 293–301 (2015).

26. Freud, A. G. et al. Evidence for discrete stages of human natural killer cell differentiation in vivo. J. Exp. Med. 203, 1033–1043 (2006).

27. Klose, C. S. N. et al. Differentiation of type 1 ILCs from a common progenitor to all helper-like innate lymphoid cell lineages. Cell 157, 340–356 (2014).

28. Setty, M. et al. Wishbone identifies bifurcating developmental trajectories from single-cell data. Nat. Biotechnol. 34, 637–645 (2016).

29. Pezzotti, N. et al. Approximated and User Steerable tSNE for Progressive Visual Analytics. IEEE Trans. Vis. Comput. Graph. (2016). doi:10.1109/TVCG.2016.2570755

30. Mjösberg, J. M. et al. Human IL-25- and IL-33-responsive type 2 innate lymphoid cells are defined by expression of CRTH2 and CD161. Nat. Immunol. 12, 1055–1062 (2011).

31. Yang, Q. et al. TCF-1 upregulation identifies early innate lymphoid progenitors in the bone marrow. Nat. Immunol. 16, 1044–1050 (2015).

32. Ishizuka, I. E. et al. Single-cell analysis defines the divergence between the innate lymphoid cell lineage and lymphoid tissue-inducer cell lineage. Nat. Immunol. 17, 269–276 (2016).

33. Bando, J. K., Liang, H.-E. & Locksley, R. M. Identification and distribution of developing innate lymphoid cells in the fetal mouse intestine. Nat. Immunol. 16, 153–160 (2015).

34. Renoux, V. M. et al. Identification of a Human Natural Killer Cell Lineage-Restricted Progenitor in Fetal and Adult Tissues. Immunity 43, 394–407 (2015).

35. Satoh-Takayama, N. et al. IL-7 and IL-15 independently program the differentiation of intestinal CD3-NKp46+ cell subsets from Id2-dependent precursors. J. Exp. Med. 207, 273–280 (2010).

36. Kotecha, N., Krutzik, P. O. & Irish, J. M. Web-based analysis and publication of flow cytometry experiments. Curr. Protoc. Cytom. Chapter 10, Unit10.17 (2010).

Supplemental information

	Antigen	Tag	Clone	Supplier	Cat.	Final dilution
1	CD127	^{165}Ho	AO19D5	DVS	3165008B	1/800
2	CCR6	^{141}Pr	G034E3	DVS	3141003A	1/200
3	CD8a	^{146}Nd	RPA-T8	DVS	3146001B	1/200
4	CD11c	^{162}Dy	Bu15	DVS	3162005B	1/200
5	CD38	^{172}Yb	HIT2	DVS	3172007B	1/200
6	CD45	^{89}Y	HI30	DVS	3089003B	1/100
7	CD117	^{143}Nd	104D2	DVS	3143001B	1/100
8	CD4	^{145}Nd	RPA-T4	DVS	3145001B	1/100
9	CD16	^{148}Nd	3G8	DVS	3148004B	1/100
10	CD25	^{149}Sm	2A3	DVS	3149010B	1/100
11	CD123	^{151}Eu	6H6	DVS	3151001B	1/100
12	CD7	^{153}Eu	CD7-6B7	DVS	3153014B	1/100
13	CD163	^{154}Sm	GHI/61	DVS	3154007B	1/100
14	CCR7	^{159}Tb	G043H7	DVS	3159003A	1/100
15	CD14	^{160}Gd	M5E2	DVS	3160001B	1/100
16	CD161	^{164}Dy	HP-3G10	DVS	3164009B	1/100
17	CD27	^{167}Er	O323	DVS	3167002B	1/100
18	CD45RA	^{169}Tm	HI100	DVS	3169008B	1/100
19	CD3	^{170}Er	UCHT1	DVS	3170001B	1/100
20	PD-1	^{175}Lu	EH 12.2H7	DVS	3175008B	1/100
21	CD56	^{176}Yb	NCAM16.2	DVS	3176008B	1/100
22	CD11b	^{144}Nd	ICRF44	DVS	3144001B	1/100
23	TCRgd	^{152}Sm	11F2	DVS	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
DVS Sciences (DVS), eBioscience (eBio) and Biolegend (BioL).						

Table S1. CyTOF antibody panel
The conjugation, validation and titration of all the antibodies which were not bought from DVS were done in house.

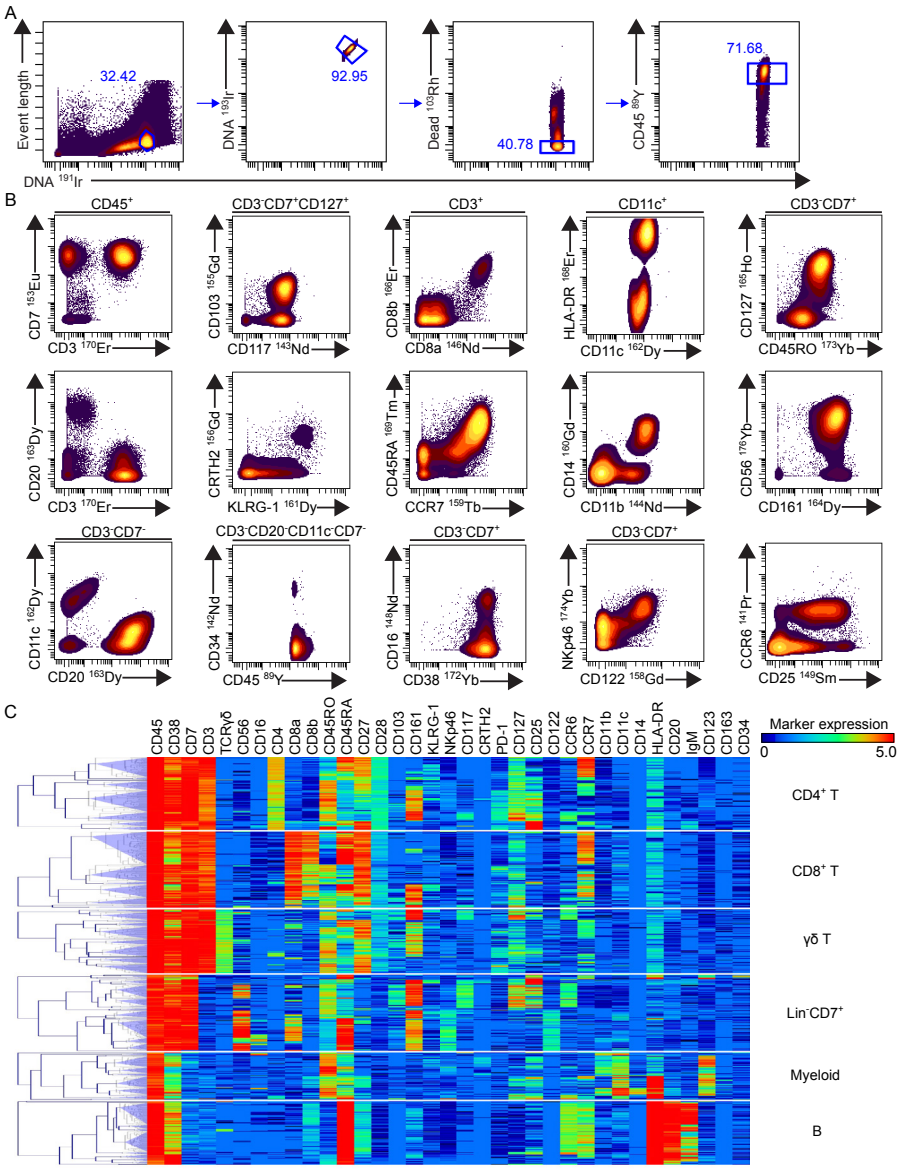


Fig. S1. Analysis of the entire immune system in the human fetal intestine using mass cytometry. (**A**) Representative biaxial density dot plots from one fetal intestine showing the gating strategy for single, live CD45+ cells with percentages analysed by mass cytometry (N=7). Event length is a mass cytometry parameter defined as signal duration for the number of scans taken to acquire a given ion cloud. (**B**) Representative biaxial density dot plots from one fetal intestine showing the typical cell staining profiles of the antibodies used in mass cytometry after gating as shown in panel A (N=7). Gated cell populations are annotated above the plots. (**C**) A heatmap showing the median ArcSinh5-transformed marker expression values of the total 907 clusters identified by analyzing the entire immune system ($CD45^+$ cells) from 7 fetal intestines using the t-SNE-ACCENSE analysis pipeline (clustering per individual sample) for each of the 6 major immune lineages individually, as described before[18], and hierarchical clustering thereof. four independent experiments).

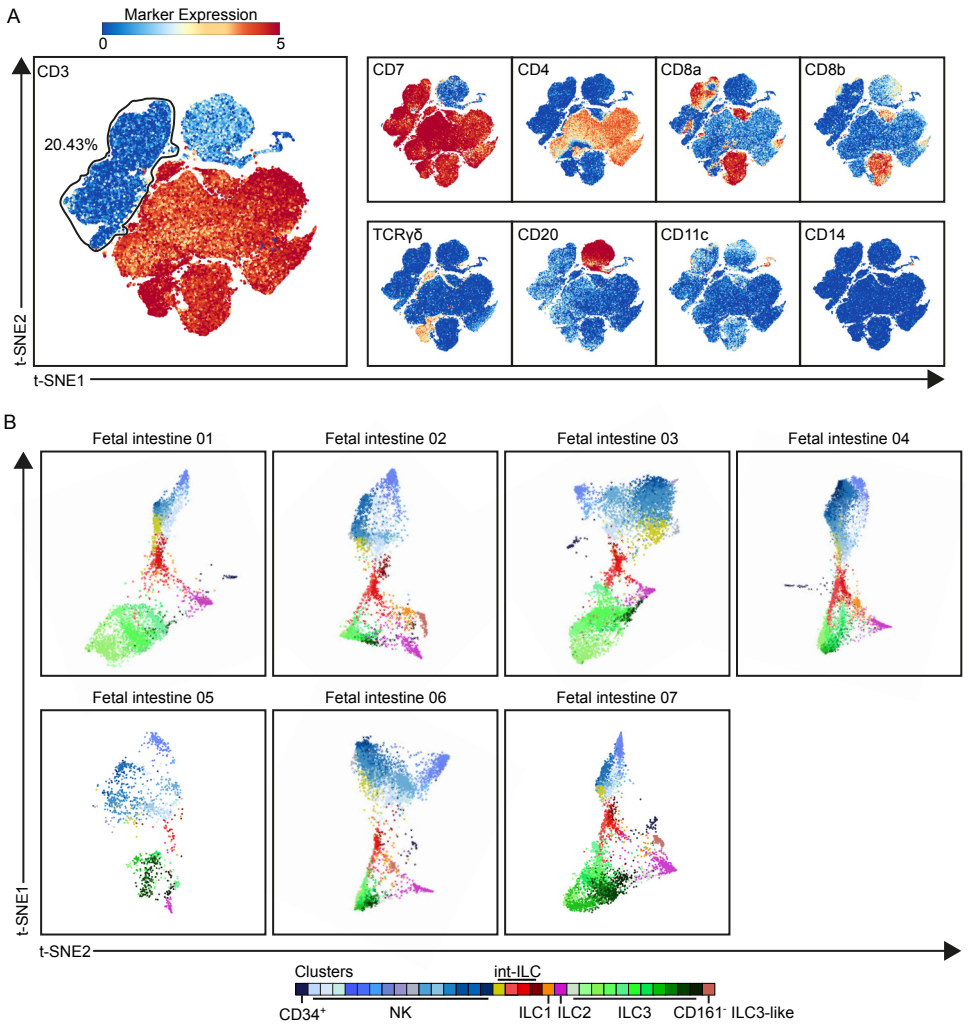
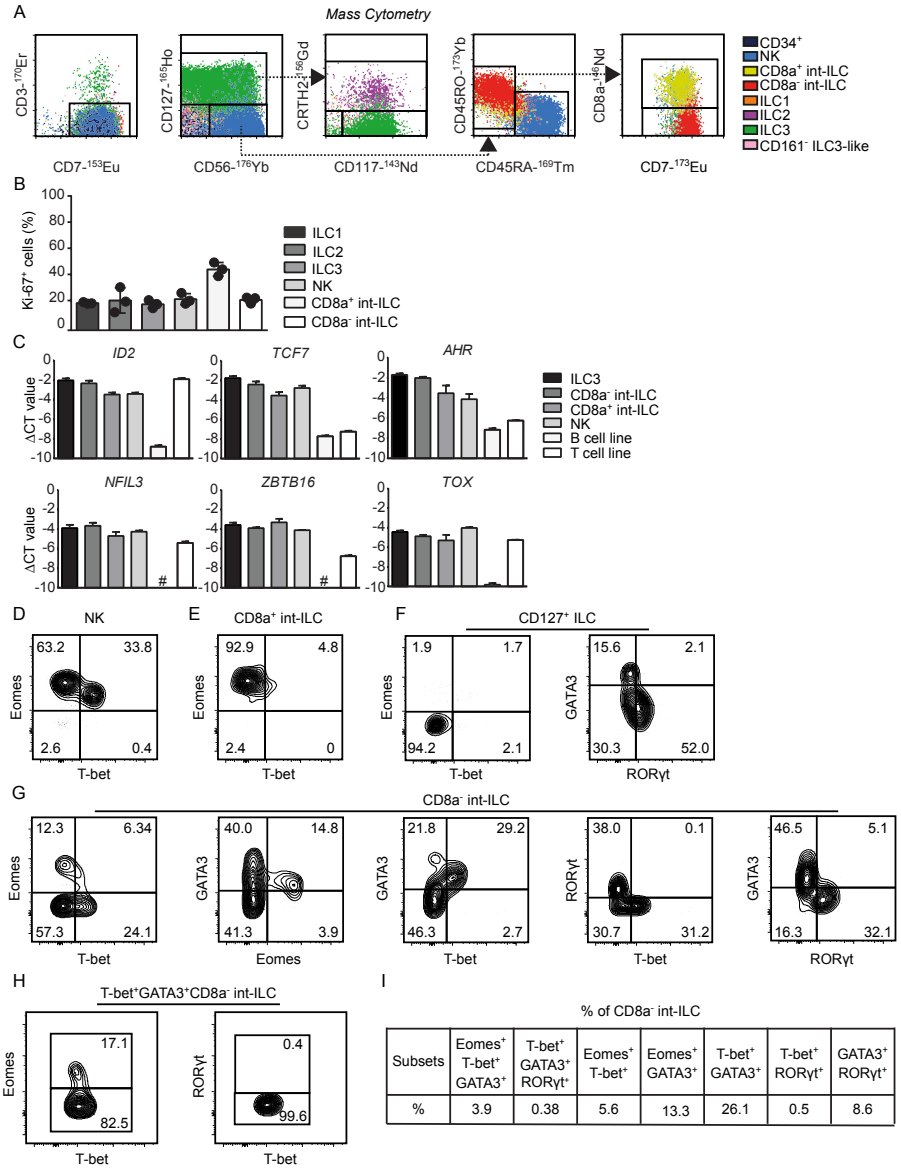


Fig. S2. t-SNE-based analysis of the Lin$^-$CD7$^+$ innate immune compartment in the human fetal intestine. (A) Gating strategy of the Lin$^-$CD7$^+$ innate cell population in the human fetal intestine. t-SNE embeddings of the collective CD45$^+$ cells (2.2x10^5 cells) from 7 fetal intestines at single-cell resolution. Colors of cells represent ArcSinh5-transformed expression values of indicated markers. (**B**) Monitoring t-SNE computation dynamics for each individual fetal intestine. t-SNE embeddings of the ILC mass cytometry data showing single cells at stage 4 of the optimization course of the t-SNE computation for each individual fetal intestine (N=7). Colors represent the cluster partitions.



Subsets	Eomes+ T-bet+ GATA3+	T-bet+ GATA3+ RORγt+	Eomes+ T-bet+	Eomes+ GATA3+	T-bet+ GATA3+	T-bet+ RORγt+	GATA3+ RORγt+
%	3.9	0.38	5.6	13.3	26.1	0.5	8.6

Fig. S3. Ki-67 and transcription factor expression profiles of fetal intestinal ILCs *ex vivo*. (**A**) Minimal antibody panel required for phenotyping ILC and int-ILC subsets in the human fetal intestine. Biaxial plots showing the expression of the indicated markers by ILC1, ILC2, ILC3, NK and int-ILC subsets based on mass cytometry data derived from 7 human fetal intestines. Color represents the different subsets identified by the t-SNE-based analysis shown in Fig. 1 B and numbers of x-axis and y-axis represent ArcSinh5-transformed expression values of indicated markers. (**B**) Quantification of Ki-67 positive cells within indicated subsets obtained from 3 different human fetal intestines (Two independent experiments). Error bar shows mean ± SD. (**C**) Relative mRNA expression of ID2, TCF, AHR, NFIL3, ZBTB16 and TOX by the purified ILC subsets, B and T cell lines analyzed by RT-PCR (Three independent experiments). GAPDH as reference gene. # indicates that the ΔCt value is below -10. Error bar shows mean ± SD. (**D**) Representative biaxial plots showing the expression of T-bet and Eomes, and GATA3 and RORγt by fetal intestinal CD127$^+$ ILC. (**E-F**) Representative biaxial plots showing the expression of T-bet and Eomes by (**E**) fetal intestinal NK cells, and (**F**) CD8a$^+$ int-ILC as defined in Fig. 3 A with flow cytometry. (**G**) Representative biaxial plots showing the combinatorial expression profiles of Eomes, T-bet, GATA3 and RORγt by fetal intestinal CD8a$^-$ int-ILC. (**H**) Expression of Eomes and RORγt by fetal intestinal T-bet$^+$GATA3$^+$CD8a$^-$ int-ILC. (**I**) Quantification of transcription factor positive cells of fetal intestinal CD8a$^-$ int-ILC. (Two independent experiments).

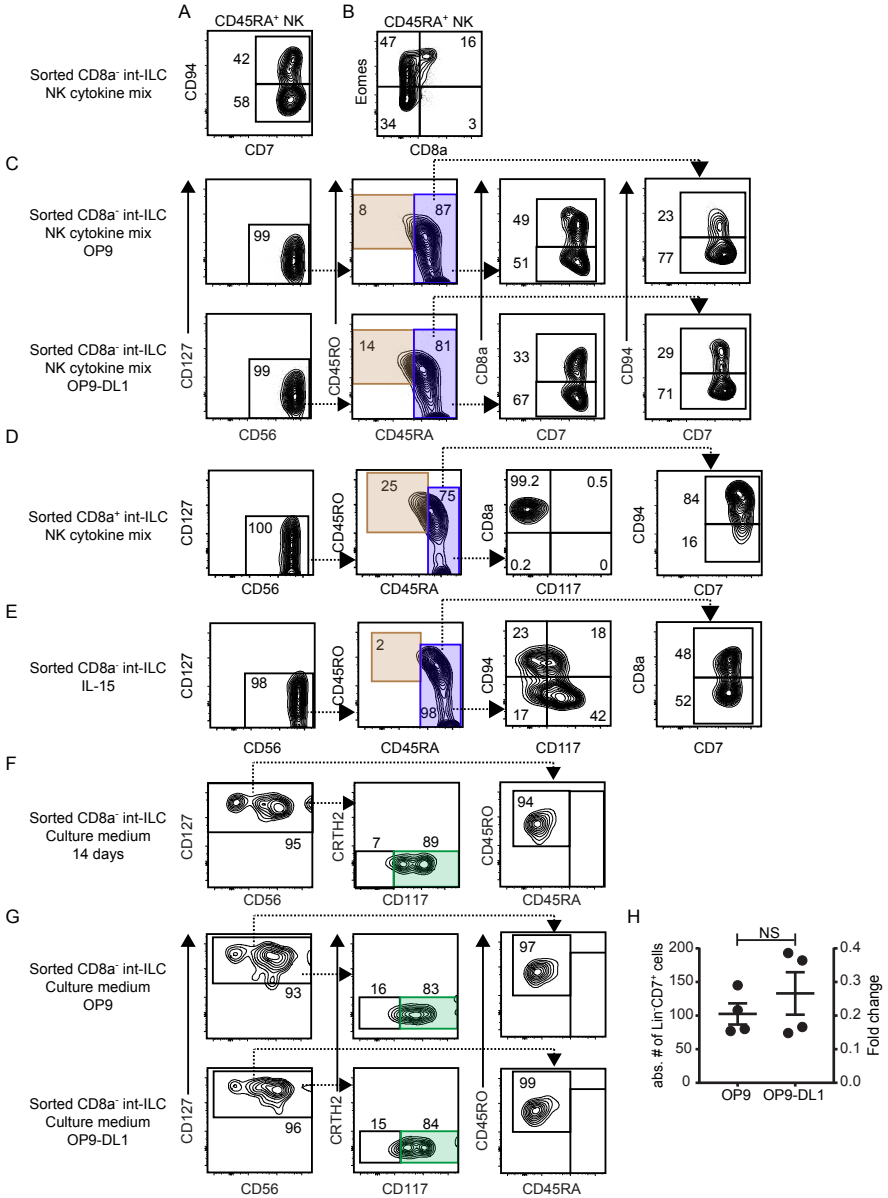
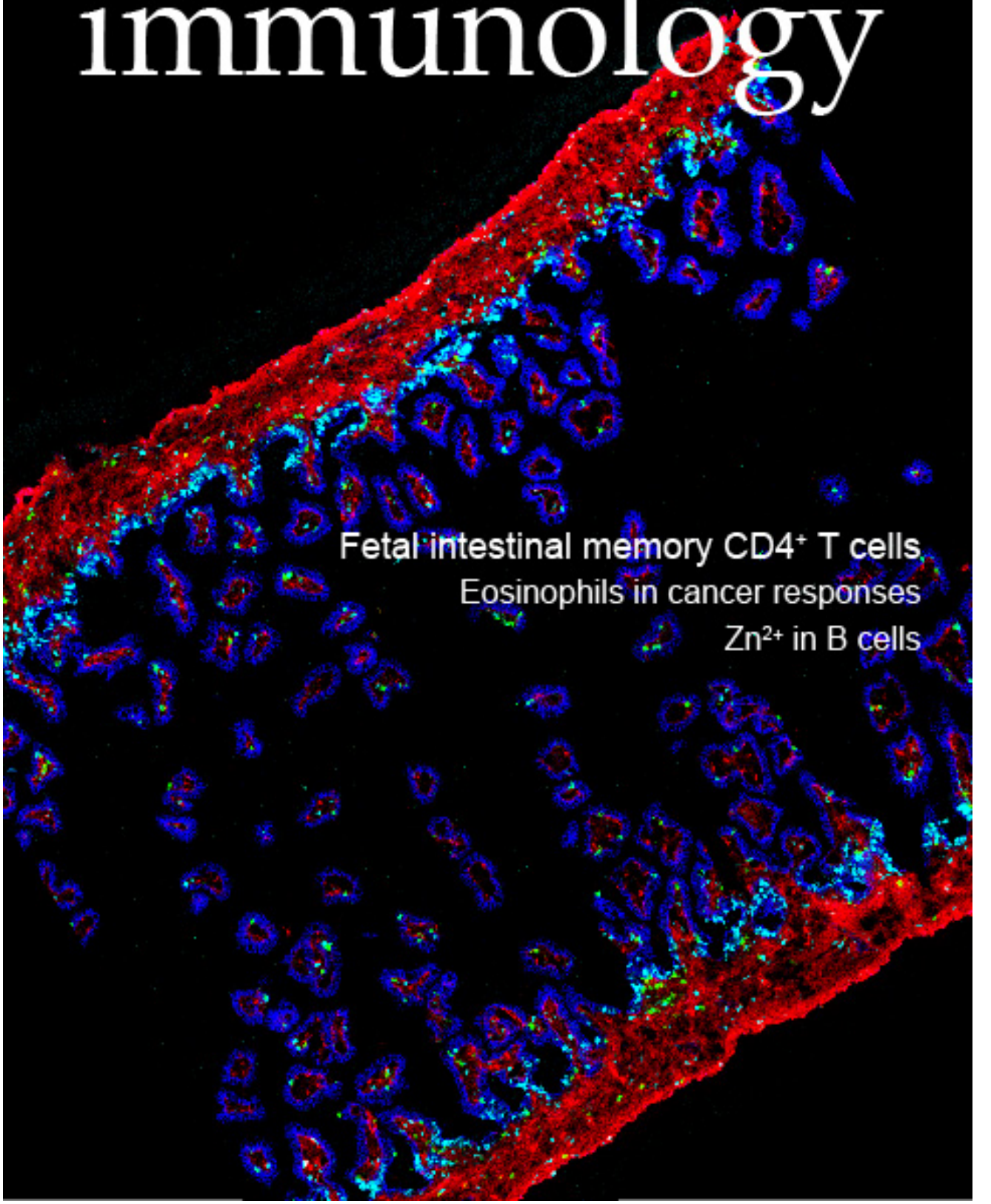


Fig. S4. CD8a$^-$ int-ILC can differentiate into CD45RA$^+$ NK Cells and ILC3. (**A-E**) Purified CD8a$^-$ int-ILC (500 cells/well) and CD8a$^+$ int-ILC (100 cells/well) populations were co-cultured with irradiated OP9-DL1 or OP9 stromal cells for 7 days either with culture medium supplemented with SCF, IL-7, IL-2, IL-15 (referred to as NK cytokine mix) or supplemented with IL-15. Generated cells were harvested and analyzed by flow cytometry. Duplicated wells were included for each condition in each experiment. Representative plots show a single duplicate. (**A and B**) Biaxial plots show the expression of CD94 (**A**) and the transcription factor Eomes (**B**) by generated CD45RA$^+$ NK cells from sorted CD8a$^-$ int-ILCs with NK cytokine mix. (Three to five independent experiments). (**C-E**) Representative biaxial plots depict the phenotypes of the generated Lin$^-$CD7$^+$ cells from (**C**) sorted CD8a$^-$ int-ILCs with NK cytokine mix with irradiated OP9-DL1 or OP9 stromal cells and (**D**) sorted CD8a$^+$ int-ILCs with NK cytokine mix and (**E**) sorted CD8a$^-$ int-ILCs with IL-15. (Two independent experiments). (**F-H**) Purified CD8a$^-$ int-ILC (500 cells/well) were co-cultured either with irradiated OP9-DL1 or with OP9 stromal cells with cytokine-free culture medium for 14 days (**F**) and 7 days (**G and H**). Generated cells were harvested and analyzed by flow cytometry (**F and G**). Representative biaxial plots depict the phenotypes of the generated Lin$^-$CD7$^+$ cells. (Two independent experiments). (**H**) Quantification in absolute cell number (left axis) and fold change (right axis) compared to the number of initially sorted cells. (Two independent experiments). Error bar shows mean ± SEM.

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

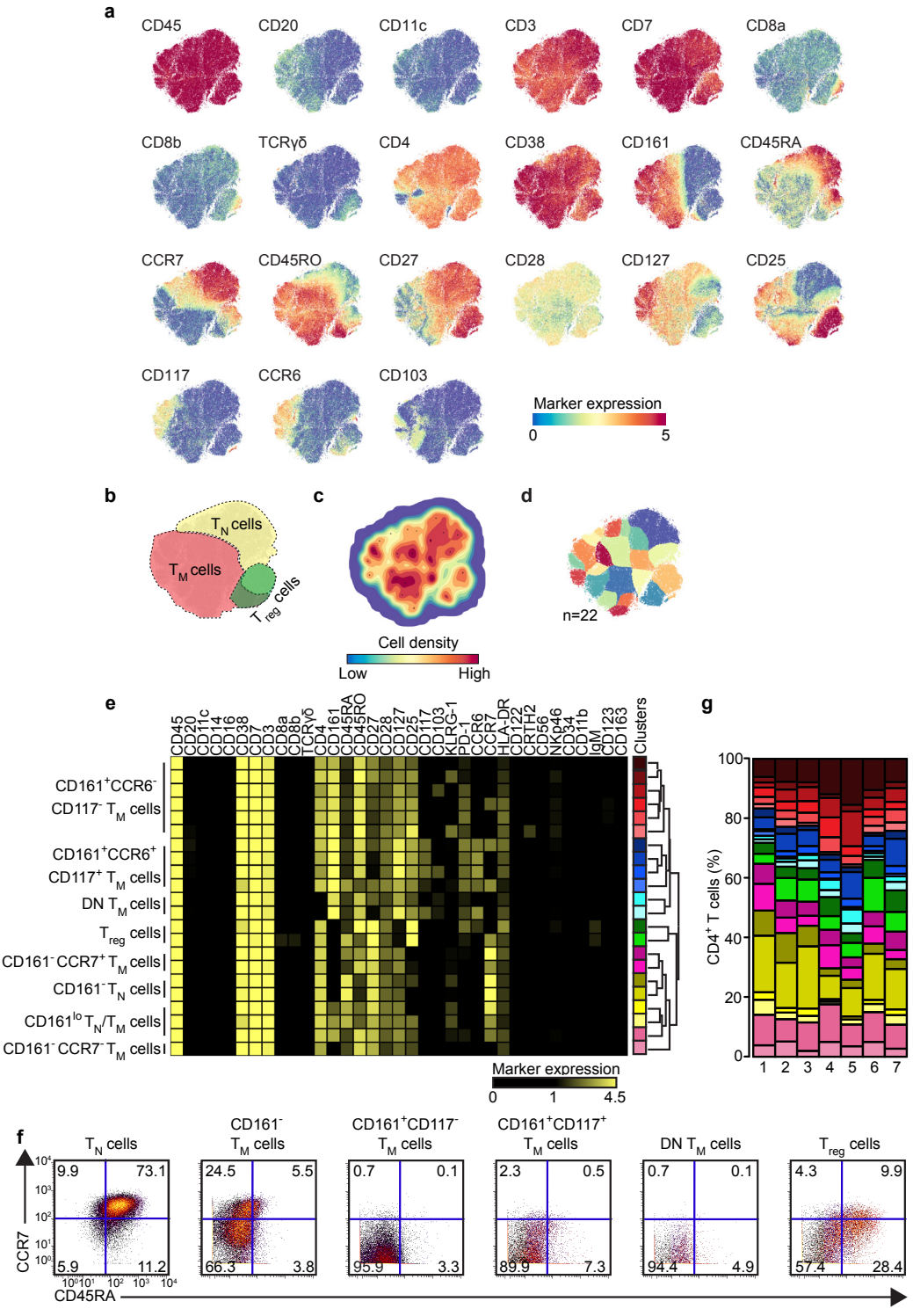
Adaptive immunity is founded on the selection and expansion of antigen-specific T cells from a clonally diverse pool of naive precursors[1]. Naive T cells recirculate among lymph nodes to survey the array of peptide epitopes bound to major histocompatibility complex (MHC) proteins on the surface of antigen-presenting 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

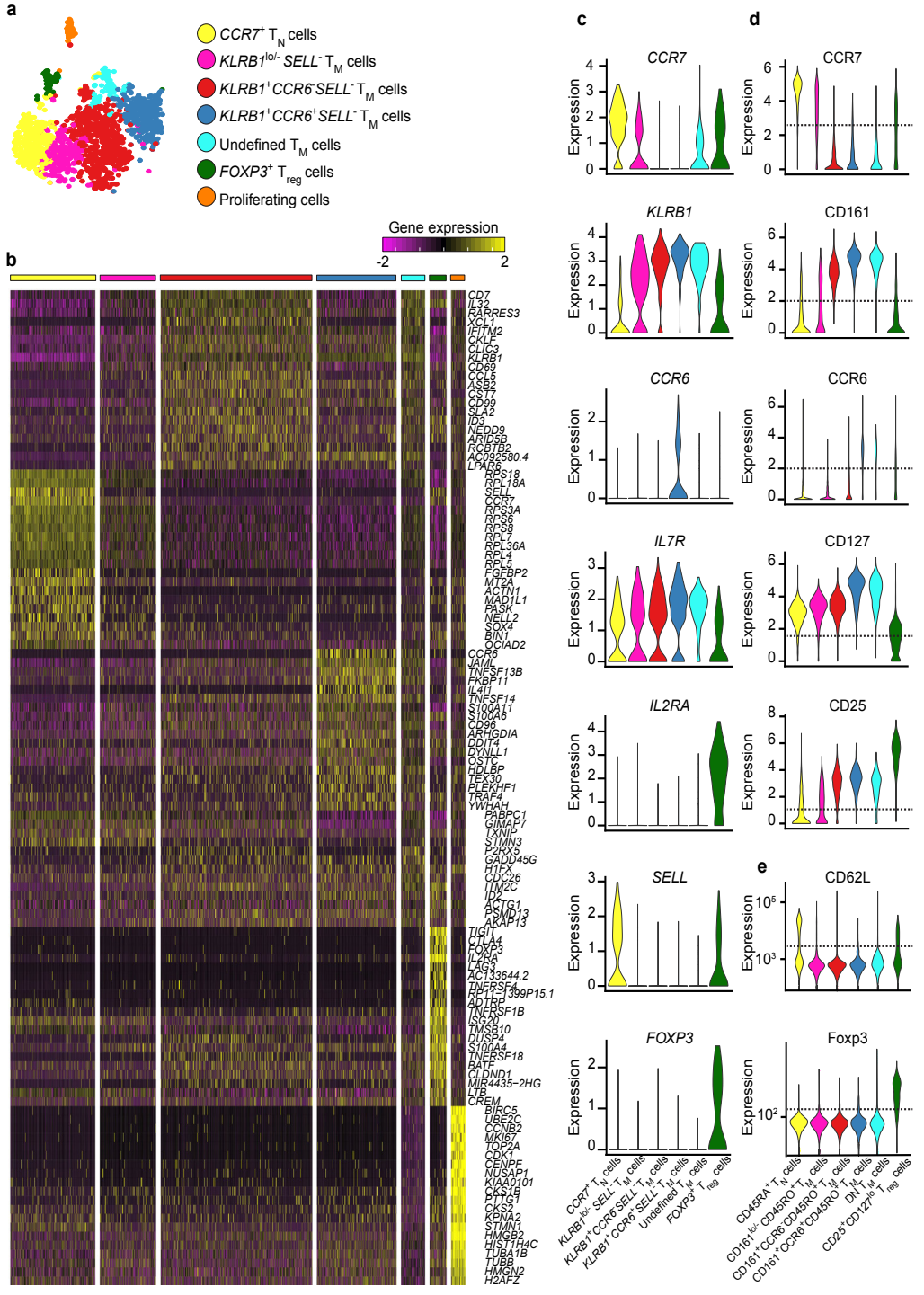
To explore the CD4+ T cell compartment in the human fetal intestine, we applied a mass cytometry panel comprising 35 antibodies (**Supplementary Table 1**) that was designed to capture the heterogeneity of the immune system to seven lamina propria samples aged 14-21 gestational weeks[12]. After data 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
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
CD161-CCR7- T M cells
Clusters
Marker expression
0
1
4.5
g
CD4+ T cells (%)
f
T N cells
CD161- T M cells
CD161+CD117- T M cells
CD161+CD117+ T M cells
DN T M cells
T reg 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, *KLRB1lo/-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), *KLRB1lo/-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

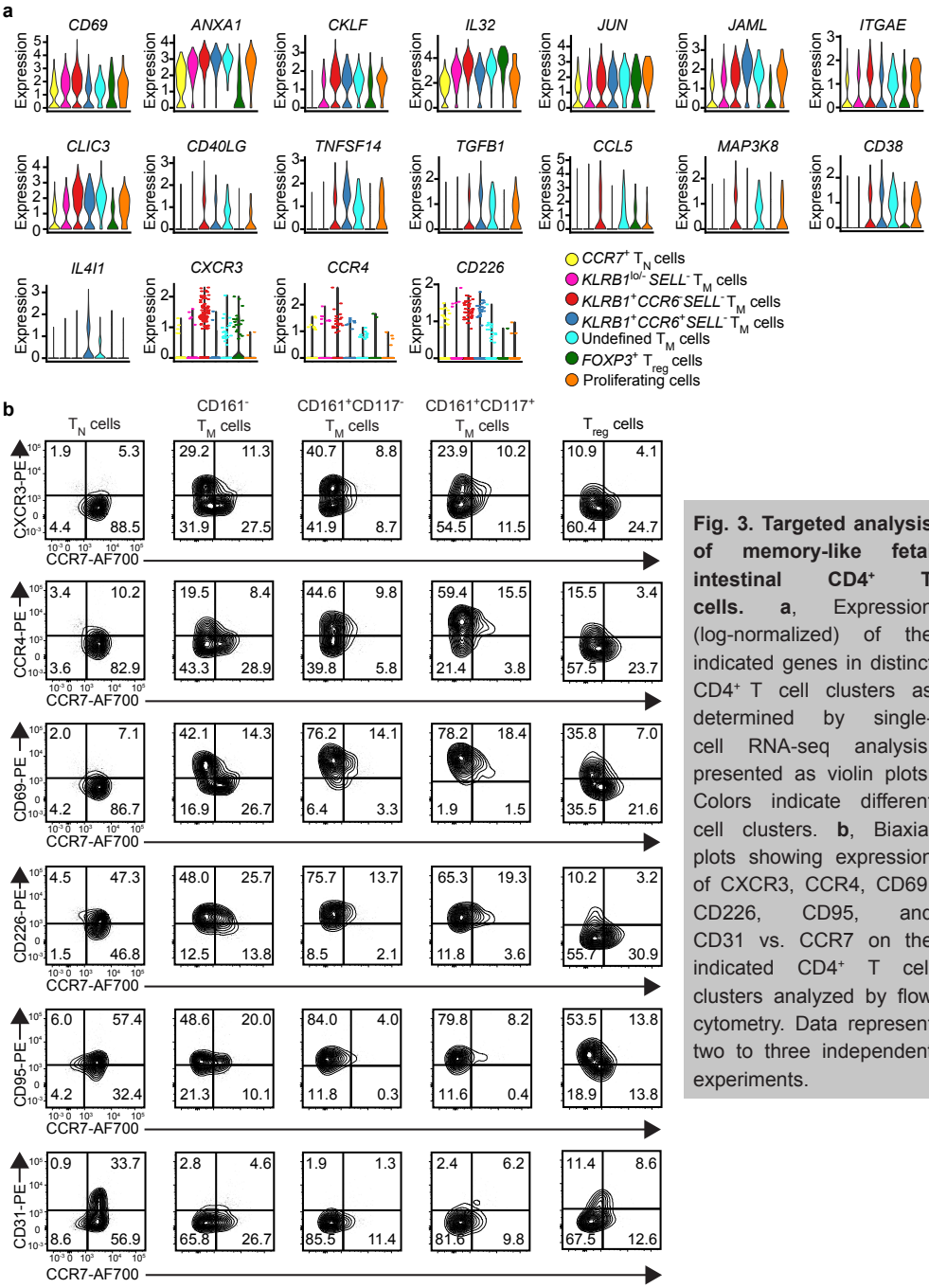


Fig. 3. Targeted analysis of memory-like fetal intestinal CD4+ T cells. **a**, Expression (log-normalized) of the indicated genes in distinct CD4+ T cell clusters as determined by single-cell RNA-seq analysis, presented as violin plots. Colors indicate different cell clusters. **b**, Biaxial plots showing 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

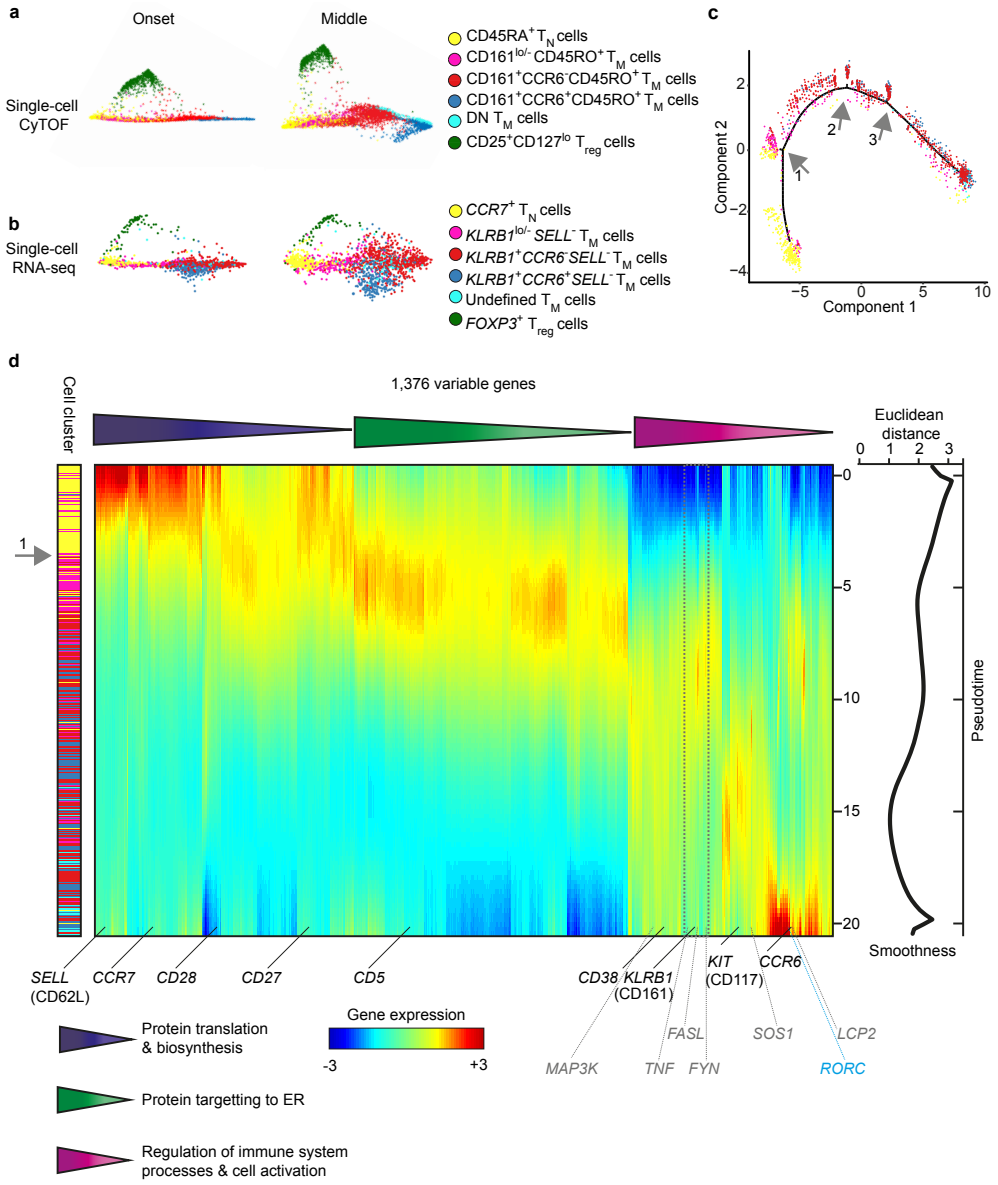


Fig. 4. Single-cell trajectory analysis of fetal intestinal CD4⁺ T cells. a–b, t-SNE embeddings of all fetal intestinal CD4⁺ T cells analyzed by (**a**) mass cytometry (n =10,436) and (**b**) single-cell RNA-seq (n = 1,743) at the onset and at the middle of the t-SNE computation. Colors indicate different 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).

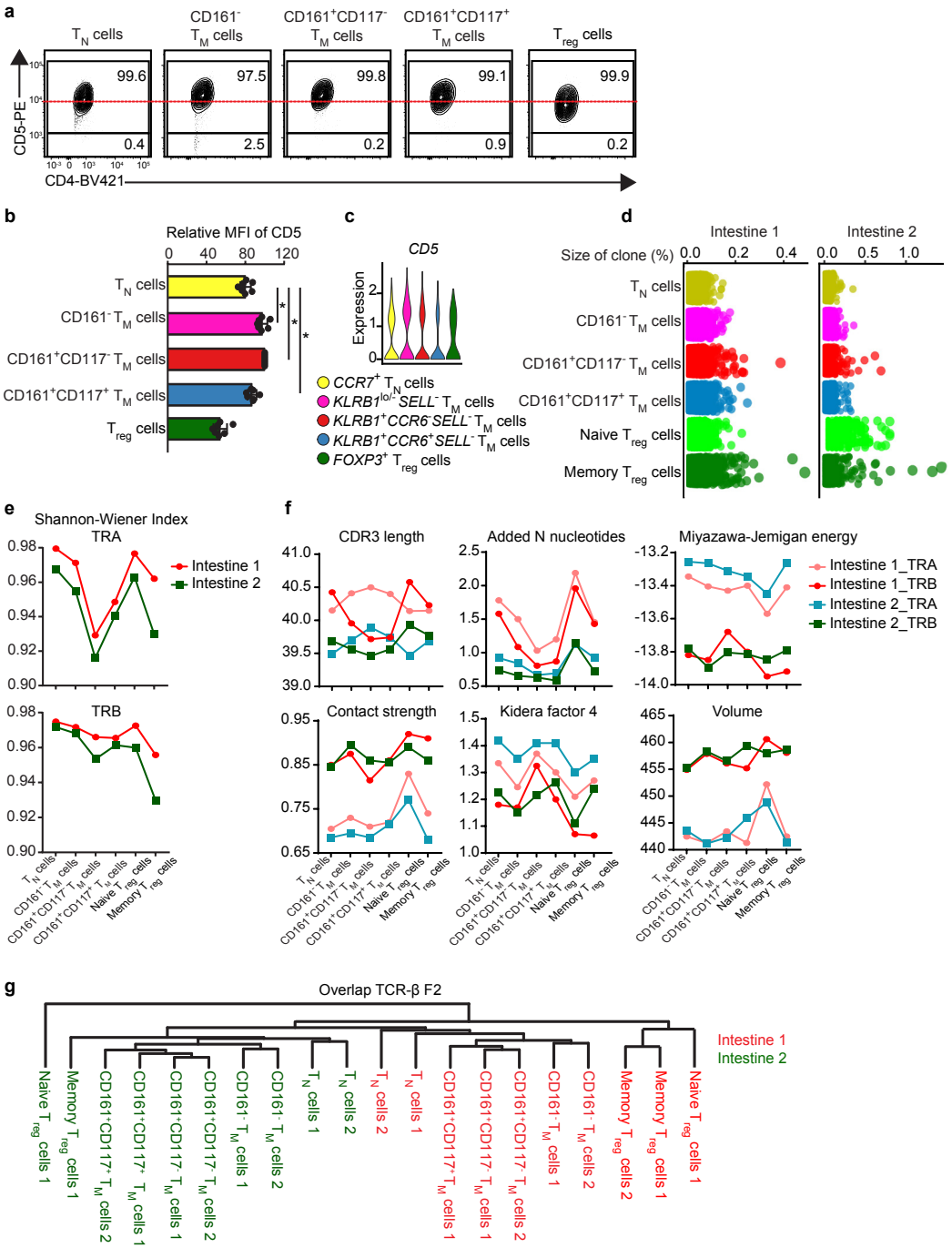
to interact physically with each other (**Supplementary Fig. 5b**). In addition, 23 genes in module 3 could be assigned to cytokine or chemokine receptor pathways, including *CCL20* and its receptor *CCR6*, the interferon receptor *IFNGR1*, TNF family members and IL-1 and IL-17 receptors (**Supplementary 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.

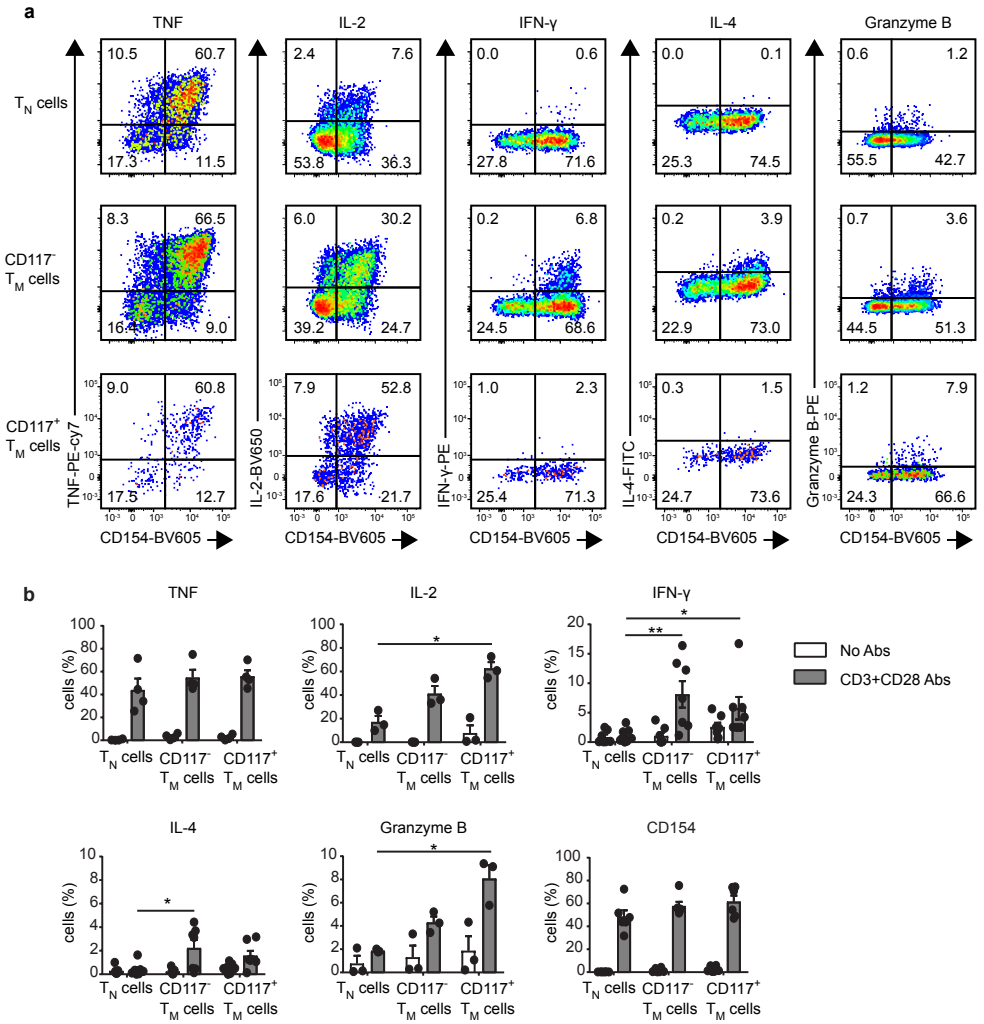


Fig. 6. Functional profiling of fetal intestinal CD4+ T cells. a–b, Purified fetal intestinal CD4+ T cells were treated with a control antibody or stimulated with anti-CD3 and anti-CD28 for 4 h. Intracellular expression of TNF, IL-2, IFN-γ, IL-4, granzyme B, and CD154 was determined for each subpopulation 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

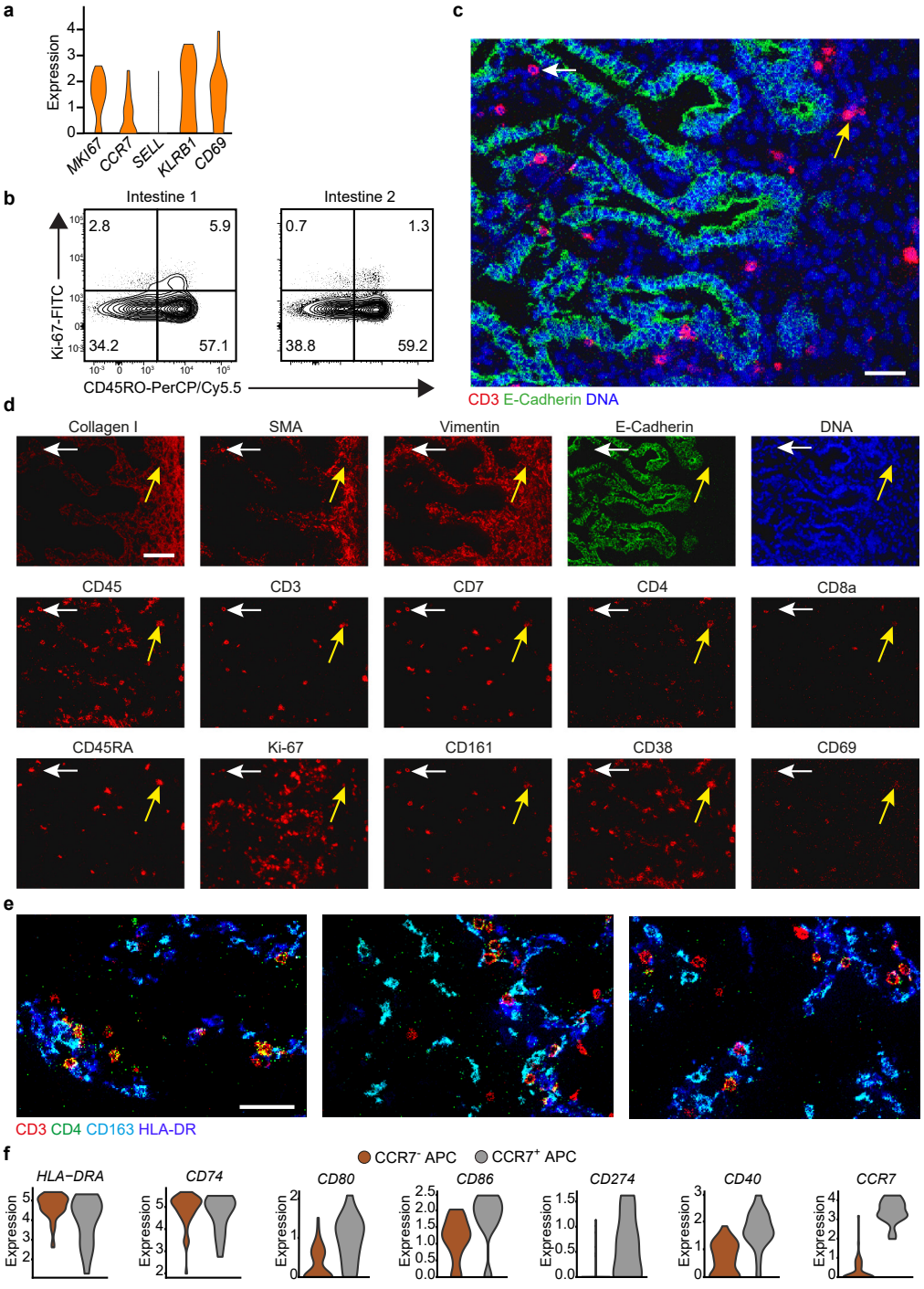
To determine the functional profiles of fetal intestinal CD4+ T cells, we flow-sorted

$CD3^+CD4^+$ 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).

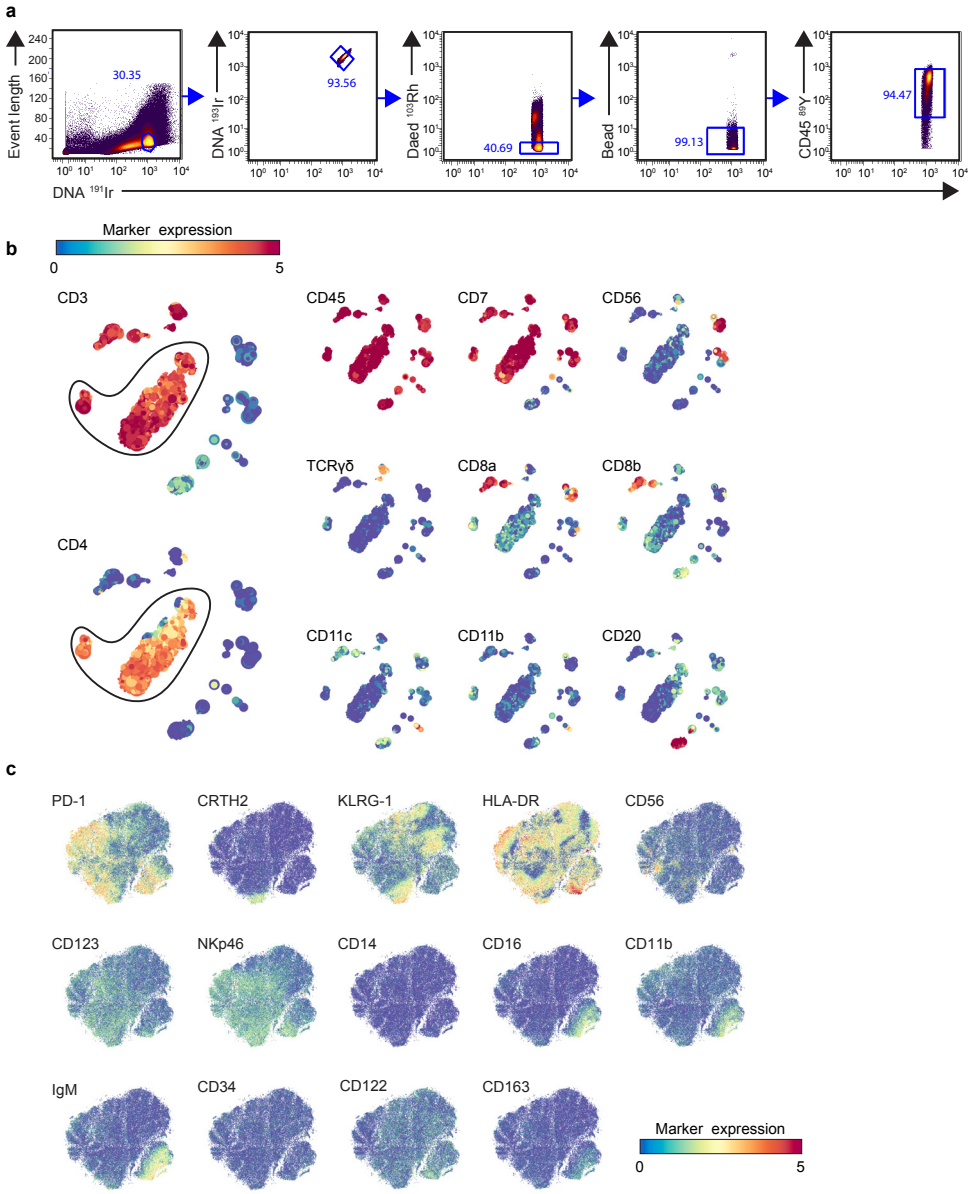
46. Quigley, M. F., Almeida, J. R., Price, D. A. & Douek, D. C. Unbiased molecular analysis of T cell receptor expression using template-switch anchored RT-PCR. *Curr. Protoc. Immunol.* **Chapter 10,** Unit10.33 (2011).

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).

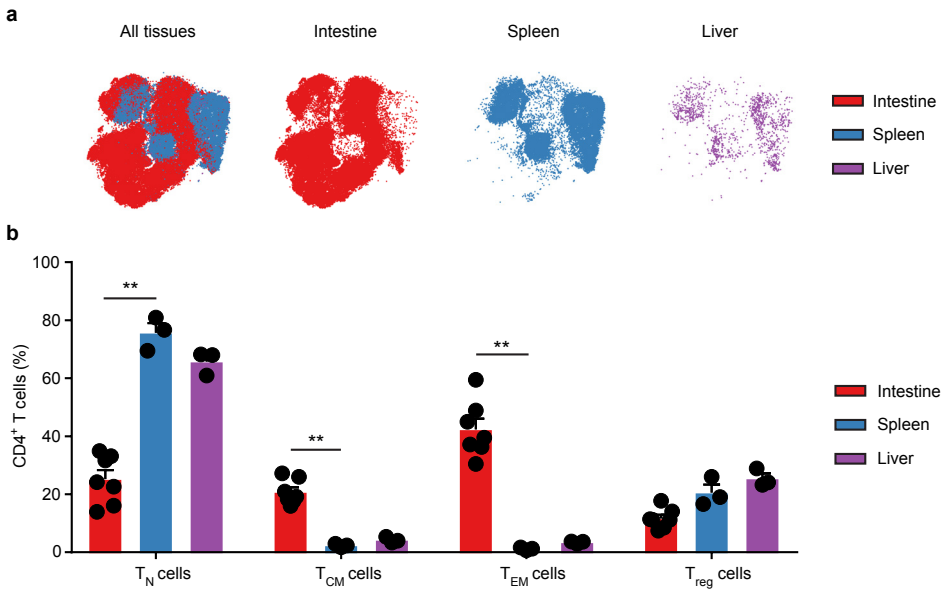
48. Shugay, M. *et al.* Towards error-free profiling of immune repertoires. *Nat. Methods* **11,** 653–655 (2014).

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

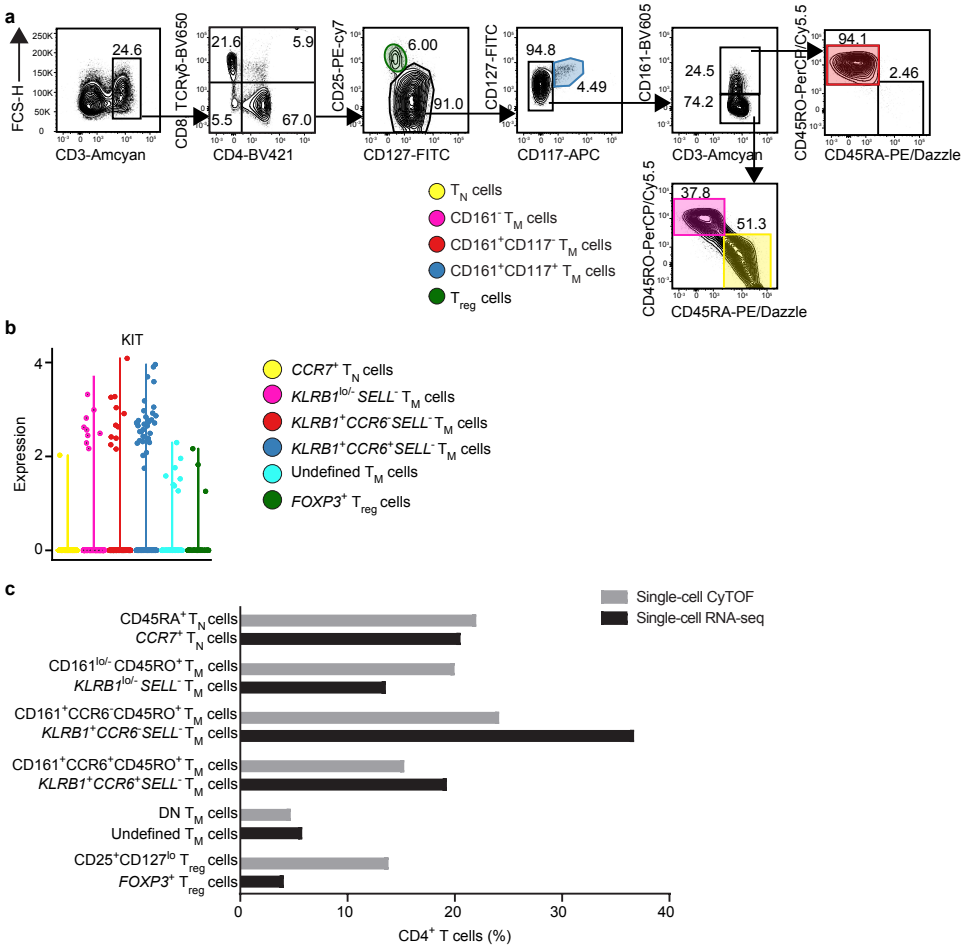
Supplemental information



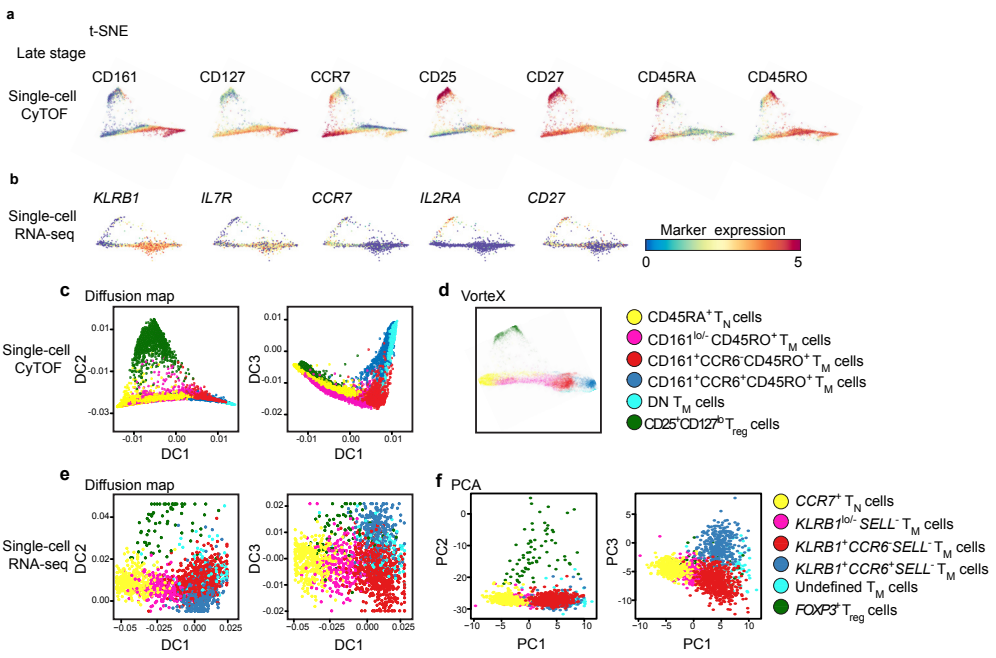
Supplementary Fig. 1. Mass cytometric analysis of fetal intestinal immune cells. (a) Biaxial plots from one fetal intestine showing the gating strategy for single, live, CD45⁺ cells analyzed by mass cytometry (n = 7). **(b)** First-level HSNE embedding of CD45⁺ immune cells (n = 224,286) derived from 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.

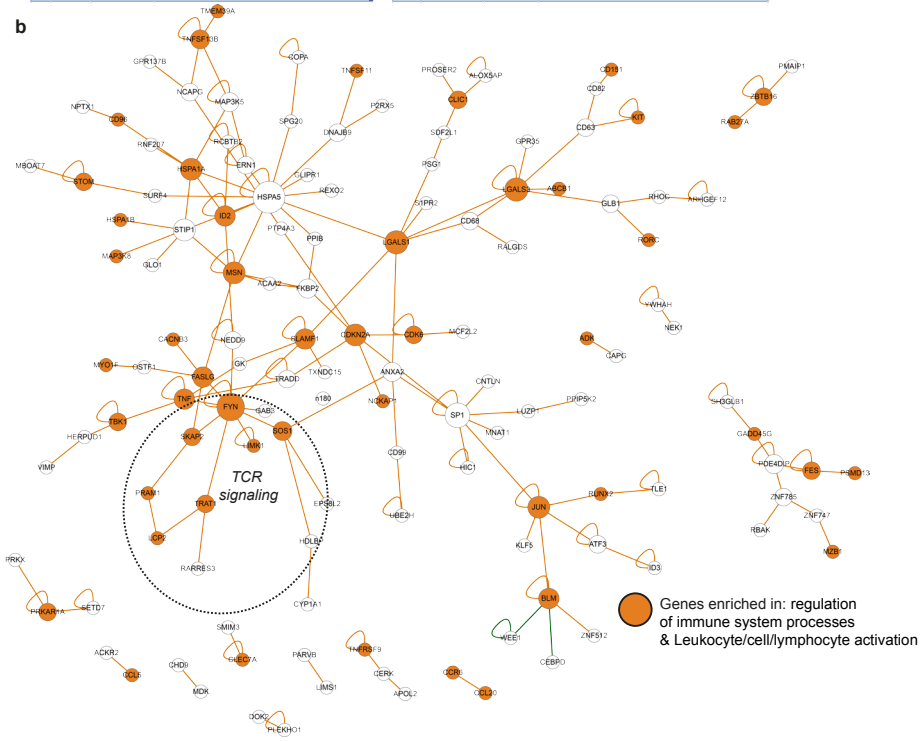
Supplementary Fig. 5. Immune-related genes in the pseudotime-dependent gene module. (**a**) Table listing 106 immune-related genes kinetically enriched in the pseudotime-dependent third gene module (**Fig. 4d**). (**b**) Interaction network of the pseudotime-dependent third gene module using Esyn. Orange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

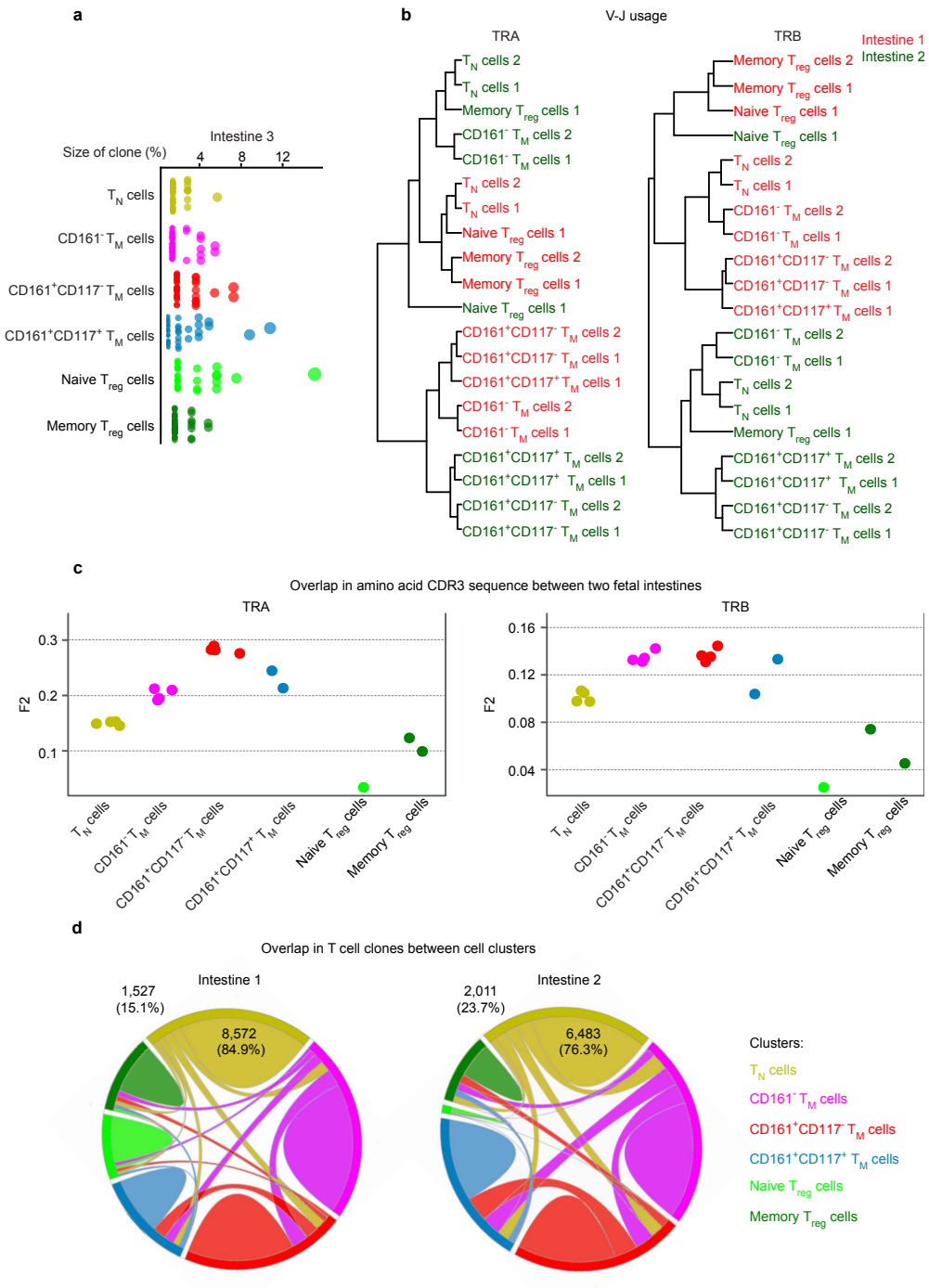
a

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

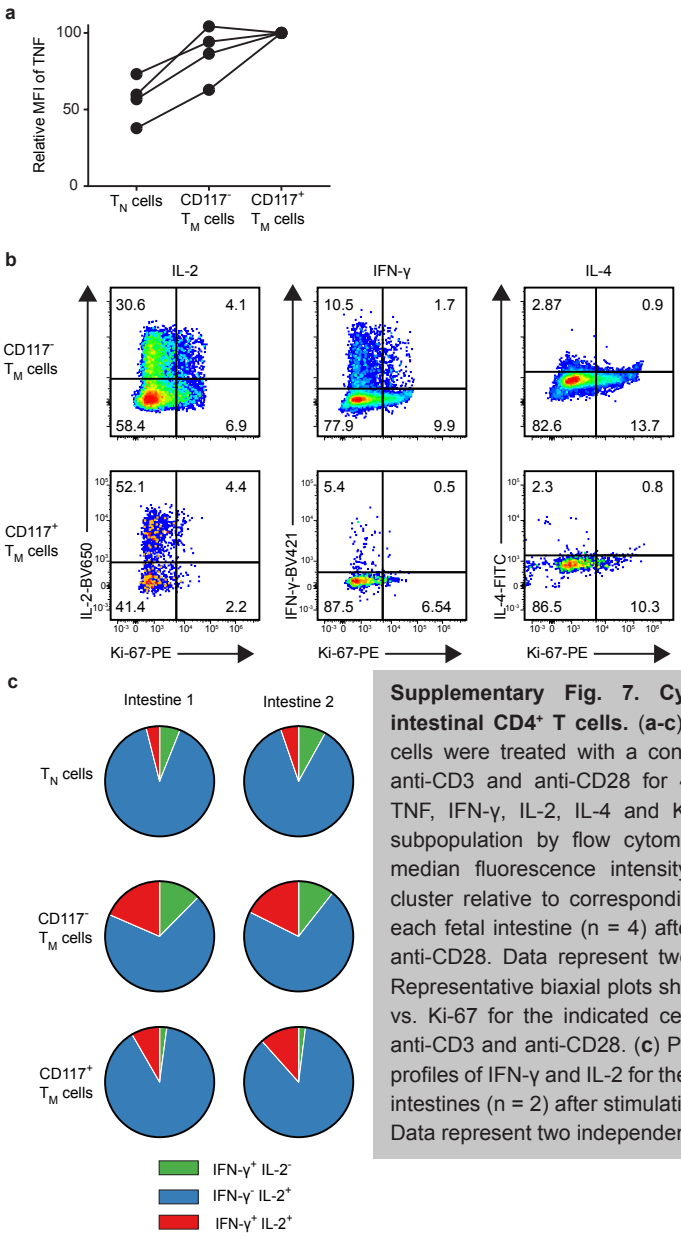
Nr	Gene Symbol	Gene Name	Original Symbol
54	4478	MSN	moesin
55	4542	MYO1F	myosin IF
56	64005	MYO1G	myosin IG
57	51237	MZB1	marginal zone B and B1 cell specific protein
58	4681	NBL1	neuroblastoma 1, DAN family BMP antagonist
59	10787	NCKAP1	NCK associated protein 1
60	114548	NLRP3	NLR family pyrin domain 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	GBP4	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	298940	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	227380	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.

Chapter 4

Early-life compartmentalization of immune cells in human fetal tissues revealed by high-dimensional mass cytometry

Na Li[1], Vincent van Unen[1], Nannan Guo[1], Tamim Abdelaal[2,3], Antonis Somarakis[4], Jeroen Eggermont[4], Ahmed Mahfouz[4], Susana M. Chuva de Sousa Lopes[5], Boudewijn P.F. Lelieveldt[3,4], Frits Koning[1]

[1] Department of Immunohematology and Blood Transfusion, Leiden University Medical Center, Leiden, Netherlands

[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]Department of Radiology, Leiden University Medical Center, Leiden, the Netherlands

[5]Department of Anatomy and Embryology, Leiden University Medical Center, Leiden, Netherlands.

Accepted by ***Frontiers in Immunology*** (2019)

Abstract

The human fetal immune system must protect the infant against the sudden exposure to a large variety of pathogens upon birth. While it is known that the fetal immune system develops in sequential waves, relatively little is known about the composition of the innate and adaptive immune system in the tissues. Here, we applied high-dimensional mass cytometry to profile the immune system in human fetal liver, spleen and intestine. With Hierarchical Stochastic Neighbour Embedding (HSNE) we distinguished 177 distinct immune cell clusters, including both previously identified and novel cell clusters. PCA analysis indicated substantial differences between the composition of the immune system in the different organs. Through dual t-SNE we identified tissue-specific cell clusters, which were found both in the innate and adaptive compartment. To determine the spatial location of tissue-specific subsets we developed a 31-antibody panel to reveal both the immune compartment and surrounding stromal elements through analysis of snap-frozen tissue samples with imaging mass cytometry. Imaging mass cytometry reconstructed the tissue architecture and allowed both the characterization and determination of the location of the various immune cell clusters within the tissue context. Moreover, it further underpinned the distinctness of the immune system in the tissues. Thus, our results provide evidence for early compartmentalization of the adaptive and innate immune compartment in fetal spleen, liver and intestine. Together, our data provide a unique and comprehensive overview of the composition and organization of the human fetal immune system in several tissues.

Introduction

The notion of phenotypical and functional differences between the fetal and adult immune system has been widely accepted. The human fetal immune system has to maintain tolerance towards the semi-allogeneic environment in utero while being prepared for the massive exposure to massive foreign antigens directly after birth [1, 2]. The ontogeny of the immune system occurs in sequential waves during gestation. Fetal hematopoiesis is initiated in the yolk sac around day 16 of the development, then transits to the fetal liver at 6 weeks until 22 weeks gestational age, where the progenitors give rise to both lymphoid and myeloid cells [3]. T cells have been identified as early as 10 weeks of gestation while $Foxp3^{+}CD4^{+}$ regulatory T (Treg) cells, whose generation is mainly driven by maternal alloantigens, have also been observed in different fetal tissues [4]. Furthermore, it has been shown that human fetal dendritic cells in spleen, skin, thymus and lung promote prenatal T-cell immune suppression [5]. Interestingly, several studies have provided evidence for the existence of memory-like T (Tm) cells in fetal spleen [6], skin [7], intestine [8, 9] and cord blood [10], which produce pro-inflammatory cytokines such as IFN-γ and TNF-α, suggesting functional maturation of T cells in utero. In line with the discovery of pro-inflammatory T cells, mucosa-associated invariant T cells as well as natural killer (NK) cells and innate lymphoid cells have also been found to be present in the fetal intestine [11, 12]. A potential link between the composition of the prenatal immune system and disease later in life has been proposed [9, 13]. Thus, the fetal immune system has both pro-inflammatory and immune suppressive capacity.

Most investigations of the human fetal immune system are based on umbilical cord blood collected at birth. However, the representation of cord blood has recently been questioned as cord blood samples were heterogeneous without clearly shared patterns in cell and plasma protein composition [14]. Although in recent years several reports [5, 8] have studied the fetal immune system in tissues, due to the scarcity of human fetal tissues and technique limitations, a system-wide and detailed characterization of the human fetal immune system is currently lacking. Mass cytometry (cytometry by time-of-flight; CyTOF) now offers the opportunity to analyze the heterogeneity of the human immune system in an unbiased and data-driven fashion by the simultaneous measurement of over 40 unique cellular markers at the single-cell level with unprecedented resolution [15]. As traditional analysis approaches for flow cytometry are not compatible with high-dimensional mass cytometry datasets, dimensionality reduction-based approaches such as t-stochastic neighbor embedding (t-SNE) [16] have been widely used as they allow users to analyze all the markers concurrently in an unbiased manner. Moreover,

hierarchical SNE (HSNE) has removed the scalability limitation of t-SNE, allowing the analysis of tens of millions of cells at single-cell resolution [17].

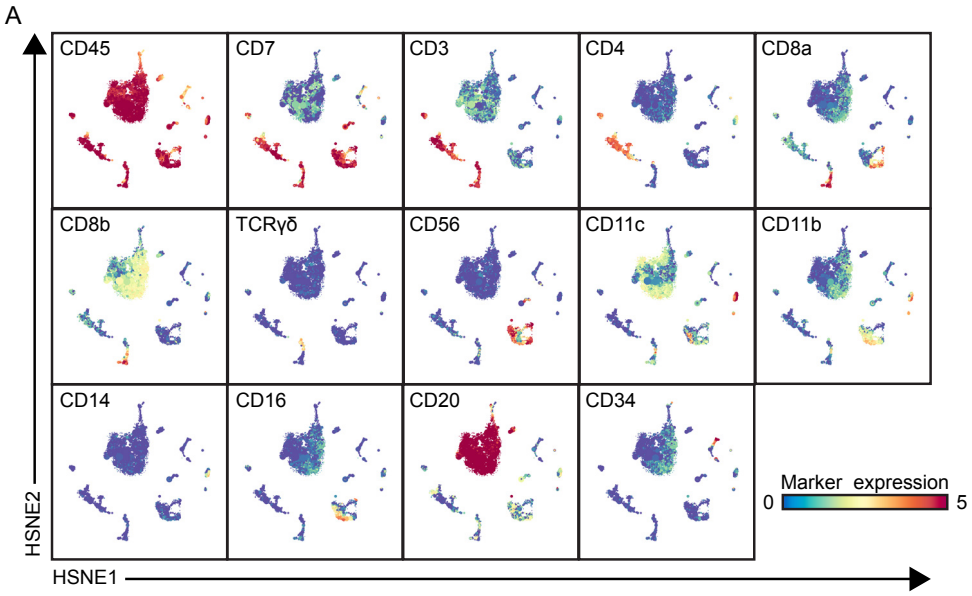
Here, we applied suspension mass cytometry to study the complexity and heterogeneity of the immune system in the human fetal intestine, liver and spleen and confirmed the existence of previously identified subsets through the unsupervised HSNE analysis. In addition, we provide evidence for the existence of previously unrecognized distinct cell clusters. Besides the heterogeneity within each tissue, our data further reveals clear tissue-specific signatures in both the innate and adaptive immune compartment as early as week 16 of gestation. Finally, we employed imaging mass cytometry to reconstruct the tissue architecture and characterize and determine the location of the immune cell clusters within the tissue context, results that underpinned the compartmentalization of the immune system in the tissues. Together these data provide a comprehensive and valuable resource for understanding the fetal immune development and linking prenatal immunity with immunity after birth.

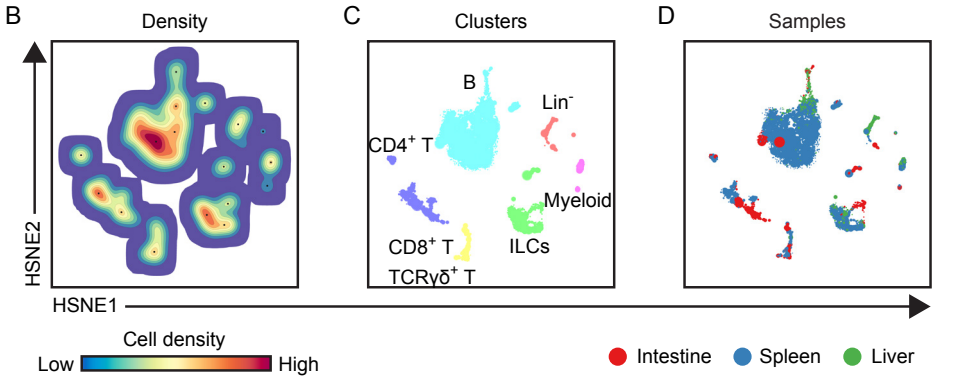
Results

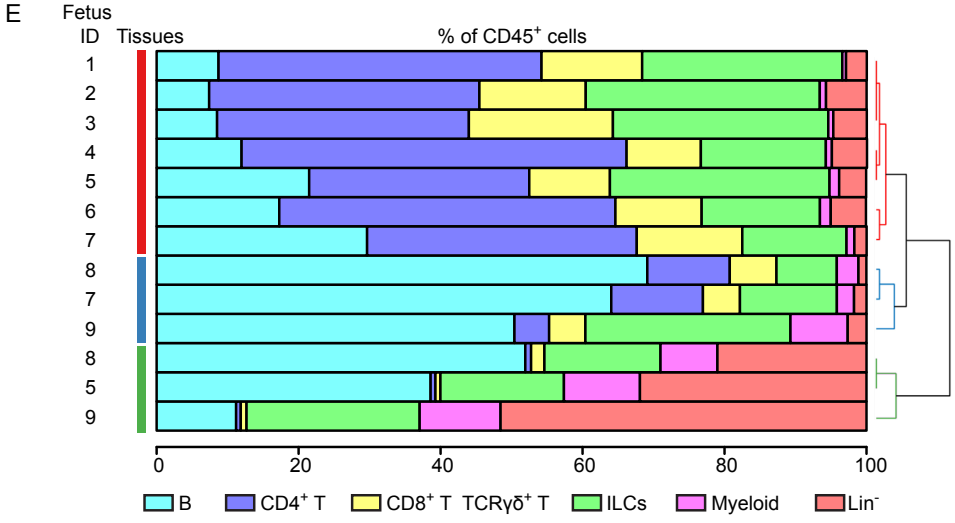
Identification of major immune lineages across human fetal tissues

To explore the immune system in the human fetus, we employed a previously described CyTOF panel (**Table S1**) consisting of 35-metal isotope-tagged monoclonal antibodies [18] designed to identify the major immune lineages (B cells, $CD4^+$ T, $CD8^+$ T, γδ T cells, innate lymphoid cells (ILCs), and myeloid cells) and determine the heterogeneity within these lineages. For this purpose, the panel consisted of lineage markers, markers specific for cell differentiation, activation, trafficking and function. With this panel, single-cell suspensions from fetal

Figure 1 HSNE analysis distinguishes major immune lineages across human fetal tissues. (A) HSNE embedding of 1.6 million immune cells derived from fetal intestines (N=7, 0.3 x 10^6 cells), fetal spleens (N=3, 1.1 x 10^6 cells), and fetal livers (N=3, 0.2 x 10^6 cells) at the overview level. Each dot represents an HSNE landmark and the size of the landmark is proportional to the number of cells that each landmark represents. Colors of the landmarks represent ArcSinh5-transformed expression values of the indicated markers. (B) A density map showing the local probability density of the embedded cells where black dots display the centroids of identified clusters using GMS clustering. (C) A HSNE plot showing main immune lineage cluster partitions in different colors. (D) HSNE embedding as shown in panel A. Colors represent different tissues. (E) The composition of major immune lineage clusters for $CD45^+$ cells in the individual fetal tissues is represented in horizontal bars where the colored segment lengths represent the proportion of cells as a percentage of $CD45^+$ cells in the sample. The dendrogram shows the hierarchical clustering of samples. Colors represent the different tissues as shown in panel D. Numbers indicate fetus ID.


A
CD45
CD7
CD3
CD4
CD8a
CD8b
TCRγδ
CD56
CD11c
CD11b
CD14
CD16
CD20
CD34
Marker expression
0
5
HSNE2
HSNE1


B
Density
HSNE2
HSNE1
Cell density
Low
High
C
Clusters
B
Lin-
CD4+ T
Myeloid
CD8+ T
TCRγδ+ T
ILCs
D
Samples
Intestine
Spleen
Liver


E
Fetus ID
Tissues
% of CD45+ cells
1
2
3
4
5
6
7
8
7
9
8
5
9
0
20
40
60
80
100
B
CD4+ T
CD8+ T
TCRγδ+ T
ILCs
Myeloid
Lin-

intestines (N=7), fetal spleens (N=3) and fetal livers (N=3) (**Table S2**) were analyzed. Single, live $CD45^+$ cells were distinguished by event length, DNA stains, and CD45 antibody stains (**Figure S1**). In the liver, but not in the spleen and intestine, three distinct subpopulations were observed based on different CD45 and DNA stainings (**Figure S1**). Here, $CD45^{hi}DNA^{low}$ cells represent lymphoid cells, while $CD45^{hi}DNA^{hi}$ and $CD45^{low}DNA^{hi}$ cells correspond to myeloid and $CD34^+$ precursor cells, respectively (results not shown). All antibodies displayed a clear separation between antibody-negative and -positive cells as described previously [18].

To determine the major immune lineages, we pooled the data (1.6×10^6 $CD45^+$ cells) derived from seven fetal intestines (39,357 cells per intestine), three fetal spleens (365,653 cells per spleen) and three fetal livers (69,928 cells per liver) and performed a 3-level HSNE analysis in Cytosplore^{+HSNE} [17] from a global overview down to the single cell level. Here, HSNE landmarks depicted the global composition of the immune system (**Figure 1**). Based on the marker expression profiles (**Figure 1A**) and density features of the embedded cells (**Figure 1B**), we identified 6 phenotypically distinct major lineage clusters at the overview level, namely $CD20^+$ B cells, $CD3^+CD4^+$ T cells, $CD3^+CD8^+$ T cells and $CD3^+TCR\gamma\delta^+$ T cells, $CD3^-CD20^-CD11c^-CD7^+$ ILCs, $CD11c^+$ myeloid cells and a $CD3^-CD20^-CD11c^-CD7^{+/-}$ lineage-negative (Lin^-) cell cluster (**Figure 1C**). In addition, the global differences in immune composition across tissues were revealed by visualizing the tissue-origin of the cells (**Figure 1D**). Next, we quantified the relative frequencies of major immune lineage clusters in each fetal tissue sample, showing that B cells were more abundant in the fetal spleen and fetal liver while Lin^- cells were most profound in the fetal liver. $CD4^+$ T, $CD8^+$ T and γδ T cells together comprised more than 50% of immune cells in the fetal intestine but were typically lower in the fetal liver. There was no significant difference in the distribution of ILCs across fetal tissues. The frequencies of myeloid cells were low in all fetal tissues, however, these cells were relatively more abundant in fetal liver (**Figure 1E**). Importantly, unbiased hierarchical clustering of cell frequencies grouped the tissue samples in a tissue-specific manner (**Figure 1E)**, indicating that despite differences between the samples the composition of the major immune lineage of each fetal tissue was relatively constant and tissue-specific.

Together, these global analyses revealed that the major immune lineages could be readily identified in each fetal tissue, and that the composition of these lineages differs between the examined human fetal tissues.

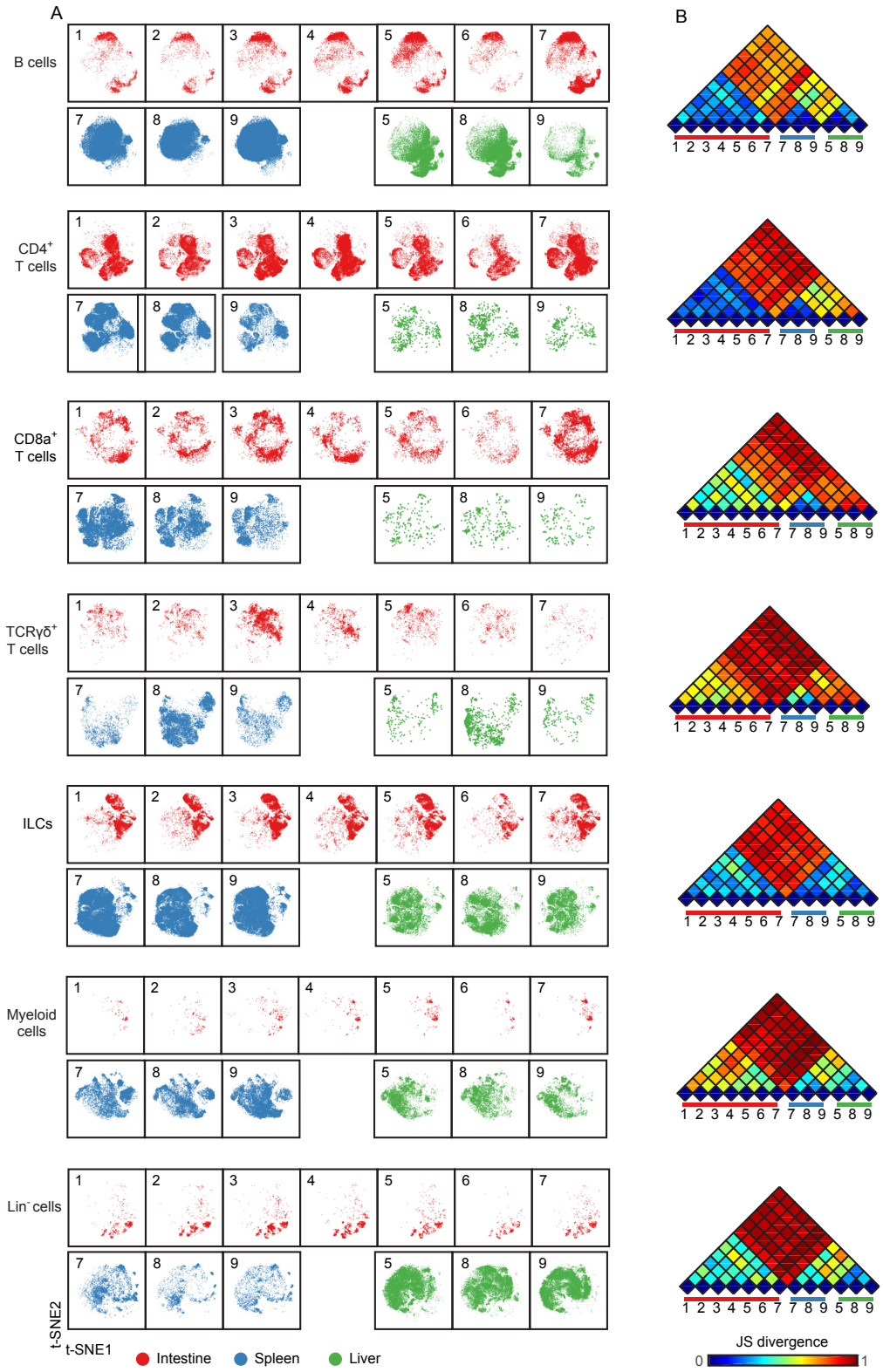

A
B
B cells
CD4+ T cells
CD8a+ T cells
TCRγδ+ T cells
ILCs
Myeloid cells
Lin- cells
t-SNE2
t-SNE1
Intestine
Spleen
Liver
1 2 3 4 5 6 7 7 8 9 5 8 9
JS divergence
0
1

Figure 2 Cellular signatures reveal the immune cell compartmentalization across human fetal tissues. (A) A collective t-SNE was performed on each major immune lineage individually and stratified for samples. The plots are showing in total 1.5 x 10^5 B cells, 2.3 x 10^5 $CD4^+$ T cells, 6.9 x 10^4 $CD8^+$ T cells, 3.6 x 10^4 γδ T cells, 2.6 x 10^5 ILCs, 6.4 x 10^4 myeloid cells and 8.7 x 10^4 Lin^- cells. Colors represent different tissues. Numbers indicate fetus ID. (B) The similarity of a pair of t-SNE plots is shown by Jensen-Shannon (JS) divergence within the major immune lineage. A higher JS divergence value indicates higher dissimilarity between pairwise t-SNE plots.

Visualization of cellular signatures across human fetal tissues

To compare immune cells between all samples and between the tissues, we next selected the major immune lineage clusters individually and embedded the clusters at the single-cell level to deduce the cellular signatures (**Figure 2**) [19]. The results demonstrate that the intestine, spleen and liver displayed a distinct cellular signature for most of the seven major immune lineages. We observed especially highly distinct differences between the intestinal samples compared to the spleen and liver samples (**Figure 2A**). Also, the results underscore that samples from the same tissue type have a highly similar cellular signature (F**igure 2A**). We next applied the Jensen-Shannon (JS) divergence method to quantify the similarities and differences between pairs of t-SNE maps (**Figure 2B**). For all major immune lineages, the JS divergence was highest between the intestinal samples and the spleen and liver samples, while considerable differences in JS divergence were also found between the spleen and the liver samples, except for the ILC compartment. In general, the composition of the B cells, $CD4^+$ T cells and ILCs was more similar within a tissue group while more variation was present among $CD8^+$ T cells, γδ T cells, and myeloid cells (**Figure 2B**). Together, the visualization of cellular signatures provided further evidence for the distinct composition of immune system in the fetal tissues.

High-dimensional analysis reveals unprecedented immune heterogeneity in human fetal tissues

To extend the analysis we next selected every main immune lineage individually and embedded them at the second level (B cells) or single-cell data level to identify the phenotypically distinct clusters, here illustrated for the $CD8^+$ T cell compartment (**Figure 3**). First, the cluster including $CD8^+$ T and γδ T cells (105,211 cells) were selected at the overview level and embedded in the second level of the HSNE analysis (**Figure 3A, B**), followed by selection of the $CD8^+$ T cells (69,269 cells) to zoom-in further (**Figure 3B**), revealing more single-cell details (**Figure 3C**). Based on the density features of the t-SNE-embedded cells, we identified 20 distinct $CD8^+$ T cell clusters (**Figure 3C**), each defined by a unique maker expression profile (**Figure 3D**). After hierarchical clustering of the

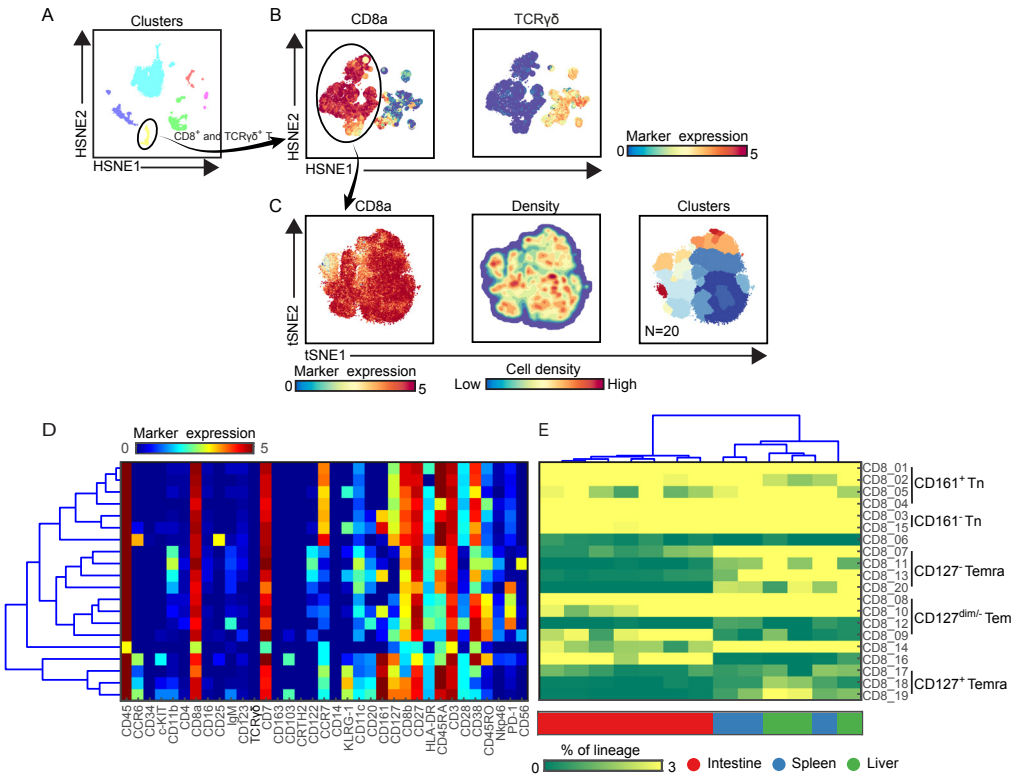


Figure 3 Identification of phenotypically distinct clusters in the CD8⁺ T cell compartment across human fetal tissues. (A) Global level of HSNE embedding of 1.6 million immune cells as shown in Figure 1C showing major immune lineage cluster partitions in different colors. (B) Second level of HSNE embedding of CD8⁺ T and γδ T cells (1.1 x 10^5 cells) derived from fetal intestines (N =7), fetal spleens (N=3) and fetal livers (N=3). Colors represent the ArcSinh5-transformed expression values of CD8a and TCR γδ. (C) A t-SNE embedding of collective CD8⁺ T cells (6.9 x 10^4 cells) derived from fetal intestines (N =7), fetal spleens (N=3) and fetal livers (N=3). Colors represent the ArcSinh5-transformed expression values of CD8a (left panel). A density map showing the local probability density of the embedded cells (middle panel). Colors represent cluster partitions (right panel). (D) Heatmap (blue-to-red scale) showing the median marker expression values of the clusters identified in panel C and hierarchical clustering thereof. (E) Heatmap (green-to-yellow scale) showing the corresponding cell frequencies of identified clusters of total CD8⁺ T cells in each sample. The dendrogram shows the hierarchical clustering of samples. Colors represent different fetal tissues. Tn = naive T cells, Tem = effector memory T cells and Temra = terminally differentiated T cells.

heatmap, these clusters grouped into five main metaclusters (**Figure 3D,E**). CD45RA⁺CCR7⁺ naive T (Tn) cells were classified into two groups based on CD161 expression while CD45RA⁻CCR7⁺ central memory T (Tcm) and CD45RA⁺CCR7⁻ terminally differentiated T (Temra) cells were detected in the CD127$^{dim/-}$ compartment. In addition, a group of CD127⁺CD161^{high}CCR6⁺KLRG-1⁺ Temra cells clustered separately from the other groups. We next quantified the cell frequencies

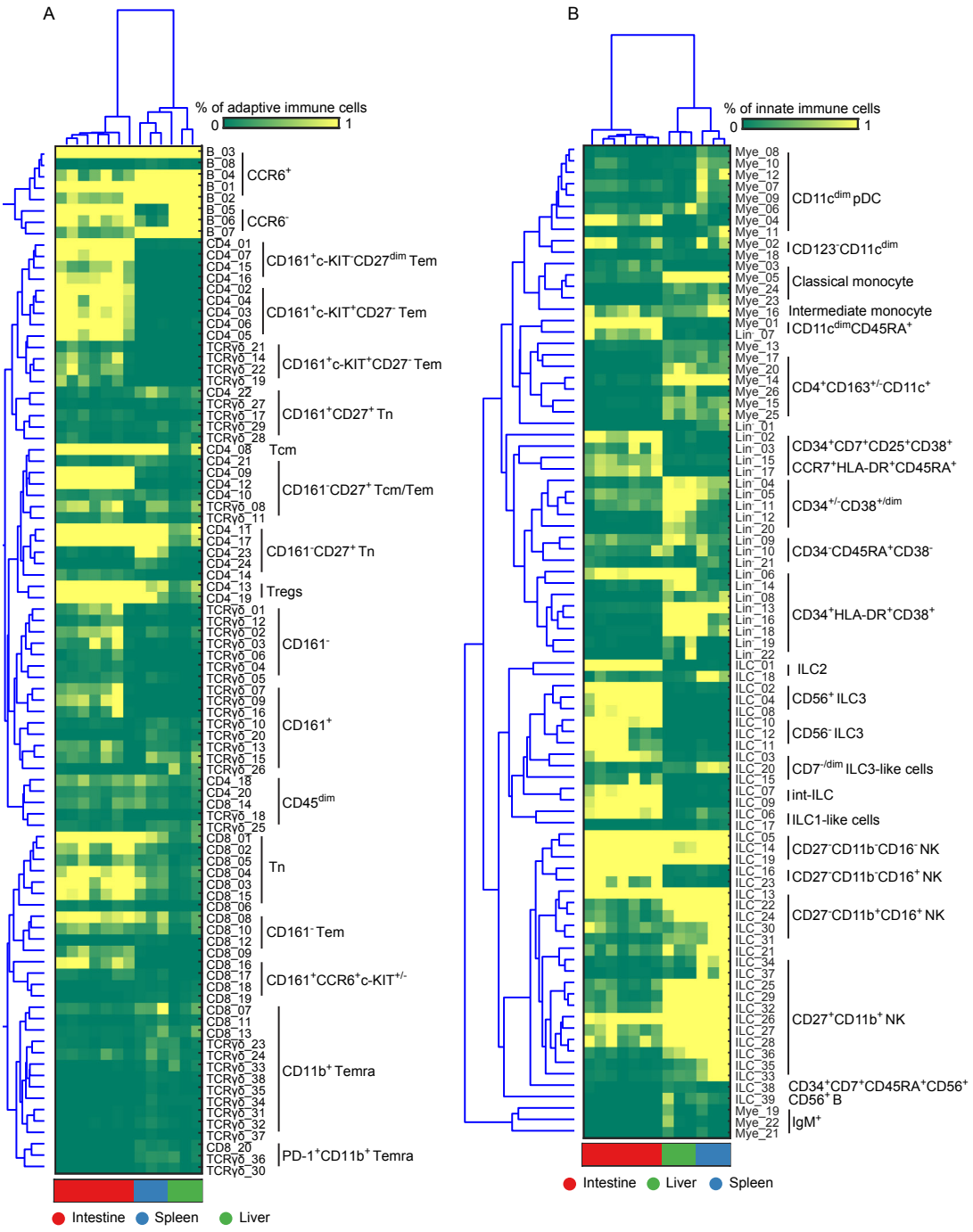


Figure 4 Immune system landscapes visualize cell cluster composition across human fetal tissues. (A and B) Heatmap showing the cell frequencies of 90 adaptive and 87 innate immune clusters identified in Figure S2 and S3 combined with hierarchical clustering of samples (Top) and phenotype of clusters (Left). Colors represent different fetal tissues as indicated. Tem = effector memory T cells, Tcm = central memory T cells, Temra = terminally differentiated T cells, and pDC = plasmacytoid dendritic cells.of each

of each cluster per sample and performed the hierarchical clustering of different samples based on the cell frequencies. In line with previous studies [20], several clusters including Tn cells (CD8_01-05, 15) and Tem cells (CD8_08, CD8_10) were present in all tissues (**Figure 3E**). $CD127^{-}CD161^{+/-}$ Temra (CD8_07, CD8_11, CD8_13 and CD8_20) mainly existed in the spleen and liver whereas $PD\text{-}1^{-}$ Tem (CD8_09) were mainly abundant in the intestine. Hierarchical clustering revealed that all the intestinal samples clustered together whereas the spleen and liver samples clustered intermixed with each other.

By applying this approach to all 7 major immune lineage clusters, we identified 177 phenotypically distinct clusters (8 B cell clusters, 24 $CD4^{+}$ T cell clusters, 20 $CD8^{+}$ T cell cluster, 38 γδ T cell cluster (**Figure S2A-C, Figure 4A**), 39 ILC clusters, 26 myeloid cell clusters and 22 Lin^{-} clusters (**Figure S3A,B, Figure 4B**). Consistent with previous reports [2], $IgM^{+}HLA\text{-}DR^{+}$ B cells were identified in all three tissues while the plasmacytoid dendritic cells (pDC)-like cells and monocytes were mainly identified in the intestine and non-intestine, respectively. In addition to Tn cells, we also identified several clusters of Tm cells, which were more prominent in the $CD4^{+}$ T cell compartment in the fetal intestine, as reported [9]. Strikingly, some of these Tm cells expressed a higher level of CD127, CD161, CCR6 and c-KIT, the latter a marker typically associated with progenitor cell types. Similarly, we identified a rare $c\text{-}KIT^{+}$ ($CD117^{+}$) $CD8^{+}$ T cell population (CD8_16, < 1% of $CD45^{+}$ cells) that is almost exclusively present in the intestine. As reported [21], helper-ILCs were more dominant in the fetal intestine than the fetal spleen and liver, where the majority of ILCs were different types of NK cells such as $CD27^{-}CD11b^{+}$, $CD27^{+}CD11b^{+}$ NK clusters. Moreover, $CD34^{+}CD38^{+}$, $CD34^{+}CD38^{dim}$, $CD34^{+}CD38^{-}$ precursor cell clusters and several previously unknown precursor-like cell clusters were mainly identified in the fetal liver (e.g., Lin^{-}_02, Lin^{-}_03, Lin^{-}_15, Lin^{-}_17) (**Figure 4**), an important organ for fetal hematopoiesis. Unbiased hierarchical clustering of cell frequencies for each sample revealed a clear distinction between the fetal intestine and the other two tissues in both the adaptive and innate immune compartment (**Figure 4**). The separation between fetal spleen and liver was readily observed in the innate immune compartment, due to the differential abundance of $CD34^{+}$ cells and certain types of ILCs, such as ILC2 and $CD7^{dim}$ ILC3-like cells (**Figure 4B**), but not in the adaptive immune compartment.

Altogether, our data reveal that there is far greater heterogeneity of the immune system in human fetal tissues than previously appreciated and provide evidence for the existence of previously unknown immune cell types.

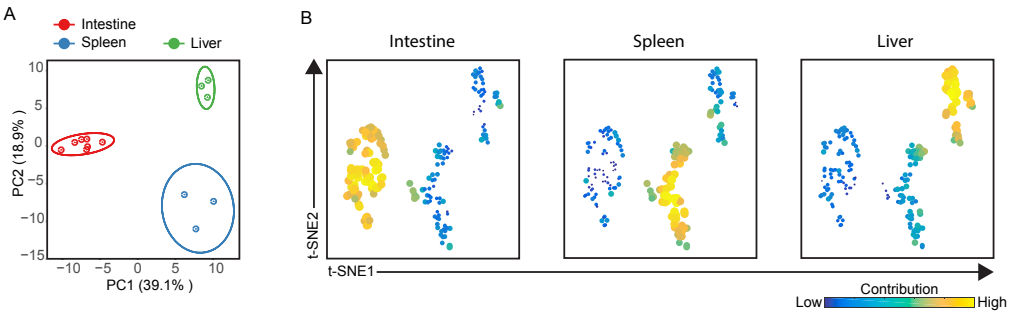
An integrated system-wide analysis of the immune system reveals early immune compartmentation across fetal tissues

In order to investigate the entire immune profiles across fetal tissues, a principal component analysis (PCA) was performed on the samples based on the cluster frequency values of $CD45^{+/dim}$ cells. The samples were clearly separated from each other based on the tissue origin, especially the intestines and the other two organs by principal component 1 (explaining 39.1 % variance) (**Figure 5A)**. In consistency with the cellular signatures (**Figure 2**), the variance of immune composition within spleens was higher than that within intestines and livers (**Figure 5A**).

To reveal which cell clusters were most strongly associated with the tissue-specificity, we performed a t-SNE analysis on 177 clusters based on the frequency values, visualizing networks of cell clusters, which determined the tissue-associated patterns (**Figure 5B)**. Moreover, the dual-tSNE identified the top ten ranked clusters that contributed to the sample clustering patterns (**Figure 5C**). This analysis yielded three distinct networks of cell clusters corresponding to the three tissue types (**Figure 5B**). Several distinct hematopoietic progenitor clusters (Lin^-_16, ILC_38, Lin^-_04, Lin^-_18, Lin^-_20 and Lin^-_12) and four myeloid populations (Mye_13, Mye_14, Mye_15 and Mye_25) were specifically found in the fetal liver. For the fetal intestines, the main types of contributing cells were lymphoid precursors (Lin^-_15 and Lin^-_17) and $CD4^+$ Tm clusters (CD4_01, CD4_08,CD4_09, CD4_06, CD4_04, CD4_05), including three previously unidentified c-KIT^+ clusters, while for the spleen a diverse pool of $CD8^+$ T cells, B cells and myeloid cells contributed to the tissue-specific signature (**Figure 5C**). Together, the integrated system-wide analysis of the immune system reveals a unique immune cell landscape and early-life immune compartmentation across human fetal tissues, suggesting different immune responses in situ during human fetal development.

Imaging mass cytometry reveals the tissue-specific signatures in situ.

To extend the understanding of the fetal immune system, we applied an imaging mass cytometry panel comprising 31 antibodies to determine the spatial distribution of the immune and stromal cells in the tissues in situ. This panel contained markers for visualizing the tissue architecture such as E-cadherin (epithelium), CD31 (endothelial cells), D2-40 (lymphatic endothelium), vimentin (intermediate filament), smooth muscle actin (SMA) and collagen I (extracelluar matrix), as well as markers to identify various immune cell subsets such as T cells (CD3, CD8, CD4, CD45RA, CD45RO), NK cells (CD7 and CD56), myeloid cells (CD163) and antigen presenting cells (HLA-DR). In addition, the proliferation



C

Top ten ranked clusters cotributing to the tissue-specific t-SNE sample signatures

Nr	Intestine	Spleen	Liver
1	CD4_01	Mye_12	Lin$^-$_16
	CD161$^+$c-KIT$^-$ Tem	IgM$^+$CD11c^{dim} pDC	CD34$^+$CD7$^-$HLA-DR$^+$CD38^{dim}
2	CD4_08	B_04	ILC_38
	CD161$^+$ Tcm	CCR7$^-$CCR6$^+$ B	CD34$^+$CD7$^+$CD38$^+$CD45RA$^+$CD56$^+$
3	CD4_14	CD8_19	Lin$^-$_04
	CCR6$^+$CD161$^-$CD27$^+$ Tn	CD103$^+$CCR6$^+$c-KIT$^-$ Tn	CD34$^-$c-KIT$^+$
4	CD4_06	Lin$^-$_01	Mye_13
	CD27^{dim}CD161$^+$c-KIT$^+$ Tem	CD34$^+$CD56$^+$CD38^{dim}HLA-DR$^+$	CD11c^{dim}CD45RA$^-$
5	CD4_09	B_03	Mye_14
	CD161$^-$CD27$^+$ Tcm	CCR7^{dim}CCR6$^+$ B	CD4$^+$ M2 macrophage
6	CD4_04	CD8_20	Lin$^-$_18
	CD161$^+$c-KIT$^+$CD103$^+$ Tem	PD-1$^+$CD11b$^+$ Temra	CD34$^+$CD7$^-$c-KIT$^+$HLA-DR$^+$CD38^{dim}
7	CD4_05	CD8_11	Mye_15
	CD161$^+$c-KIT$^+$ DN	CD56$^+$PD-1$^-$CD11b$^+$ Temra	CD4$^+$ pDC
8	Mye_01	ILC_34	Lin$^-$_20
	CD45RA^{dim}CD123^{dim}CD11c^{dim}	CD16$^+$CD11b$^+$CD11c$^+$CD27$^+$ NK	CD34$^+$CD38^{dim}c-KIT$^+$HLA-DR$^-$
9	Lin$^-$_15	CD8_12	Mye_25
	CD34$^+$CD7$^+$CCR7$^+$CD45RA$^+$CD38^{dim}	PD-1$^+$CD161$^-$ Tem	CD4$^+$CD123^{dim}CD45RA$^+$
10	Lin$^-$_17	B_07	Lin$^-$_12
	CD34$^+$CD7$^+$CCR7^{dim}CD45RA$^+$CD38$^+$	CCR6$^-$CD38^{dim} B	CD34^{dim}CD38$^+$CD7^{dim}HLA-DR$^-$

Figure 5 Integrated analysis of immune cluster composition reveals highly discriminatory tissue-specific signatures. (A) Principal component analysis (PCA) of 13 fetal samples. PCA was computed on the cell frequencies of 177 immune clusters per sample (as the percentage of CD45$^+$ cells) and top 2 principal components are shown describing 58% of total variance. One dot represents a single sample. Colors represent different fetal tissues. (B) t-SNE embedding of 177 immune clusters derived from 13 samples showing the tissue-specific signatures. Every dot represents a single immune cluster. The size of the dot is proportional to the value of cell frequency, which is more similar across tissues when the clusters are together closer. (C) Table depicting top ten ranked clusters contributing to tissue-specific t-SNE sample signatures with biological annotation. Tem = effector memory T cells, Tcm = central memory T cells, Temra = terminally differentiated T cells, pDC = plasmacytoid dendritic cells and NK = natural killer cell.

marker Ki-67 was included (**Table S3**). With this panel, tissue sections derived from the fetal intestine, spleen and liver were analysed. All antibodies showed clear positive signals as illustrated in one of the regions of interest (ROIs) derived from the fetal intestine (**Figure S4**).

To obtain further information on the phenotype and localization of immune cell subsets within the tissue context, we visualized combinations of specific markers (**Figure 6**). Staining with E-cadherin, SMA and DNA were used to show the tissue structure, reflecting the overall morphological differences between the tissues and the distribution of the immune cells was revealed by CD45 staining. Most CD45$^+$

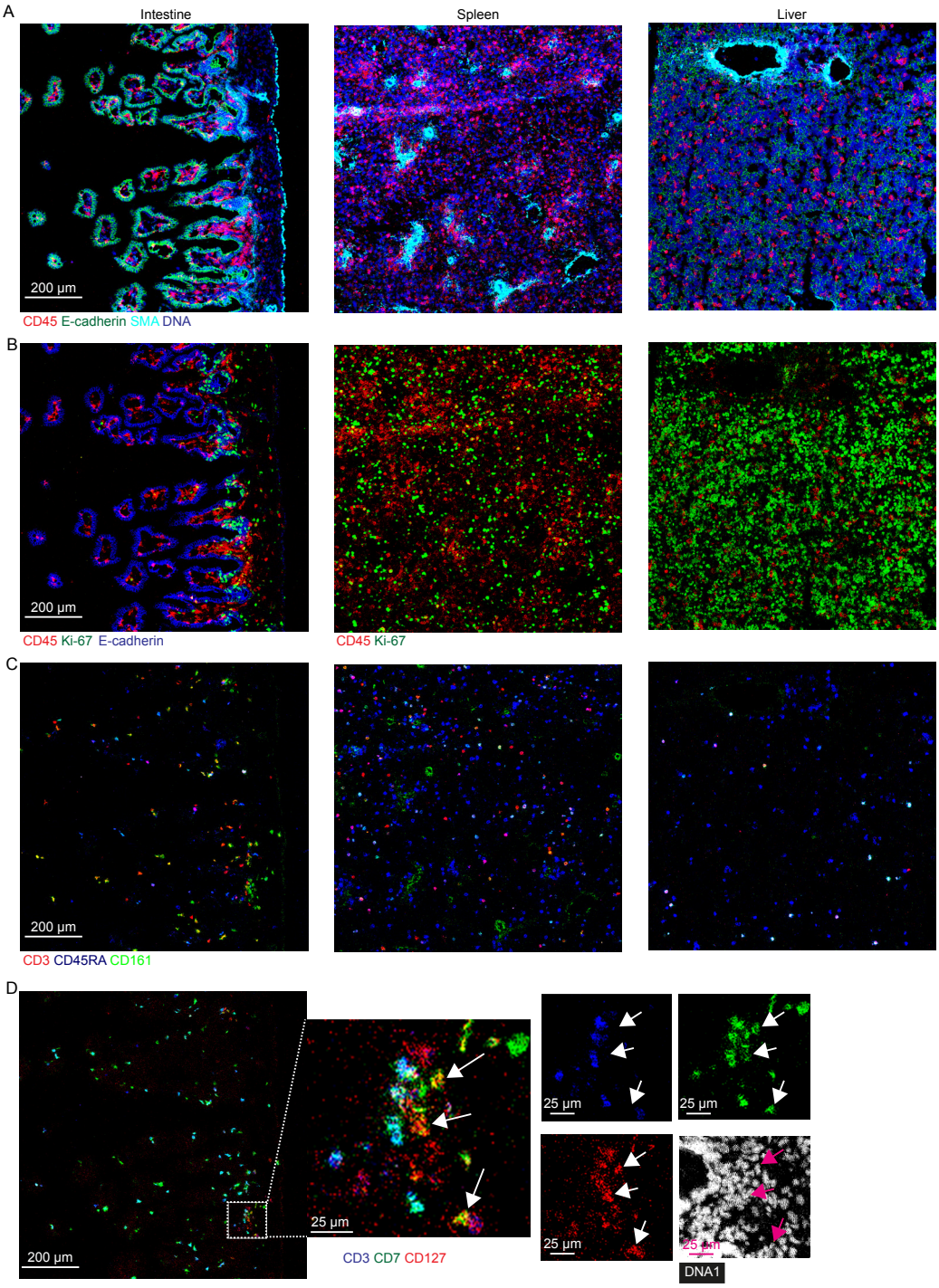

A
Intestine
Spleen
Liver
200 µm
CD45 E-cadherin SMA DNA
B
200 µm
CD45 Ki-67 E-cadherin
CD45 Ki-67
C
200 µm
CD3 CD45RA CD161
D
200 µm
25 µm
CD3 CD7 CD127
25 µm
25 µm
25 µm
DNA1

Figure 6 Spatial localization of immune and stromal cell subsets across tissues. (A-D) Representative mass cytometry images of a fetal intestine, spleen and liver, displaying the overlay of the indicated makers. Scale bar: 200 µm. Colors and scale bars are identical in panel A-C. (A) CD45 in red, E-cadherin in green, SMA in cyan and DNA in blue. (B) CD45 in red, Ki-67 in green and E-cadherin in blue. (C) CD3 in red, CD45RA in blue and CD161 in green. (D) CD3 in red, CD7 in green, CD127 in blue and DNA in white. White and purple arrows indicate the $CD3^{-}CD7^{+}CD127^{+}$ ILCs.

immune cells were in the E-cadherin^{-} lamina propria in the intestine whereas these cells were scattered in the fetal spleen and liver (**Figure 6A)**. While the majority of $CD45^{+}$ immune cells lacked expression of Ki-67 in all three tissues, many non-immune cells in the fetal liver and to a lesser extent in the fetal spleen were stained by the Ki-67-specific antibody, indicative of cell division and likely linked to the development of these organs at this stage (**Figure 6B**). In agreement with earlier work, Ki-67^{+} cells in the intestines resided in the crypt (**Figure 6B**), where stem cells and progenitor cells are enriched [22]. Consistent with findings using suspension mass cytometry, both $CD45RA^{+}$ Tn and $CD45RA^{-}$ Tm cells were identified in all three tissues. Compared with the spleen and liver, the $CD45RA^{-}$ Tm cells were most abundant in the intestine, especially $CD161^{+}CD45RA^{-}$ T cells (**Figure 6C**). $CD161^{-}CD45RA^{-}$ T cells and $CD161^{+}CD45RA^{+}$ T cells were more predominate in the spleen and liver as compared with the other two tissues, respectively (**Figure 6C**). Finally, simultaneous visualization of CD3, CD7 and CD127 reveals that $CD3^{+}$ T cells and $CD3^{-}CD7^{+}CD127^{+}$ ILCs (**Figure 6D, white and purple arrows**) were located in close proximity of each other in the lamina propria of the intestine.

Together, these data confirmed the existence of tissue-specific subsets in situ and underpinned the distinctness of the immune system in the tissues at the second trimester during pregnancy.

Discussion

While recent studies have investigated the heterogeneity and functionality of the developing immune system in prenatal life [5, 9, 10], knowledge of the immune system development remains sparse due to the relative scarcity of fetal tissues. In the current study, we applied high-dimensional mass cytometry to dissect both the innate and adaptive immune compartment in the human fetal intestine, spleen and liver in an unbiased and data-driven manner. For data analysis, we used HSNE [17], an in-house developed software, which allows the identification of phenotypically distinct clusters in datasets consisting of millions of cells. In line with previous work [5, 6, 9, 10, 21], we were able to readily define the major cell lineages

and substantial heterogeneity therein, and reveal the identity of previously unrecognized cell clusters. Moreover, our results provide evidence for tissue-specific compartmentalization of both the innate and adaptive immune compartment early in the second trimester of pregnancy. Altogether, our results confirm and extend previous studies that have analyzed the immune subsets in mucosal and lymphoid tissues from human fetus with immunohistochemical analysis [8, 23] or traditional multi-color flow cytometry [5, 24]. Our comprehensive analysis provides a valuable resource that may aid future studies into the development of the human immune system during gestation.

Consistent with previous reports, our analysis confirmed the distinctness of the human fetal liver, with enrichment of myeloid and $CD34^+$ precursor cell clusters. Six out of the top ten ranked clusters contributing to liver-specific signatures were $CD34^+$ precursor cells. Also, we observed extensive diversity within the $CD34^+$ cell population due to differential expression of several markers including c-KIT and HLA-DR, possibly reflecting differentiation pathways and linked to the crucial role of the fetal liver in haematopoiesis in prenatal life.

In addition, several other features distinguished the immune systems found in the tissues examined. Here, $CD27^-CD11b^-$ NK cells exhibiting dominant inhibitory signals [25] were found primarily in the intestine while $CD27^-CD11b^+$ NK and $CD27^+CD11b^+$ NK cells with dominant activating signals [25] predominate in the spleen and liver and contribute to the tissue partitions and demarcate different functions in the tissues examined. Also, helper-ILCs were mainly present in the fetal intestine, as reported [21]. Moreover, we have previously reported that memory-like T cells are located in the fetal intestine [9] but to a much lower degree in fetal spleen and liver indicative of in situ immune priming and maturation of T cells in the developing intestine. Intriguingly, previously unidentified $CCR6^+$c-$KIT^+CD45RO^+$ subpopulations of T cells were identified in all lineages in the fetal intestine, but not in other organs. Expression of c-KIT, a receptor for stem cell factor, has been shown to demarcate a subset of human $CD8^+$ Tm cells with self-renewal properties [26]. Moreover, flow-sorted c-KIT^+ Tm cells cocultured with OP9 stromal cells in the presence of stem cell factor and IL-7 proliferated in vitro (data not shown). These observations suggest a potential role for $CCR6^+$c-KIT^+ Tm cells in the conservation of mucosal T cell memory.

CD161 has previously been shown to define a transcriptional and functional phenotype across human T cell lineages with an innate-like ability to respond to cytokines [27] while another study has described that CD161 mediates prenatal immune suppression [28]. In our study, CD161-expressing T cells are more

pronounced in the human fetal intestine as compared to fetal spleen and liver, consistent with a unique function of CD161$^+$ cells within the fetal intestine. HLA-DR$^+$ myeloid cells and B cells were found in all three organs, however, all B cells lacked expression of CD27, a classical memory B cell marker.

We have used imaging mass cytometry to gain insight into the localization of specific immune subsets within the tissue context. The results reiterate the highly distinct organization of the immune compartment in the three tissues investigated and allow the simultaneous analysis of cellular activity, i.e. cell proliferation. We demonstrate that this can be used to confirm results obtained by single-cell suspension mass cytometry while simultaneously obtaining information on the co-localization of specific immune subsets. For example, we obtained evidence that in the intestinal lamina propria CD3$^+$ T cells are found in close proximity of CD3$^-$CD7$^+$CD127$^+$ ILC, possibly pointing towards "crosstalk" between these cell subsets. This will be the topic of future studies.

While this study is one of the first to study the immune landscape in human fetal tissues with high-dimensional analysis at the single-cell level, our study has limitations. First, we did not observe much heterogeneity in the B population as our antibody panel lacked several markers required for this. It would thus be important to include other antibodies such as IgD, CD10 and CD5 to dissect the B compartment further. Second, we had only access to fetal material from the second trimester (16-21 gestational weeks). In future studies, it would thus be valuable also to include samples from the first and third trimester. Third, while we identified 177 known and unknown cell clusters, the functional properties of these subsets need to be clarified.

Nevertheless, our mass cytometric analysis provides the first global, comprehensive and detailed description of the immune landscape in the developing fetus in several tissues, reveals tissue-specific signatures and demonstrates a clear immune compartmentalization in the tissues in prenatal life. Recent studies have indicated that events shaping the immune system in utero and in early life can have a significant impact on the development of diseases in adult life [29, 30, 31]. Detailed understanding of the early development of the immune system is crucial for the development of strategies to prevent such diseases and our study may help to achieve that goal.

Materials and methods

Human fetal tissues and cell isolation

Human fetal tissues from elective abortions were obtained after informed consent. The gestational age ranged from 16 to 21 weeks. Single-cell suspensions from different tissues were prepared as previously described [18]. Briefly, the mesentery and meconium were removed from the fetal intestine. The intestines were then cut into small fragments and treated with 1 mM dithiothreitol (Fluka) in 15 mL of HBSS (Sigma-Aldrich) for 2 x 10 min (replacing buffer) at room temperature to dissolve the mucus and subsequently with 1 mM ethylenediaminetetraacetic acid (Merck) in 15 mL of Hank's balanced salt solution (ThermoFisher Scientific), under rotation for 2 x 1 h (replacing buffer) at 37 °C to separate the epithelium from the lamina propria fraction. To obtain single-cell suspensions from the lamina propria, the intestines were rinsed with HBSS and incubated with 15 mL Iscove's Modified Dulbecco's Medium (IMDM; Lonza) supplemented with 10% fetal calf serum (FCS), 10 U/mL collagenase IV (Worthington), 200 µg/mL DNAse I grade II (Roche Diagnostics), at 37 °C overnight, after which cell suspensions were filtered through a 70 µm nylon cell strainer. Finally, the immune cells were isolated with a Percoll (GE Healthcare) gradient. Fetal liver and spleen tissues were cut into small pieces then filtered through a 70 µm nylon cell strainer and the immune cells were isolated with Ficoll-Paque™ density gradient (provided by the pharmacy of Leiden University Medical Center). All 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.

Mass antibodies and antibody conjugation

Antibodies used for suspension and imaging mass cytometry are listed in **Table S1** and **Table S3**, separately. Conjugation of the purified antibodies lacking carrier protein with metal reporters was performed with the MaxPar X8 antibody labeling kit (Fluidigm Sciences) according to the manufacturer's instruction.

Suspension mass cytometry staining and data acquisition

Procedures for suspension mass cytometry antibody staining and data acquisition were carried out as previously described [18]. Briefly, cells from different fetal tissues were incubated with 1 mL 500x diluted 500 µM Cell-ID intercalator-103Rh (Fluidigm Sciences) for 15 min at room temperature to identify dead cells. Cells were then stained with metal-conjugated antibodies for 45 min at room temperatur. After staining, cells were labeled with 1 mL 1,000x diluted 125 µM Cell-ID intercalator-Ir (Fluidigm Sciences) to stain all cells in MaxPar Fix and Perm

Buffer (Fluidigm Sciences) overnight at 4 °C. Finally, cells were acquired by CyTOF 2™ mass cytometer (Fluidigm Sciences). Data were normalized by using EQ Four Element Calibration Beads (Fluidigm Sciences) with the reference EQ passport P13H2302 during the course of each experiment.

Imaging mass cytometry staining and data acquisition

Snap-frozen human fetal splenic, intestinal and liver biopsies were sectioned at a thickness of 5 µm. All sections were fixed by incubating with 1% paraformaldehyde for 5 min at room temperature followed by 100% methanol for 5 min at −20 °C. After fixation, tissue sections were washed in Dulbecco's phosphate-buffered saline (ThermoFisher Scientific) containing 0.5% bovine serum albumin and 0.05% Tween, rehydrated in additive-free Dulbecco's phosphate-buffered saline. After washing again, tissue sections were blocked with Superblock Solution (ThermoFisher Scientific) for 30 min in a humid chamber. Tissue sections were then stained with a metal-conjugated antibody master mix overnight at 4 °C, washed and incubated with 125 nM Cell-ID Intercalator-Ir for 30 min at room temperature. After a further wash, tissue sections were dipped in Milli-Q water (Merck Millipore) for 1 min and dried for 20 min at room temperature. The acquisition was performed 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 $1 \times 1\,mm^2$. All data were stored as MCD files and txt files.

Data analysis

Data for single, live $CD45^{+/dim}$ cells gated from each sample individually using Cytobank as shown in **Figure S1**, were sample tagged and hyperbolic arcsinh transformed with a cofactor of 5 within Cytosplore^{+HSNE} [17]. The major immune lineages in **Figure 1** were then identified by performing hierarchical stochastic neighbor embedding (HSNE) in Cytosplore^{+HSNE} software [17]. HSNE was carried out with default parameters (perplexity: 30; iterations: 1000). For the cluster identification, each cluster contains at least 100 cells. All HSNE plots were generated in Cytosplore [32]. Cellular signatures of immune cells for each sample individually were generated in Cytosplore. Due to homogeneity and abundance of B cells in the fetal spleen, B cells from each spleen were downsampled to 50,000 cells to deduce the cellular signatures. The similarity between two paired t-SNE maps was quantified by Jensen-Shannon (JS) divergence by measuring the similarity between corresponding probability density distribution after converting t-SNE maps to two-dimensional probability density functions in Matlab R2015b. Hierarchical clustering of the phenotype heatmap was created with Euclidean

correction and average linkage clustering while the cell frequency heatmap with Spearman correction and average linkage clustering in Matlab R2015b. Principal component analysis (PCA) was computed with the cluster frequencies of $CD45^{+/dim}$ cells in the individual samples using "prcomp" function, and the result was visualized using "ggbiplot" function in R software. The cluster t-SNE map in **Figure 5B** was performed as our previous study [19]. Briefly, the data matrix with cluster frequencies of $CD45^{+/dim}$ cells in the individual samples as input variance was normalized and computed to select the top ten highest variance principal components as input to the t-SNE analysis. Hence, the cluster with similar profiles clustered together in the t-SNE map. Images in **Figure 6** and **Figure S4** were generated using MCD Viewer software v1.0.560 (Fluidigm).

Data availability

Mass cytometry data will be available via Flow Repository.

Author contributions

NL and FK conceived the study and wrote the manuscript. NL performed most experiments with the help of NG. NL performed most of the analyses with the help of VvU, TA, AS, JE, AM and BPFL. SCdSL provided human fetal tissues. All authors discussed the results and commented on the manuscript.

Acknowledgements

We thank the Center for Contraception, Abortion and Sexuality (CASA) in Leiden and The Hague for their efforts in collecting and providing the fetal material, Dr. M Bialecka and M Nieveen for the dissection of the fetal tissues and ME Ijsselsteijn and NFCC de Miranda for optimizing the imaging mass cytometry panel. This research was supported by the Leiden University Medical Center, The Netherlands Organization for Scientific Research (ZonMW grant 91112008), NWO Applied Technical Sciences grant 12721 (VANPIRE) and the European Commission under an MSCA-ITN award (grant no. 675743/ISPIC). N Li and N Guo were supported by the China Scholarship Council.

References

1. Frascoli, M. et al. Alloreactive fetal T cells promote uterine contractility in preterm labor via IFN-gamma and TNF-alpha. Sci Transl Med 10 (2018).

2. Steiniger, B., Ulfig, N., Risse, M. & Barth, P.J. Fetal and early post-natal development of the human spleen: from primordial arterial B cell lobules to a non-segmented organ. Histochem Cell Biol 128, 205-215 (2007).

3. Zhang, X., Zhivaki, D. & Lo-Man, R. Unique aspects of the perinatal immune system. Nat Rev Immunol 17, 495-507 (2017).

4. Mold, J.E. et al. Maternal alloantigens promote the development of tolerogenic fetal regulatory T cells in utero. Science 322, 1562-1565 (2008).

5. McGovern, N. et al. Human fetal dendritic cells promote prenatal T-cell immune suppression through arginase-2. Nature 546, 662-666 (2017).

6. Michaelsson, J., Mold, J.E., McCune, J.M. & Nixon, D.F. Regulation of T cell responses in the developing human fetus. J Immunol 176, 5741-5748 (2006).

7. Schuster, C. et al. Phenotypic characterization of leukocytes in prenatal human dermis. J Invest Dermatol 132, 2581-2592 (2012).

8. 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).

9. Li, N. et al. Memory CD4(+) T cells are generated in the human fetal intestine. Nat Immunol (2019).

10. 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, 238ra272 (2014).

11. Leeansyah, E., Loh, L., Nixon, D.F. & Sandberg, J.K. Acquisition of innate-like microbial reactivity in mucosal tissues during human fetal MAIT-cell development. Nat Commun 5, 3143 (2014).

12. Hazenberg, M.D. & Spits, H. Human innate lymphoid cells. Blood 124, 700-709 (2014).

13. Williams, T.J., Jones, C.A., Miles, E.A., Warner, J.O. & Warner, J.A. Fetal and neonatal IL-13 production during pregnancy and at birth and subsequent development of atopic symptoms. J Allergy Clin Immunol 105, 951-959 (2000).

14. Olin, A. et al. Stereotypic Immune System Development in Newborn Children. Cell 174, 1277-1292 e1214 (2018).

15. Bandura, D.R. et al. Mass cytometry: technique for real time single cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry. Anal Chem 81, 6813-6822 (2009).

16. van der Maaten, L. & Hinton, G. Visualizing Data using t-SNE. J Mach Learn Res 9, 2579-2605 (2008).

17. 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).

18. Li, N. et al. Mass cytometry reveals innate lymphoid cell differentiation pathways in the human fetal intestine. J Exp Med (2018).

19. 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).

20. Howie, D. et al. Extrathymic T cell differentiation in the human intestine early in life. J Immunol 161, 5862-5872 (1998).

21. Artis, D. & Spits, H. The biology of innate lymphoid cells. Nature 517, 293-301 (2015).

22. Howell, J.C. & Wells, J.M. Generating intestinal tissue from stem cells: potential for research and therapy. Regen Med 6, 743-755 (2011).

23. Hwang, S.E., Kim, J.H., Yu, H.C., Murakami, G. & Cho, B.H. Lymphocyte subpopulations in the liver, spleen, intestines, and mesenteric nodes: an immunohistochemical study using human fetuses at 15-16 weeks. Anat Rec (Hoboken) 297, 1478-1489 (2014).

24. Gunther, U. et al. Phenotypic characterization of CD3-7+ cells in developing human intestine and an analysis of their ability to differentiate into T cells. J Immunol 174, 5414-5422 (2005).

25. Fu, B.Q., Tian, Z.G. & Wei, H.M. Subsets of human natural killer cells and their regulatory effects. Immunology 141, 483-489 (2014).

26. Turtle, C.J., Swanson, H.M., Fujii, N., Estey, E.H. & Riddell, S.R. A distinct subset of self-renewing human memory CD8+ T cells survives cytotoxic chemotherapy. Immunity 31, 834-844 (2009).

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

28. Halkias, J. et al. CD161 mediates prenatal immune suppression of IFNγ-producing PLZF⁺ T cells. bioRxiv, 305128 (2018).

29. Chen, J.C. et al. Fetal Phagocytes Take up Allergens to Initiate T-Helper Cell Type 2 Immunity and Facilitate Allergic Airway Responses. Am J Respir Crit Care Med 194, 934-947 (2016).

30. Prescott, S.L. et al. Transplacental priming of the human immune system to environmental allergens: universal skewing of initial T cell responses toward the Th2 cytokine profile. J Immunol 160, 4730-4737 (1998).

31. Renz, H. et al. The neonatal window of opportunity-early priming for life. J Allergy Clin Immunol 141, 1212-1214 (2018).

32. Hollt, T. et al. Cytosplore: Interactive Immune Cell Phenotyping for Large Single-Cell Datasets. Comput Graph Forum 35, 171-180 (2016).

Supplemental information

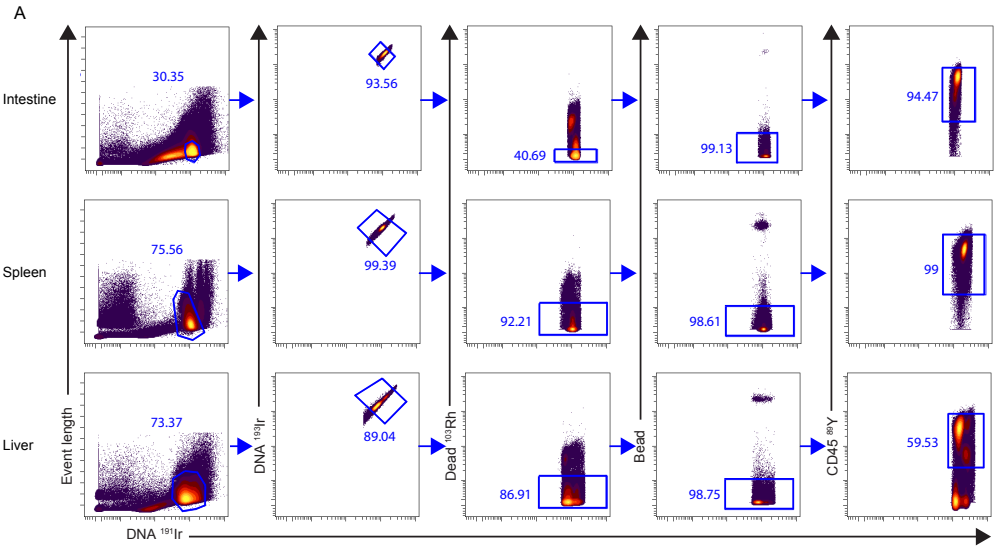


Figure S1 Gating strategy for single, live CD45⁺ cells across fetal tissues. Representative biaxial plots from a fetal intestine (top panel, N =7), a fetal spleen (middle panel, N =3) and a fetal liver (bottom panel, N=3) depicting the sequential gates for single, live CD45⁺ cells with percentages.

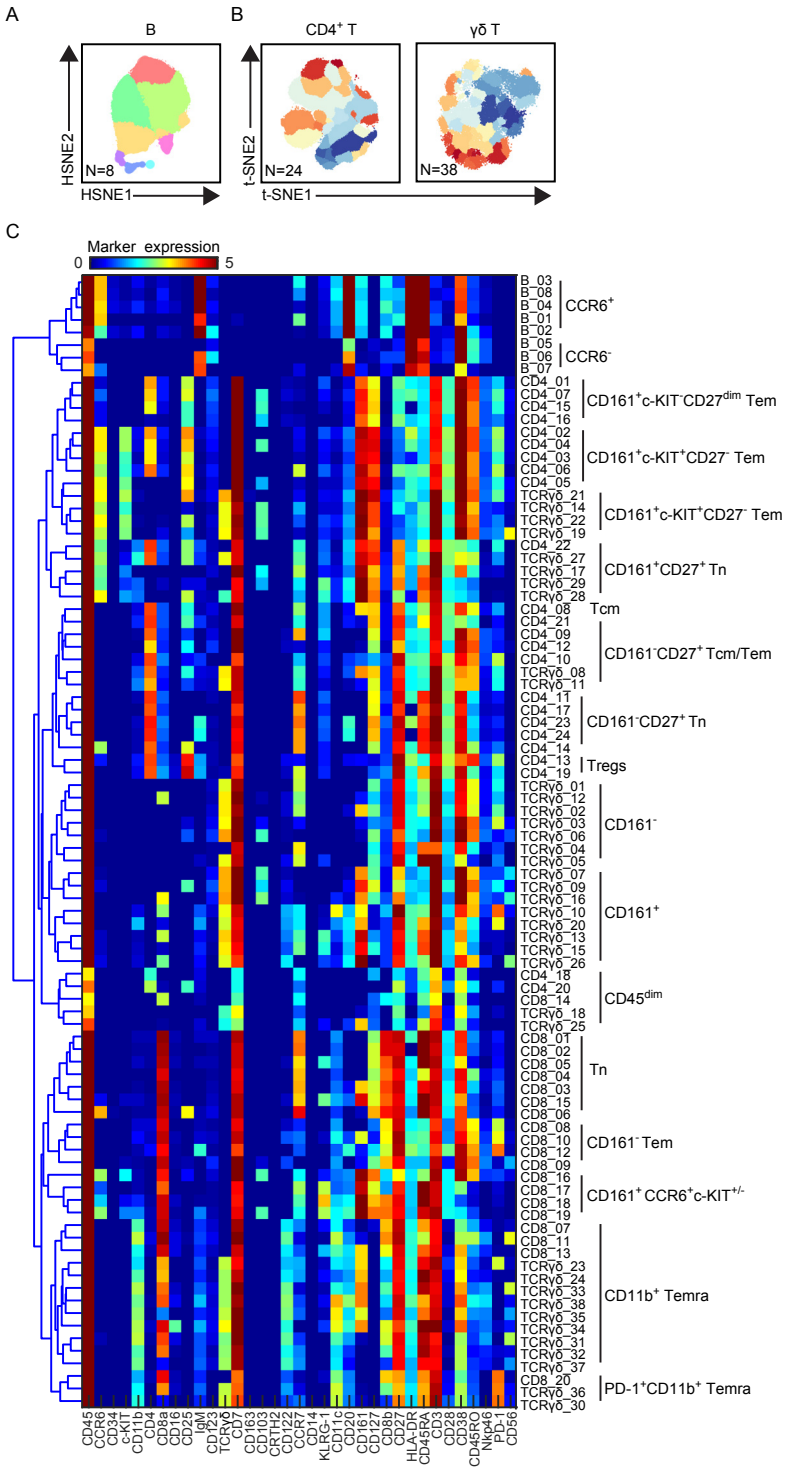


Figure S2 Analysis of the adaptive immune system across human fetal tissues using mass cytometry. (A) HSNE embedding of B cells (8.2 x 10^5 cells) at the second level. Colors indicate cluster partitions. (B) t-SNE embedding of $CD8^+$ T (6.9 x 10^4 cells) and γδ T cells (3.6 x 10^4 cells) at the single-cell level. Colors indicate cluster partitions. (C) Heatmap showing the median marker expression values of the 90 clusters identified in B, $CD4^+$ T, $CD8^+$ T and γδ T cells from seven fetal intestines, three spleens and three livers, and hierarchical clustering thereof. Tem = effector memory T cells, Tcm = central memory T cells and Temra = terminally differentiated T cells.

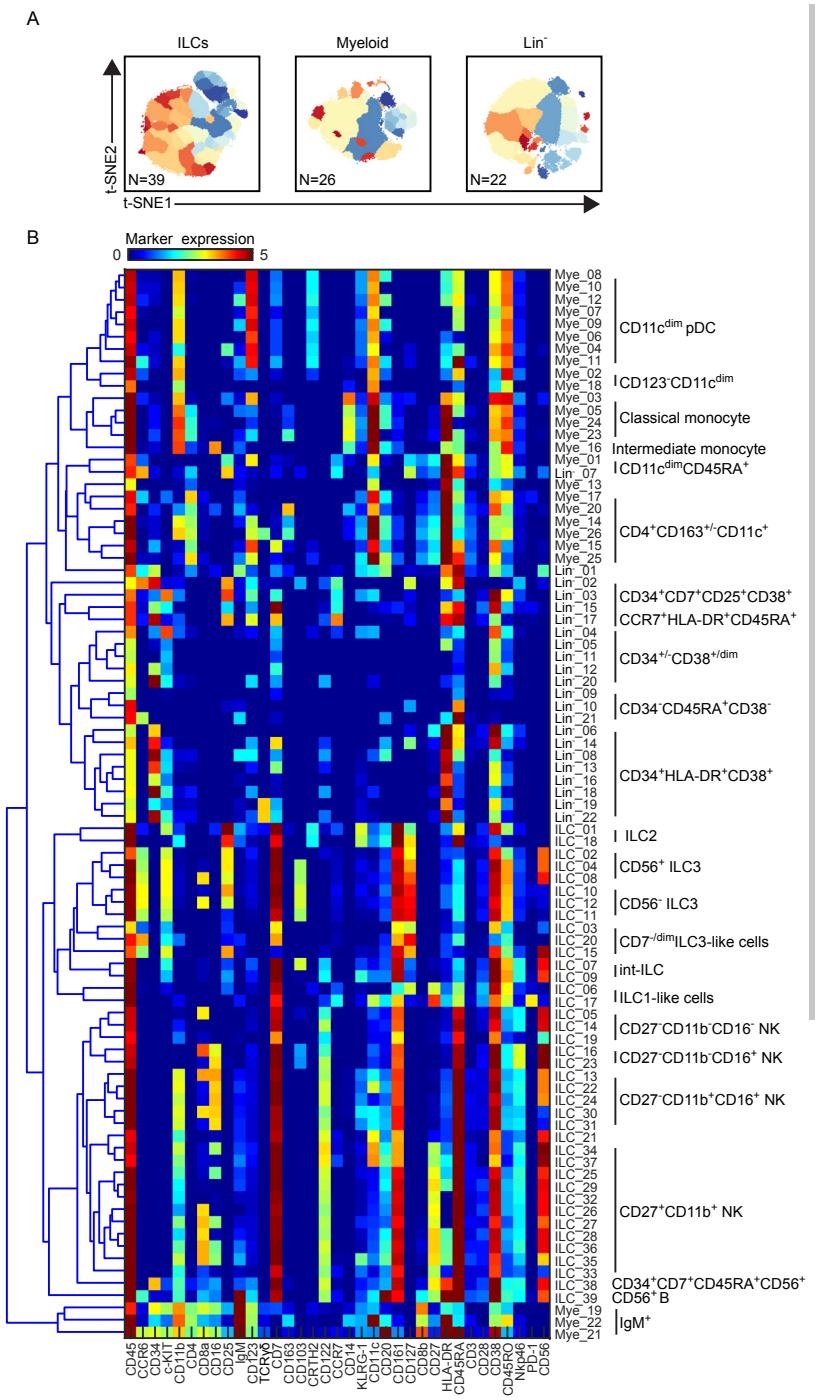


Figure S3 Analysis of the innate immune system across human fetal tissues using mass cytometry. (A) t-SNE embedding of ILCs (2.6 x 10^5 cells), myeloid cells (6.4 x 10^4 cells) and Lin$^-$ (8.7 x 10^4 cells) at the single-cell level. Colors indicate cluster partitions. (B) Heatmap showing the median marker expression values of the 87 clusters identified in ILCs, myeloid cells and Lin$^-$ from seven fetal intestines, three spleens and three livers, and hierarchical clustering thereof. pDC = plasmacytoid dendritic cells.

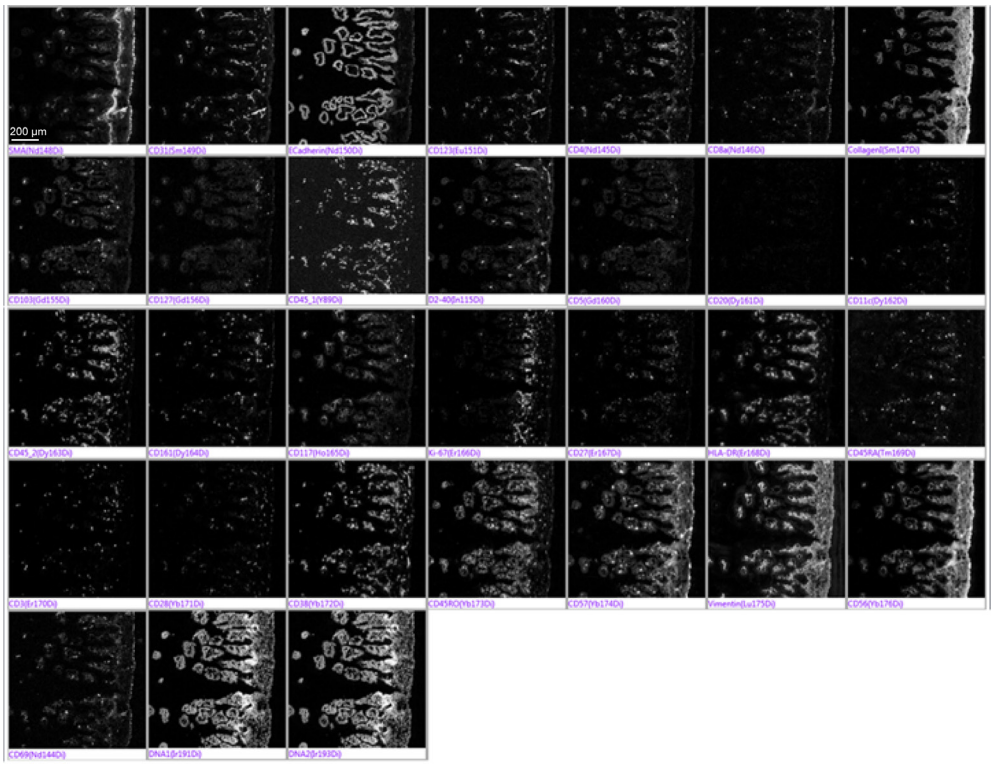


Figure S4 Validation of the imaging mass cytometry antibody panel on human fetal intestine. Representative mass cytometry images of a human fetal intestine showing the expression of the indicated markers. Scale bar: 200 µm. Scale bars are identical in all the images.

	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).						

Table S1 Suspension mass cytometry antibody panel

Fetus ID	Gestational age, wk	Intestine	Liver	Spleen
1	19	+	-	-
2	16	+	-	-
3	19	+	-	-
4	19	+	-	-
5	21	+	+	-
6	21	+	-	-
7	21	+	-	+
8	21	-	+	+
9	16	-	+	+

Table S2 Sample information

	Antigen	Tag	Supplier	Final dilution
1	CD8a	^{146}Nd	Flui	1/50
2	CD11c	^{162}Dy	Flui	1/50
3	CD117	^{165}Ho	BioL	1/50
4	CD38	^{172}Yb	Flui	1/50
5	CD69	^{144}Nd	Flui	1/50
6	CD45	^{89}Y	Flui	1/50
7	CD4	^{145}Nd	Flui	1/50
8	SMA	^{148}Nd	CST	1/200
9	CD31	^{149}Sm	CST	1/100
10	CD123	^{151}Eu	Flui	1/50
11	CD7	^{153}Eu	Flui	1/100
12	CD163	^{154}Sm	Flui	1/100
13	CD161	^{164}Dy	Flui	1/50
14	CD27	^{167}Er	Flui	1/50
15	CD45RA	^{169}Tm	Flui	1/100
16	CD3	^{170}Er	Flui	1/100
17	Vimentin	^{175}Lu	CST	1/200
18	CD56	^{176}Yb	Flui	1/100
19	FOXp3	^{142}Nd	CST	1/100
20	CD5	^{160}Gd	BioL	1/25
21	HLA_DR	^{168}Er	BIoL	1/600
22	E-cadherin	^{150}Nd	CST	1/50
23	CD103	^{155}Gd	Abcam	1/50
24	CD127	^{156}Gd	Beckman	1/50
25	CD45	^{163}Dy	CST	1/100
26	CD28	^{171}Yb	BioL	1/50
27	CD45RO	^{173}Yb	BioL	1/50
28	Ki-67	^{166}Er	CST	1/200
29	CD57	^{174}Yb	Abcam	1/100
30	Collagen1	^{147}Sm	Millipore	1/100
31	D2-40	^{115}In	BioL	1/50

Fluidigm (Flui) and Biolegend (BioL).

Table S3 Imaging mass cytometry antibody panel

Chapter 5

Mass Cytometry of the Human Mucosal Immune System Reveals Tissue- and Disease-Associated Immune Subsets

Vincent van Unen,[1] Na Li,[1] Ilse Molendijk,[2] Mine Temurhan,[3] Thomas Höllt[7], Andrea E. van der Meulen-de Jong,[2] Hein W. Verspaget,[2] M. Luisa Mearin,[4] Chris J. Mulder,[6] Jeroen van Bergen,[1] Boudewijn P. F. Lelieveldt,[5,7] Frits Koning.[1]

[1] Department of Immunohematology and Blood Transfusion, Leiden University Medical Center, Leiden, Netherlands

[2]Department of Gastroenterology, Leiden University Medical Center, Leiden, Netherlands

[3]Department of Center for Proteomics, Leiden University Medical Center, Leiden, Netherlands

[4]Department of Pediatrics, Leiden University Medical Center, Leiden, the Netherlands

[5]Department of LKEB Radiology, Leiden University Medical Center, Leiden, Netherlands.

[6]Department of LKEB Gastroenterology, VU Medical Center, Amsterdam, The Netherlands.

[7]Department of Pattern Recognition and Bioinformatics Group, Delft University of Technology,Delft, The Netherlands

Immunity 44:5, 1227-1239 (2016)

Abstract

Inflammatory intestinal diseases are characterized by abnormal immune responses and affect distinct locations of the gastrointestinal tract. Although the role of several immune subsets in driving intestinal pathology has been studied, a system-wide approach that simultaneously interrogates all major lineages on a single-cell basis is lacking. We used high-dimensional mass cytometry to generate a system-wide view of the human mucosal immune system in health and disease. We distinguished 142 immune subsets and through computational applications found distinct immune subsets in PBMCs and intestinal biopsies that distinguished patients from controls. In addition, mucosal lymphoid malignancies were readily detected as well as precursors from which these likely derived. These findings indicate that an integrated high-dimensional analysis of the entire immune system identifies immune subsets that are associated with the pathogenesis of complex intestinal disorders. This may have important implications for diagnostic procedures, immune-monitoring and treatment of intestinal diseases and mucosal malignancies.

Introduction

The intestinal immune system protects us from bacterial, viral and parasitic infections. Disruption of intestinal homeostasis, however, can lead to a variety of autoinflammatory intestinal diseases, including celiac disease (CeD) and Crohn's disease (CD), which together have a prevalence of 1,500 per 100,000 adults in the Western world [1, 2]. Both diseases are multifactorial and encompass a broad spectrum of clinical phenotypes and ages of onset. CeD is a disease of the small intestine caused by pro-inflammatory $CD4^+$ T cell responses specific for dietary gluten and concomitant destruction of the epithelium due to activation of intraepithelial $CD8^+$ T cells. The introduction of a strict gluten-free diet constitutes a highly effective treatment for CeD but nevertheless 2-5% of patients develop refractory CeD type II (RCDII) with persistent inflammation. RCDII is characterized by a monoclonal outgrowth of aberrant intra-epithelial lymphocytes (IELs) from which an aggressive enteropathy-associated T cell lymphoma (EATL) evolves in 40% of patients [3]. In contrast, CD affects the terminal ileum and/or colon and results from aberrant immune responses against the microbiota [4]. CD is usually treated with the use of lifelong pharmacotherapy [5], including biologicals (e.g. anti-TNF) to reduce chronic inflammation and to accomplish sustained remission. Despite achieving states of remission, perianal fistulas occur in 25% of CD patients and this is accompanied by multiple relapses and a poor prognosis due to insufficient healing [6, 7].

Although the role of several immune subsets in driving intestinal pathology has been studied in CeD [8], RCDII [9] and CD [10], a system-wide approach that simultaneously interrogates immune subsets across all major lineages on a single-cell basis is currently lacking. High-dimensional mass cytometry (cytometry by time-of-flight; CyTOF) now offers the possibility to analyze many cellular markers simultaneously, providing an opportunity to analyze the mucosal immune system with unprecedented resolution [11]. Novel computational tools have been developed to handle the high-dimensional single-cell datasets that originate from mass cytometry [12-14]. In the current study we applied mass cytometry to analyze the composition of the immune compartment present in intestinal biopsies and paired peripheral blood mononuclear cell (PBMC) samples of patients with inflammatory intestinal diseases and controls. We identified 142 distinct immune cell subsets and through computational applications we found immune subsets in PBMCs and intestinal biopsies that distinguished patients with inflammatory diseases from controls. In addition, mucosal lymphoid malignancies were readily detected as well as the precursors from which these likely derived. Thus, mass cytometry unveiled previously unappreciated heterogeneity in the immune system and our

observations may help to develop improved diagnostic and therapeutic approaches for inflammatory bowel diseases.

Results

SPADE analysis identifies major immune lineages in peripheral blood and intestine

We designed a CyTOF panel of 32 metal isotope-tagged monoclonal antibodies, which was designed to obtain a global overview of the heterogeneity of the innate and adaptive immune system (**Table S1**). For this purpose, the panel contained lineage markers that distinguish the major adaptive and innate immune cell populations. In addition, markers were included to distinguish naïve from memory cells, resting from activated cells, and to identify homing properties and potential responsiveness to humoral factors. With this panel, we analyzed single-cell suspensions from biological samples including duodenum biopsies (N=36), rectum biopsies (N=13), perianal fistulas (N=6) and peripheral blood mononuclear cells (PBMC) from control individuals (N=15) and from patients with inflammatory intestinal diseases (CeD, N=13; RCDII, N=5; EATLII, N=1 and CD, N=10) (**Table S2**). The large majority (N=28) of antibodies allowed clear discrimination of antibody-positive and -negative cells (**Figure S1**). To monitor the robustness of the measurements we included a standardized PBMC sample at regular intervals in the acquisition sessions during the entire 9-month study period. These consecutive control samples yielded highly similar results (**Figure S2A-C**), demonstrating the reproducibility of the data acquisition. We discriminated live, single $CD45^+$ cells with DNA stains and event length. (**Figure S2D**). From the intestinal biopsies we acquired 27,500 duodenal, 17,500 rectal, and 76,500 perianal fistulous $CD45^+$ cells on average (**Figure 1A**), and 76,500 $CD45^+$ cells from the PBMC samples (data not shown). We visualized the global cellular heterogeneity by pooling all the acquired data on $5.2x10^6$ cells and applying unsupervised hierarchical clustering with minimum spanning tree projection (SPADE), grouping the cells into a pre-defined number of nodes based on phenotypic similarity [15]. A dendrogram displayed the corresponding higher-order relatedness between those nodes (**Figure 1B**). The major branches in this dendrogram corresponded to $CD4^+$ T cells, $CD8^+$ T cells, TCRγδ cells, B cells, innate lymphocytes (referred to as $CD3^-$ $CD7^+$ cells hereafter) and myeloid cells (**Figure 1C**). The cell frequencies of these major cell lineages obtained through SPADE were confirmed by traditional gating procedures using two-parameter dot plots (**Table S3**).

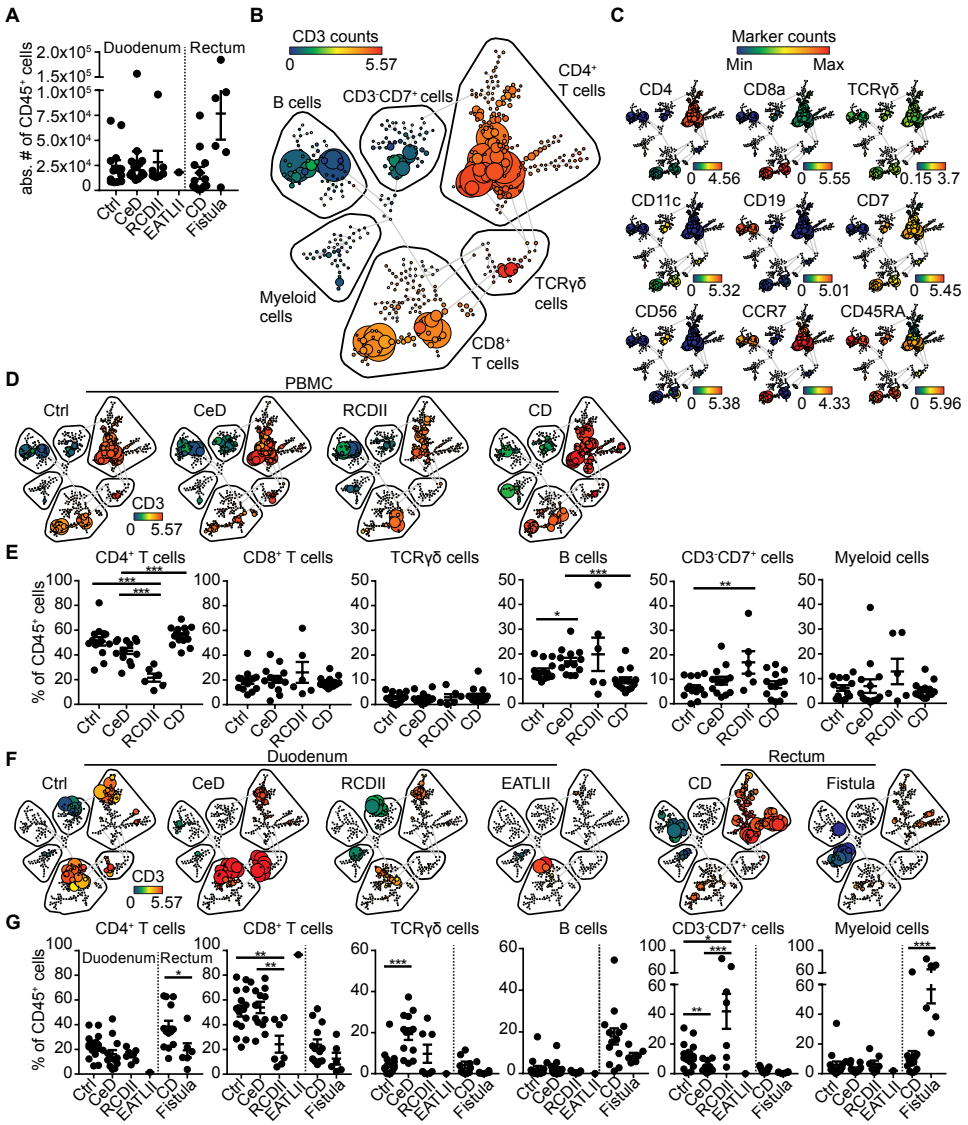
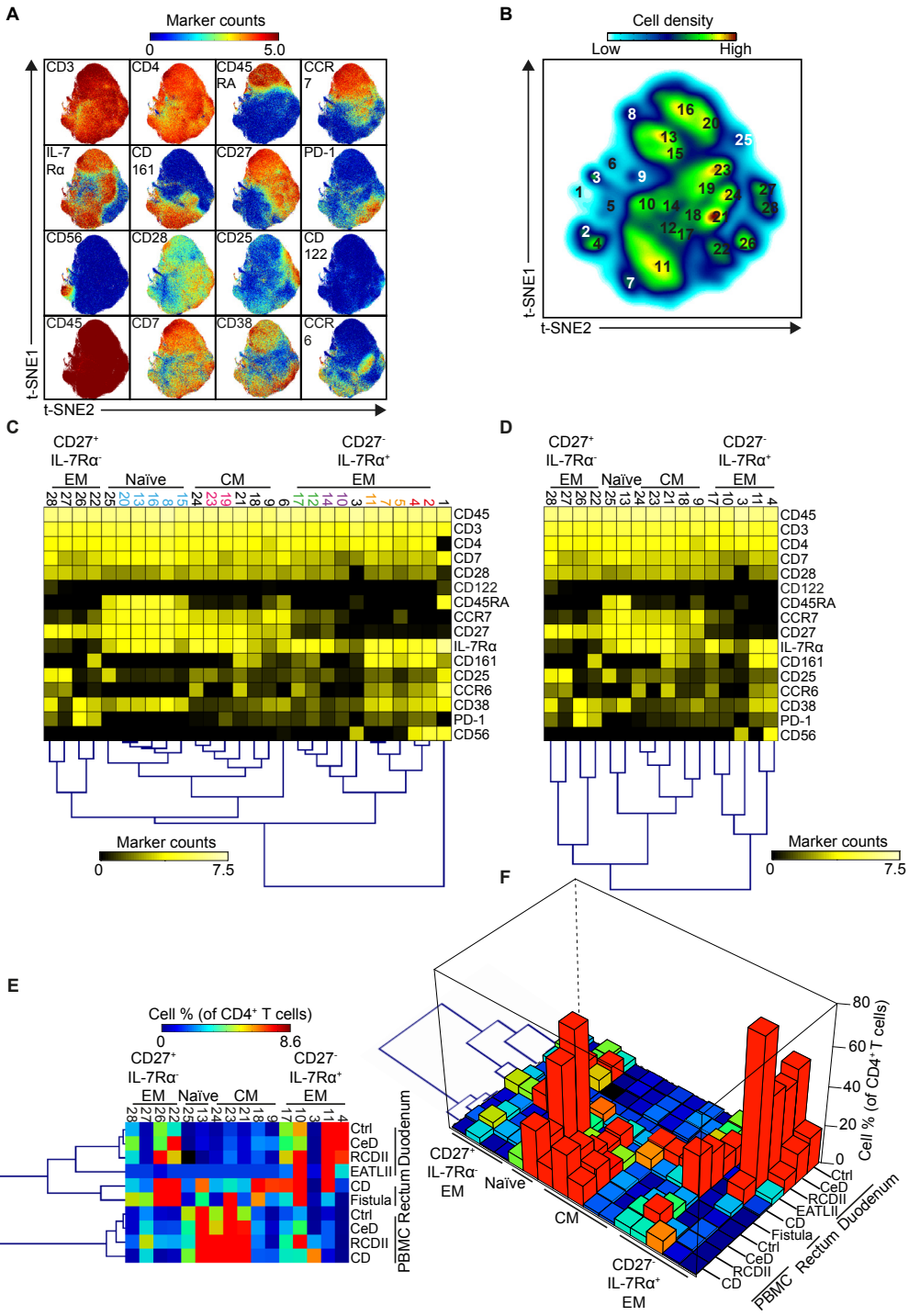


Figure 1. Collective SPADE analysis distinguishes major immune lineages in peripheral blood and intestine. (A) Live single CD45$^+$ absolute cell number acquired for 55 intestinal biopsies. (B) A SPADE tree of a PBMC sample after analysis of the combined 102 sample dataset containing 5.2 x 10^6 cells. Size of the nodes is proportional to the respective number of clustered cells. Color bars represent ArcSinh5-transformed values for CD3 marker expression. The identities of major immune lineages are annotated on the basis of lineage marker expression. (C) Color of the PBMC sample represents expression values for each marker as shown. (D) Representative SPADE trees showing an individual PBMC sample from a control and 3 patients with intestinal diseases. Color represents CD3 marker expression as described in panel B. (E) Comparisons of cellular frequencies for major immune lineages from 47 PBMC samples. (F) Representative SPADE trees showing an individual intestinal biopsy from a control and 5 patients with intestinal diseases. (G) Comparisons of cellular frequencies for major immune lineages from 55 intestinal biopsies. Data are plotted as single values (each data point represents an individual sample). *P <0.05; **P <0.01; ***P <0.001, using Mann-Whitney U test. Error bars show means ± s.e.m.


A
Marker counts
0
5.0
CD3
CD4
CD45 RA
CCR 7
IL-7 Rα
CD 161
CD27
PD-1
CD56
CD28
CD25
CD 122
CD45
CD7
CD38
CCR 6
t-SNE1
t-SNE2
B
Cell density
Low
High
t-SNE1
t-SNE2
C
CD27+ IL-7Rα- EM
Naïve
CM
CD27- IL-7Rα+ EM
CD45
CD3
CD4
CD7
CD28
CD122
CD45RA
CCR7
CD27
IL-7Rα
CD161
CD25
CCR6
CD38
PD-1
CD56
Marker counts
0
7.5
D
CD27+ IL-7Rα- EM
Naïve
CM
CD27- IL-7Rα+ EM
Marker counts
0
7.5
E
Cell % (of CD4+ T cells)
0
8.6
CD27+ IL-7Rα- EM
Naïve
CM
CD27- IL-7Rα+ EM
Ctrl
CeD
RCDII
EATLII
CD
Fistula
Ctrl
CeD
RCDII
CD
PBMC
Rectum
Duodenum
F
Cell % (of CD4+ T cells)
80
60
40
20
0
CD27+ IL-7Rα- EM
Naïve
CM
CD27- IL-7Rα+ EM
Ctrl
CeD
RCDII
EATLII
CD
Fistula
Ctrl
CeD
RCDII
CD
PBMC
Rectum
Duodenum

In general, the subset distribution between PBMC samples from controls and patients was quite similar although a decrease in numbers of $CD4^+$ T cells in patients with RCDII and some variability in the numbers of B cells and $CD3^-CD7^+$ cells was detected (**Figures 1D and 1E)**. In contrast, substantial differences were evident between the PBMC and intestinal samples. For example, duodenal $CD4^+$ T cell, $CD8^+$ T cell, $CD3^-CD7^+$ cell and rectal myeloid cell subsets were distinct from those in peripheral blood (**Figures 1D-1G**). In addition, the SPADE dendrograms revealed disease-associated signatures (**Figures 1D-1G**) exemplified by the disappearance of $CD3^-CD7^+$ cells and an increase in TCRγδ cells in CeD relative to the control duodenal biopsies. Also, compared to controls and patients with CeD, an increase in $CD3^-CD7^+$ cells in the duodenum of patients with RCDII was observed. Of note is the dominant presence of a $CD8^+$ T cell cluster in the duodenum of a patient with enteropathy-associated T cell lymphoma type II (EATLII). Finally, a highly diverse $CD4^+$ T cell compartment was found in rectal biopsies of patients with CD and a dominant presence of myeloid cells in perianal fistulas. Thus, this global analysis indicated that there are immune-system-wide differences in subset composition between peripheral blood and intestinal samples, and between duodenal samples from patients and controls.

t-SNE-ACCENCE analysis identifies 142 phenotypically distinct immune subsets

While SPADE analysis provides an overview of the heterogeneity and the relatedness of subsets within the major immune lineages it does not allow analysis at the single-cell level and consequently rare cells are difficult to visualize. Therefore, we applied t-Distributed Stochastic Neighbor Embedding (t-SNE) analysis [16, 17] which generates a two-dimensional map where cells with similar multidimensional phenotypes are placed close to each other, while maintaining single-cell resolution [12]. To ensure a similar impact of the cells from PBMC and intestinal samples on the t-SNE analysis the number of cells incorporated from those two compartments

Figure 2. t-SNE-ACCENSE analysis pipeline identifies tissue-specific $CD4^+$ T cell subsets. (A) Collective t-SNE dimensionality reduced $CD4^+$ T cell single-cell data ($4.7x10^5$ cells) derived from 102 samples analyzed are plotted. Every dot represents a single cell and the color of the cells indicates ArcSinh5-transformed expression values for a given marker analyzed. (B) A density map depicting the local probability density of cells as embedded in panel A, computed using a kernel based transformation. Numbers in this map represents centers of phenotypic subsets and were identified using a peak detection algorithm. (C) A heatmap summary of median ArcSinh5-transformed expression values of T cell markers expressed by 16 $CD4^+$ T cell subsets identified and hierarchical clustering of subsets with description of four categories. (D) A heatmap summary of average subset frequencies across tissues and disease states. Mean frequencies obtained from 102 samples analyzed. (E) A 3D heatmap as described in panel D. EM = effector memory and CM = central memory. A detailed heatmap showing 102 samples is described in **Figure S4**

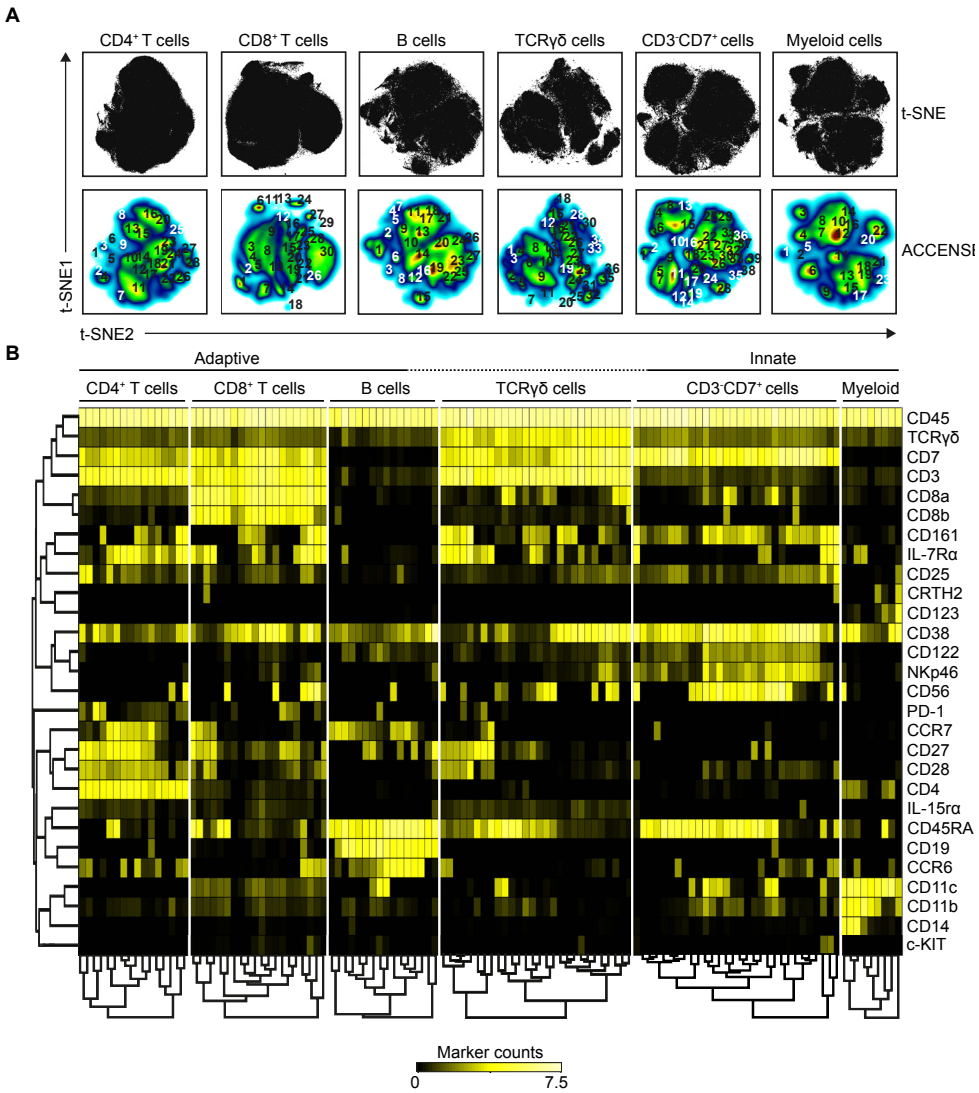


Figure 3. t-SNE-ACCENSE analysis delineates phenotypic distinct immune subsets in peripheral blood and intestine. (A) Identification of 142 cell subsets within the six major immune lineages through t-SNE-ACCENSE analysis. Cells from PBMCs were randomly sampled to match cell numbers with those from intestinal biopsies for each immune lineage individually. t-SNE plots are showing $4.7x10^5$ $CD4^+$ T cells, $9.3x10^5$ $CD8^+$ T cells, $1.8x10^5$ B cells, $1.8x10^5$ TCRγδ cells, $1.9x10^5$ $CD3^-CD7^+$ cells and $2.2x10^5$ myeloid cells of the combined 102 sample dataset. (B) Heatmap showing characterization of 119 cell clusters (16 $CD4^+$ T cell subsets, 20 $CD8^+$ T cell subsets, 16 B cell subsets, 28 TCRγδ cell subsets, 30 $CD3^-CD7^+$ cell subsets and 9 myeloid cell subsets). Shown are median ArcSinh5-transformed values of marker expression (black-to-yellow scale) and hierarchical clustering of markers and subsets within their major immune lineage

were matched. We applied the t-SNE approach for every major lineage individually, here showing the $CD4^+$ T cell compartment where over 440,000 cells were

incorporated in the analysis (**Figure 2**).

The t-SNE analysis revealed expected types of marker distributions on the $CD4^+$ T cells, such as broadly expressed markers (CD7, IL-7Rα), markers that were expressed by subpopulations of the cells (CD56, PD1, CD25) and markers that were co-expressed (CD45RA and CCR7), but also unanticipated distributions were revealed, like largely mutually exclusive marker expression patterns (i.e. CD27 and CD161) (**Figure 2A**). Next, we incorporated a kernel density-peak detection algorithm on the t-SNE map (ACCENSE) [14] which automatically identified 28 $CD4^+$ T cells subsets (**Figure 2B**) where each subset was defined by its marker expression profile (**Figure 2C)**.

To reduce the complexity implied by the ACCENSE analysis we merged computationally-derived subsets with highly similar expression profiles, which resulted in 16 cell clusters that each express a distinct set of markers (**Figure 2D**). They fell within four major $CD4^+$ T cell categories: naïve ($CD45RA^+CCR7^+$), $CD27^+$IL-7Rα$^-$ effector memory (EM; $CD45RA^-CCR7^-$), $CD27^-$IL-7Rα$^+$ EM and central memory (CM; $CD45RA^-CCR7^+$), and within those categories additional heterogeneity was present. For example, the highly similar $CD161^+CD27^-$IL-7Rα$^+$ EM cell subsets 4 and 11 were distinguished from each other by the expression of CD56 (**Figure 2D**). We next analyzed the subset distribution of the $CD4^+$ T cells in the various tissues included in the analysis by plotting the relative frequencies of the subsets for all samples analyzed (**Figures 2E and 2F**). In line with the SPADE analysis, the $CD4^+$ T cell subsets of the duodenum, rectum and PBMC samples clustered to distinct locations in the cell frequency heatmap, also when examining the 102 samples individually (**Figure S4**). Thus, we could effectively delineate cell populations in a data-driven manner, and this approach revealed distinct signatures in the cellular composition of the $CD4^+$ T cells in biopsy material and peripheral blood.

By applying the t-SNE-ACCENSE analysis to all 6 major cell lineages individually we identified 142 subsets in the entire immune system (**Figure 3A**), 23 of which contained rare cells that did not fulfill lineage phenotypic criteria (data not shown). The distinct phenotypes of the remaining 119 subsets are summarized in a heatmap (**Figure 3B**) where the subsets are clustered according to their phenotypic hierarchy within their lineage and clustered based on marker expression. The analysis identifies relatively few subsets within the B cell and myeloid compartments, likely due to the composition of the antibody panel which was designed to capture the heterogeneity of the $CD7^+$ lymphoid cells. In this context, our antibody panel identified many distinct subsets within the TCRγδ

and $CD3^{-}CD7^{+}$ immune lineages, a diversity that was even greater than that detected in the $CD4^{+}$ and $CD8^{+}$ T cell lineages despite the fact that the latter were generally more abundant in the samples included. Collectively, the combined t-SNE-ACCENSE approach on high-dimensional cytometry data can effectively identify phenotypic distinct subsets in an unbiased and data-driven manner, and the analysis indicates that the heterogeneity of the immune system is far greater than previously appreciated.

Visualization of cellular 'finger-print' signatures across tissues and disease states

To visualize the distribution of immune cells based on tissue-origin and disease state we used the t-SNE maps to deduce cellular 'finger-print-like' signatures of immune cells in the 6 major immune lineages (**Figure 4A**). The 'finger-print' gives a visual representation of the position of a collection of cells from particular (tissue) samples in the t-SNE map of the collective dataset. As such it gives a unified overview of the distinctness of cells within the samples analyzed stratified for tissues and disease states. In all 6 immune lineages, the duodenum, rectum and PBMC samples displayed a distinct cellular signature (**Figure 4A**). Moreover, on the basis of these signature maps, we were able to identify phenotypically distinct cell clusters that were either specifically present or abundant in certain diseases (highlighted by red boxes and arrows in **Figure 4A**). For example, the lineage $(Lin)^{-}$ $CD3^{-}CD7^{+}$ cells that expanded monoclonally in patients with RCDII (purple arrow; **Figure 4A**) were distinguished by the expression of CD45RA in 6 out of 7 patients (**Figures 4B and 4C**). Moreover, these aberrant Lin^{-} $CD3^{-}CD7^{+}$ cells were also detectable in PBMC samples of 3 out of 6 patients (**Figures 4A and 4D**), indicating a systemic spread of the pre-malignant cells. Similarly, a massive expansion of $CD56^{+}CD161^{+}$ $CD8^{+}$ T cells was observed in a patient with an established lymphoma of EATL-type 2 (green arrow; **Figure 4A**)**.** Rare cells (0.1-0.2% of $CD8^{+}$ T cells) displaying an identical phenotype were detected in 25% of the other duodenum samples and these may therefore represent the precursor from which the lymphoma arose. In addition, two distinct IL-$7R\alpha^{+}$ innate lymphoid cell type 3-like (ILC3-like) [18] cell clusters were identified that were either chemokine receptor $CCR6^{+}$ or $CCR6^{-}$ (blue arrow; **Figures 4A and 4B**). While the $CCR6^{+}$ ILC3-like cells represented 50% of the $CD3^{-}CD7^{+}$ cells in the rectum of CD patients, its $CCR6^{-}$ ILC3-like counterpart, which is associated with an inflammatory phenotype in CD [19], was more abundant in the fistulas (**Figure 4C**). Also, 94% of the myeloid cells within the inflammatory perianal fistulas (red arrow; **Figure 4A)** displayed a $CD11b^{+}CD11c^{+}$ dendritic cell-like phenotype (**Figure 4B**) and they comprised 50% of the accumulated immune infiltrate (**Figure 4C**).

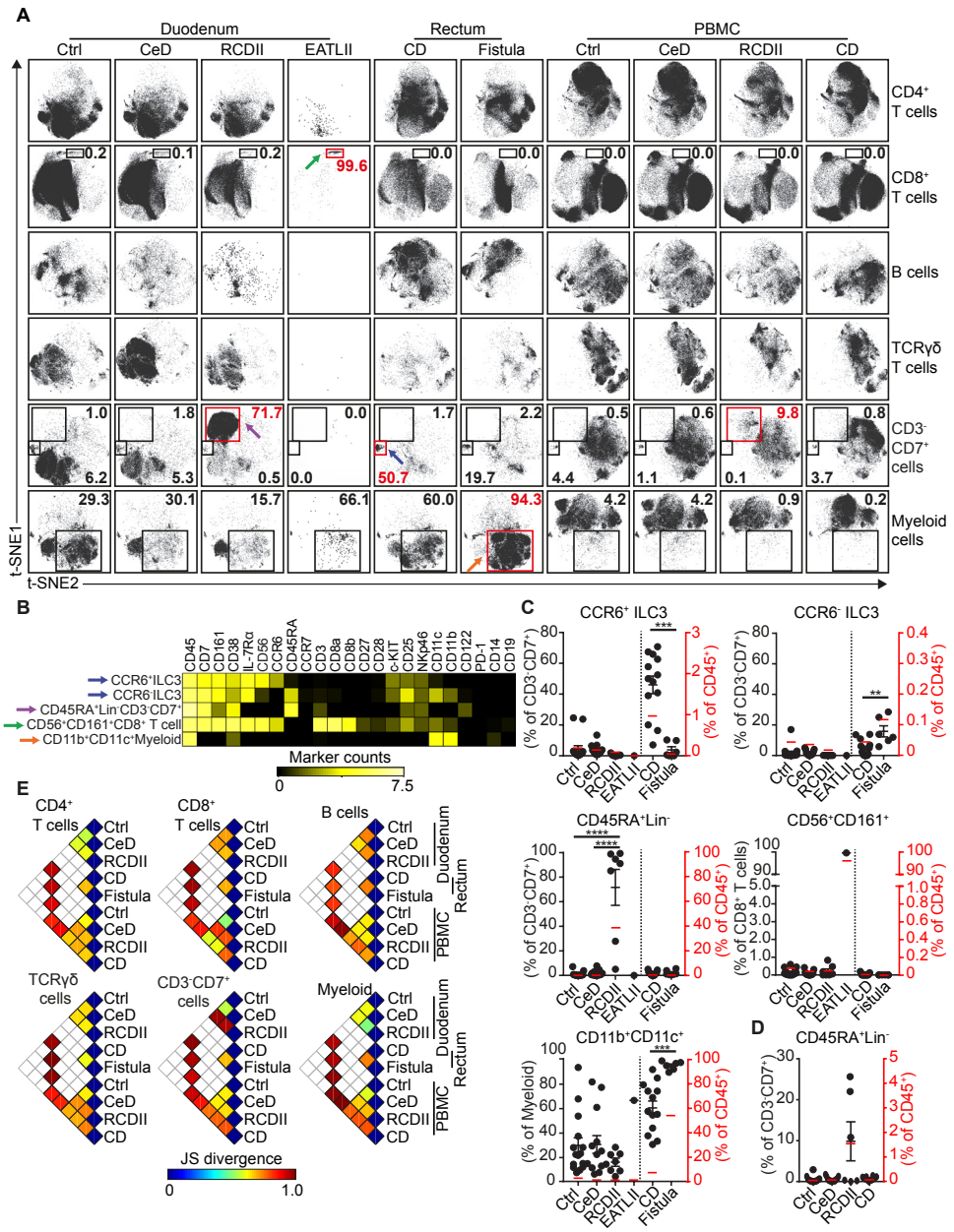


Figure 4. Stratified t-SNE plots identify distinct cellular 'finger-print' signatures across tissues and disease states. (A) Collective t-SNE dimensionality reduced single-cell data from all 102 samples analyzed are plotted showing six major immune lineages stratified for tissues and disease states. Red boxes and arrows indicate t-SNE location of phenotypically distinct disease-associated clusters of cells (B) Heatmap summary of median ArcSinh5-transformed expression values of cellular markers expressed by gated subsets and annotation for each subset. Arrows as in panel A. (C) Comparisons

of cellular frequencies for subsets from intestinal biopsies. (D) Comparisons of cellular frequencies for the CD45RA$^+$Lin$^-$ CD3$^-$CD7$^+$ subset from PBMC samples. (E) Pairwise Jensen-Shannon (JS) divergence plots of the collective t-SNE maps from all 102 samples analyzed showing six major immune lineages. A higher JS divergence value indicates more dissimilarity between a pair of t-SNE maps as shown in panel A. White squares indicate invalid comparisons. Data are plotted as single values. Red lines indicate mean value as % of CD45$^+$ cells. (each data point represents an individual sample). *P <0.05; **P <0.01; ***P <0.001; ****P <0.0001, using Mann-Whitney U test. Error bars show means ± s.e.m.

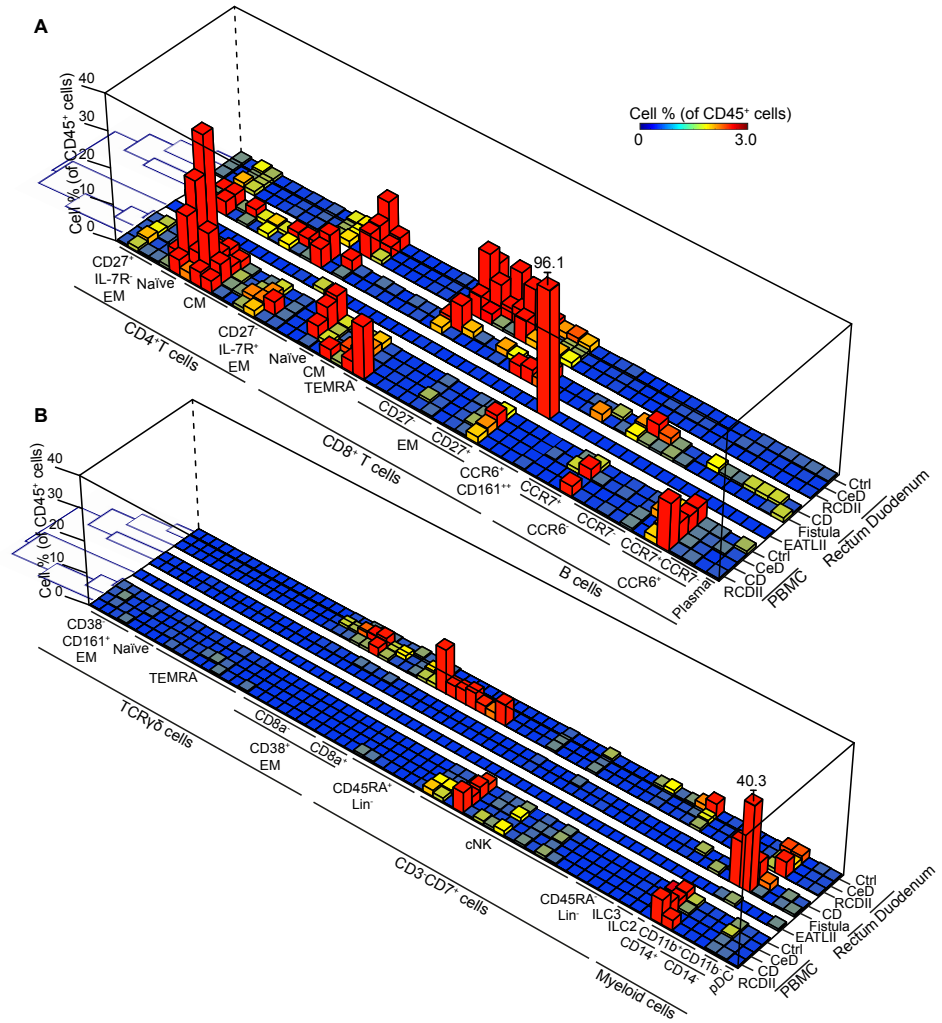


Figure 5. Immune system landscape visualizes subset composition. A 3-dimensional heatmap summary showing average frequencies of 119 immune subsets in the 102 biological samples combined with hierarchical clustering of samples and description of tissue type, disease state and biological assignment of the subsets. Color scale and z-axis indicate percentage of CD45$^+$ cells. EM = effector memory, CM = central memory, TEMRA = terminally differentiated, Lin = lineage, cNK = conventional NK cells, ILC = innate lymphoid cells and pDC = plasmacytoid dendritic cells.

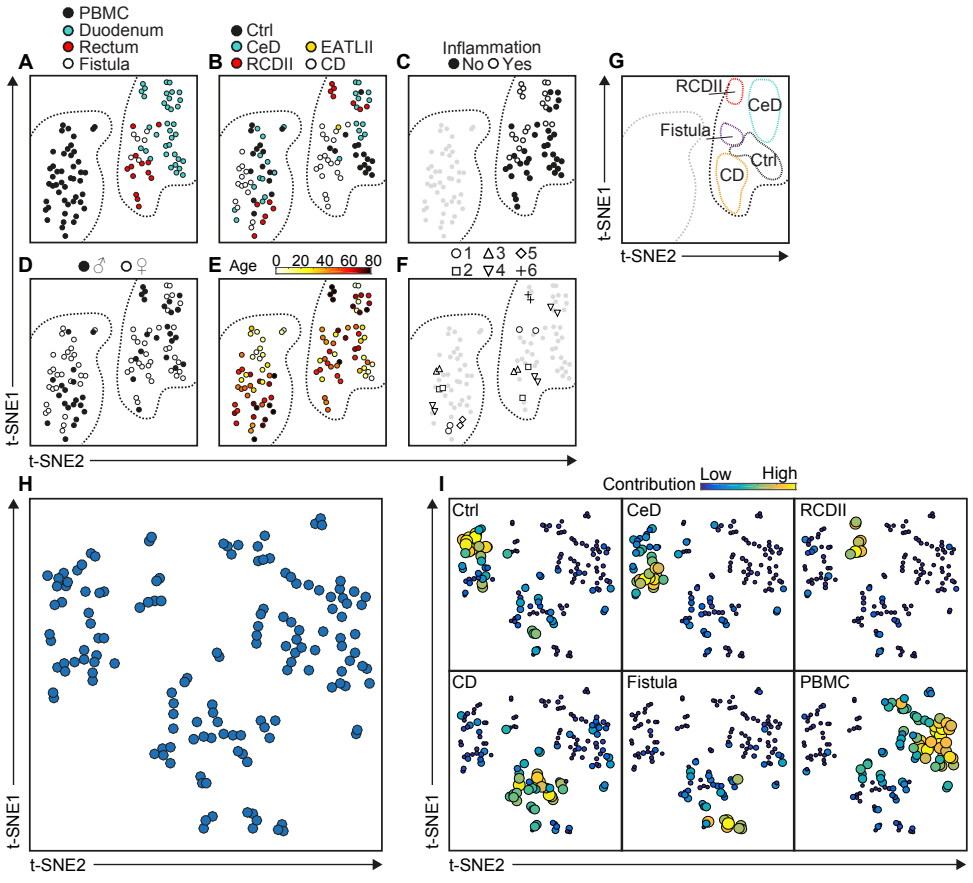
We next used the Jensen-Shannon (JS) divergence to quantify similarities and dissimilarities between pairs of t-SNE maps (**Figure 4E**). In these plots the JS divergences between samples from the intestine and peripheral blood were high for every major lineage (**Figure 4E**), indicative of dissimilarity of cellular signatures. In addition, the JS divergences reveal disease-associated cellular profiles exemplified by similarity of RCDII and CeD duodenal myeloid cells compared with those in duodenal controls, distinct $CD8^+$ T cells in RCDII patient blood and distinct B cells, $CD3^-CD7^+$ cells and myeloid cells in CD patient blood compared with other blood samples.

Furthermore, we visualized the immune composition as an immune landscape where the distribution of the subsets in the various tissue and blood samples is shown (**Figure 5**). A 2-dimensional representation of the immune landscape along with the phenotypes of the associated immune subsets is shown in **Figure S5**, visualizing the distinct cellular phenotypes of the immune subsets and their occurrence in the various tissues. Together these findings illustrate that in most of the major immune lineages, cellular subsets could be identified that were exclusively present or enriched in defined tissue samples only. In peripheral blood, cells were mainly defined by expression of interleukin 7 receptor α (IL-7Rα), CCR7, CD27 and CD28 for $CD4^+$ and $CD8^+$ T cells, CCR6 for B cells, CD56 for $CD3^-CD7^+$ cells and CD14 for myeloid cells. Similarly, many mucosal cells were defined by expression of CD161 for $CD4^+$ and $CD8^+$ T cells, CD38 for TCRγδ cells and IL-7Rα for $CD3^-CD7^+$ cells

Together these analyses demonstrate that by deducing cellular 'finger-print' signatures of immune cells we were able to visualize and quantify the immune subset distribution in the samples analyzed. In addition, we were able to identify immune subsets that are associated with disease states.

An integrated system-wide view of the immune system reveals disease-associated networks of immune subsets

Finally, we investigated whether the identified immune-system-wide cellular patterns could be integrated collectively and used to characterize samples according to tissue location or disease state by visualizing them in relation to several clinical variables. For this purpose we visualized the immune composition of all identified subsets from all included biological samples in a single graph by applying the t-SNE algorithm on cell frequency values. As expected, the PBMC and intestinal biopsy samples formed two distinct clusters (**Figure 6A**). In addition, the rectum and peri-anal fistula biopsies separated from the duodenal biopsies (**Figures 6A and 6B**) and the duodenal biopsies from patients with CeD and RCDII clustered away from the duodenal control biopsies (**Figures 6A and 6B**).



J

Top five ranked subsets contributing to the disease-specific t-SNE sample signatures

Nr	Ctrl	CeD	RCDII	CD	Fistula	PBMC
1	$CD3^-CD7^+$ 12	TCRγδ 9	$CD3^-CD7^+$13	CD4 18	B cell 7	CD4 13
	$CD45RA^-Lin^-$	$CD8a^-CD38^+$EM	$CD45RA^+Lin^-$	$CCR6^-$CM	$CCR6^-CCR7^+$	Naïve
2	$CD3^-CD7^+$17	TCRγδ 8	$CD3^-CD7^+$4	CD4 22	Other 15	CD4 21
	$CD45RA^-$cNK	$CD8a^-CD38^+$EM	$CD45RA^+Lin^-$	$PD\text{-}1^+CD27^+$EM	$CD19^-$ B cells	$CCR6^+$CM
3	CD4 11	TCRγδ 5	$CD3^-CD7^+$15	$CD3^-CD7^+$ 1	Myeloid 18	CD4 25
	$CD56^-CD27^-$EM	$CD8a^+CD38^+$EM	$CD45RA^+Lin^-$	$CCR6^+$ILC3	$CD14^-CD11b^+$	$CD25^+$Naïve
4	$CD3^-CD7^+$11	CD8 10	$CD3^-CD7^+$ 8	CD4 10	Other 12	CD4 27
	$CD45RA^-$ Lin^-	$CD27^-$EM	$CD45RA^+Lin^-$	$CCR6^-CD27^-$EM	$CD19^-$ B cells	$CD25^+CD27^+$EM
5	$CD3^-CD7^+$ 7	CD8 8	$CD3^-CD7^+$ 6	B cell 9	B cell 4	$CD3^-CD7^+$ 28
	$CD45RA^-$ Lin^-	$CD27^-$EM	$CD45RA^+Lin^-$	$CCR6^-CCR7^-$	$CCR6^-CCR7^+$	$CD11c^+$cNK

Figure 6. Integrated analysis of immune subset composition reveals tissue- and disease-associated clustering of biological samples. Collective t-SNE dimensionality reduced cell percentage data (as % of $CD45^+$ cells) of 142 subsets for 102 samples analyzed are plotted. Every dot represents a single sample and the color of the samples shows the corresponding clinical information: (A) tissue, (B) disease, (C) biopsy inflammation, (D) gender, (E) age in years, (F) samples from six patients that were sampled twice. The left dashed border represents the PBMC cluster and the right dashed border represents the intestinal cluster. (G) Deducing disease-specific signatures in the t-SNE map based on the clustering patterns of the samples. (H) Collective t-SNE dimensionality reduced cell percentage data (as % of $CD45^+$ cells) of 142 subsets for 102 samples analyzed are plotted. Every dot represents a single immune subset. The closer the subsets are together the more similar the cell frequency values are across

the samples. (I) Disease-specific subset signatures (average subset values per disease cluster in panel G, encoded with varying dot color and size). (J) Table showing top five ranked subsets contributing to the disease-specific t-SNE sample signatures (as shown in panel I) displaying both major lineage subset number and biological assignation. Lineage (Lin), conventional NK cell (cNK), effector memory (EM), central memory (CM) and innate lymphoid cells (ILC). Average cell frequency values are shown in **Figure 5.**

Also, the duodenal biopsies of the four RCDII patients with the most severe inflammation (**Table S2**) mapped far from the other duodenum samples (**Figures 6B and 6C**). Furthermore, three RCDII biopsies, two of which were from patients in remission, clustered with the other CeD biopsies, suggesting a persisting CeD immune profile. The inflammation state of the biopsies was reflected in the cluster structure of intestinal samples in general (**Figure 6C**), while this was not the case for gender (**Figure 6D**). Moreover, the age of the patients and controls from which the samples were derived is reflected in the clustering of the samples, particularly in peripheral blood (**Figure 6E**). The PBMC and intestinal samples from six patients that were biopsied twice, with a 3- to 6-months time interval, clustered tightly together (**Figure 6F**) highlighting the reproducibility and robustness of this unbiased approach.

In order to reveal which cellular subsets were associated with the disease-associated patterns (**Figure 6G**), we performed a second t-SNE analysis on the subsets of the same dataset (instead of the samples) visualizing networks of cellular subsets that determine disease-specificity (**Figures 6H and 6I**) and identified the top 5 ranked subsets contributing to these clustering patterns (**Figure 6J**). In this context, expected types of health- and disease-associated subsets were identified in the intestinal mucosa, such as $CD45RA^{-}Lin^{-}$ $CD3^{-}CD7^{+}$ cells in control individuals, $CD8a^{+}$ and $CD8a^{-}$ TCRγδ cells in CeD, $CD45RA^{+}Lin^{-}$ $CD3^{-}CD7^{+}$ cells in RCDII, $CCR6^{+}$ILC3 in CD and $CD14^{-}CD11b^{+}$ myeloid cells in fistulas. In addition, previously unidentified subset associations were revealed as well, such as $CD45RA^{-}$ cNK cells and $CD56^{-}CD27^{-}$EM $CD4^{+}$ T cells in controls, $CD27^{-}$EM $CD8^{+}$ T cells in CeD, $CRTH2^{+}$ myeloid cells in RCDII (ranked 6), $CCR6^{-}$CM and PD-$1^{+}CD27^{+}$EM $CD4^{+}$ T cells in CD, and $CCR6^{-}CCR7^{+}$ B cells in the fistulas. Thus, by integrating data-driven approaches highly specific disease-associated immune signatures across innate and adaptive major lineages in the intestine were readily identified. Figure S6 gives an overview of the developed integrated analysis pipeline developed in the current study.

Discussion

Mass cytometry offers the opportunity to simultaneously analyze dozens of single-cell markers on complex cellular samples resulting in highly complex datasets. Conventional approaches for flow cytometry data analysis are not suitable for such datasets, suffer from individual user bias, and require prior knowledge of the cell type of interest. SPADE [13] was originally applied to handle mass cytometric data. More recently, an unbiased analysis pipeline has been developed combining t-SNE [12, 16] and ACCENSE [14] with mass cytometry to visualize and delineate phenotypically distinct subsets [20]. In the current study we used a 32 antibody panel that was specifically designed to detect heterogeneity within the major adaptive and innate cell lineages. We applied this antibody panel to a variety of PBMC and intestinal biopsy samples and combined this with the newly available unbiased computational approaches to unravel the complexity of the human mucosal immune system. We used the Barnes-Hut implementation of t-SNE, a recently developed, computationally efficient t-SNE optimization algorithm [17] to accommodate the large datasets. In addition, we provide novel applications of the t-SNE-based analysis allowing the visualization of cellular 'finger-print' signatures of immune cells, and by clustering samples based on their immune composition while visualizing the cluster-contributing subsets in parallel to highlight tissue- and disease-associated patterns.

Our results demonstrate that the mass cytometry-based analysis was robust and reproducible as identical control PBMC samples that were included during the entire 9 month study period provided highly similar results. Moreover, we obtained biopsy and PBMC specimens from a number of patients twice with a time interval between 3 to 6 months and in the final sample visualization analysis these specimens clustered close together, demonstrating a high degree of reproducibility. Also, we readily observed changes in the composition of the immune compartment that are known to correlate with disease such as the increase in TCRγδ cells in CeD [21].

Both the SPADE and the combined t-SNE-ACCENSE analysis demonstrate that duodenal, rectal and PBMC samples grouped into different clusters, to a large extent due to substantial differences in the $CD4^+$ and $CD8^+$ T cell compartments. In total, 142 subsets in the immune system were defined of which 120 displayed distinct marker expression profiles on the basis of which these subsets were delineated. Compared to the $CD4^+$ and $CD8^+$ T cell lineages, a larger degree of heterogeneity was detected in the TCRγδ and $CD3^-CD7^+$ immune lineages. While we cannot exclude the possibility that by the use of another antibody panel additional heterogeneity within the $CD4^+$ and $CD8^+$ lineages may be revealed, the distinct marker expression profiles of the TCRγδ and $CD3^-CD7^+$ subsets may

correlate with distinct and potentially location-specific functional properties. This will be the subject of future investigations. Thus, based on 28 markers mass cytometry visualized system-wide cellular differences in subset composition in the samples obtained from the various anatomical locations. Further studies including control rectum samples are required to determine how these differences relate to the anatomical site and/or disease state.

Recent studies have described a crucial role for ILCs in CD [22]. They reside mainly in mucosal tissues and are functionally specialized cells characterized by the expression of lineage-defining transcription factors. Even without these markers in our antibody panel we were able to distinguish ILC-like subsets on the basis of cell surface phenotype through machine-learning cell cluster detection approaches. We identified a $CRTH2^+$ ILC2-like subset in PBMC samples and in line with previous reports such cells were not found in intestinal samples [22]. Moreover, we detected an ILC3-like subset that was exclusively present in rectum biopsies of CD patients in remission. This subset corresponds with the IL-22-producing anti-inflammatory $CD25^-CD56^+$ ILC subset previously found in the intestine of CD patients [19, 22]. Moreover, we observed that their $CD56^-$ ILC counterpart was enriched in fistulas, and these cells were previously shown to produce the inflammatory cytokine IL-17A [19]. The $CD56^-$-ILC3-like cells in fistulas unexpectedly showed expression of CD11c, a marker that has been used as 'dump channel' in the ILC field. Those cells may thus have been discarded from datasets of previous studies, highlighting the importance of unbiased data-driven approaches as used in the current study. In addition, the $CD4^+$ T cell compartment in the CD biopsies was highly heterogeneous. Further studies are required to determine a possible relationship with the highly variable disease symptoms in CD.

Mucosal lymphoid malignancies were readily detected in patients with RCDII and in a patient with an EATL-type 2 lymphoma, along with the distinct cellular phenotypes that distinguish these lymphoma's. The latter information could be used to identify the potential precursor cells in the healthy mucosa from which these malignancies are likely to originate. In the case of RCDII these precursors are $Lin^-CD3^-CD7^+$ cells, confirming previous results [23]. Moreover, our current analysis indicates that these precursors can be distinguished from their malignant counterpart by the lack of expression of CD45RA. Conversion of this subset to a CD45RA positive phenotype may thus predict development of RCDII and could constitute a novel prognostic marker. Also, extremely low numbers of $CD56^+CD161^+CD8a^+CD8b^+$ T cells were found in mucosal biopsies of healthy individuals, a phenotype which matches that of the $CD8^+$ T cell malignancy in a patient with EATLII. Thus, the t-SNE analysis is highly suitable for the identification of mucosal malignancies

and their likely precursors in healthy individuals, information that may be used to develop therapeutic approaches based on cellular characteristics.

Compared to control duodenal samples we observed the disappearance of $CD3^-CD7^+$ cells and the increase in TCRγδ cells in CeD, both well-described disease hallmarks [21, 23]. In the global analysis of the entire cell frequency dataset this resulted in the formation of clusters that distinguish duodenal biopsies derived from CeD patients from those of controls. By transposing the cell frequency dataset the disease cluster-associated subsets and their relative contribution to the clustering could be visualized and quantified. This tSNE application thus provided detailed information on the disease-associated networks of immune subsets. The identification of mucosal immune signatures that correlated with health and disease may potentially lead to the development of unbiased diagnostic procedures based on a single mass cytometric analysis.

Perianal fistulas in CD remain a substantial clinical challenge, causing pain, discharge, and abscess formation [24]. Achieving complete fistula healing is difficult and accompanied by multiple relapses, and despite the best available therapies durable remission rates of perianal fistulas remain disappointingly low [7]. In this respect it will be important to unravel the function of the $CD11b^+CD11c^+$ myeloid cells that dominate in the fistula, where the immune composition is distinct from that in the adjacent rectum biopsies. This will be addressed in future studies.

In such follow-up studies our approach can be further refined as four antibodies included in the original 32 antibody panel (CD103, IL21-R, CD34 and TCRαβ) were not informative. In particular the inclusion of antibodies specific for lineage-defining transcription factors and cytokines may provide further insight into the relationship between cell lineages and their function in the mucosal immune system. Moreover, the inclusion of an alternative metal reporter for the CD103-specific antibody in future studies may allow discrimination between cells derived from the epithelium and the lamina propria. Also, with regard to CD and perianal fistula we could not draw definitive conclusions regarding disease-specific changes as we lacked healthy control rectum samples. It would thus be highly valuable to characterize mucosal biopsies obtained from various intestinal locations within the same patients and controls, allowing direct comparisons. Finally, by combining the analysis of the mucosal immune system with an analysis of the stromal cell compartment a more integrated view of disease-specific changes may be obtained, optimizing opportunities to develop more effective personalized treatment modalities.

In conclusion, the mass cytometric analysis of the mucosal immune system

revealed heterogeneity that was greater than previously appreciated. Also, our results indicate that disease-specific leukocytes reside mainly in the affected organ and are much less readily detectable in PBMC. The identification of disease-associated changes in immune composition offers opportunities to determine cellular parameters that correlate with disease and predict response to treatment, an important step towards personalized and cost-effective treatment.

Experimental procedures

Human Samples

Samples were collected from patients who were undergoing routine diagnostic endoscopies, and the curettage material of perianal fistulas were obtained at surgical intervention. The clinical characteristics of the patients are shown in **Table S2**. All samples were obtained after informed consent, medical ethical commission approval, in accordance with the local ethical guidelines of the VU Medical Center in Amsterdam (adult duodenal biopsies) and the Leiden University Medical Center (pediatric duodenal biopsies, CD rectum biopsies, and perianal fistulas), and in accordance with the declaration of Helsinki.

Isolation of cells from intestinal and PBMC samples

Cells from the epithelium were isolated from two or three intestinal biopsies by treatment with 10 mL of HBSS (Sigma-Aldrich, St. Louis, United States) containing 1 mM EDTA (Merck, Darmstadt, Germany) under rotation for 2 hr at 37°C. To obtain cells from the lamina propria, the biopsies were washed with PBS containing 0.5% fetal calf serum (FCS) and incubated with 5 mL of a collagenase mix containing IMDM culture medium (Lonza, Basel, Switzerland) with 20% FCS, 1,000 U/mL collagenase IV (Worthington, Lakewood, United States) and 10 mg/mL DNAseI grade II (Roche Diagnostics, Basel, Switzerland) for 2 hr at 37°C. The cell suspension was then filtered through a 70 µm nylon cell strainer and centrifuged in 0.5% FCS/PBS. Curettage material of perianal fistulas obtained at surgical intervention were minced with fine scissors and incubated with 10 mL of HBSS containing 1 mM EDTA under rotation for 2 hr at 37°C. Peripheral blood mononuclear cells (PBMC) were isolated from up to 5 mL of freshly drawn heparin anticoagulated blood using Ficoll-Paque™ density-gradient centrifugation. PBMC samples from CD patients were cryopreserved and stained after thawing. Cell suspensions were washed with 0.5% FCS/PBS and kept at 4°C until antibody staining.

Antibodies

Antibodies, manufacturers, and concentrations are listed in **Table S1**. Primary antibody metal-conjugates were either purchased or conjugated using a total of 100 µg of carrier-free formulations of purified antibody combined with the MaxPar X8 antibody labeling kit (Fluidigm Sciences, Toronto, Canada) according to the manufacturer's instruction. Following conjugation, antibodies were diluted to 200 µL in Candor PBS Antibody Stabilization Buffer (Candor Bioscience GmbH, Wangen, Germany) and stored at 4°C.

Antibody staining and data acquisition

Procedures for mass cytometry antibody staining and data acquisition were carried out as previously described [13]. Briefly, directly after biopsy processing cells were resuspended in cell staining buffer (CSM; 1x PBS with 0.5% bovine serum albumin and 0.02% sodium azide, Fluidigm Sciences) and incubated with 1 mL of 1:500 diluted 500 µM rhodium DNA intercalator (Fluidigm Sciences) for 15 min to stain dead cells at room temperature (rT). Cells were washed with CSM and surface stained for 45 min at rT with a mixture of metal isotope-conjugated antibodies using predetermined concentrations (**Table S1**). Antibody staining reactions were performed in 100 µL final volume. After staining, cells were washed twice with CSM and then resuspended in 1 mL of 1:1000 diluted 125 µM iridium DNA intercalator (DVS Sciences) in Fix and Perm Buffer (PBS with 1.6% paraformaldehyde, Fluidigm Sciences) for 45 min at rT to discriminate single cells. Cells were stored overnight at 4°C. Finally, cells were washed twice in CSM and once in distilled water at rT. Prior to data acquisition, cell pellets were diluted in distilled water containing 1:10 diluted EQ Four Element Calibration Beads (Fluidigm Sciences) to the concentration of $0.4x10^6$ cells/mL to achieve an acquisition rate of 500 events/s on the CyTOF 2™ mass cytometer (Fluidigm Sciences) [25]. CyTOF data were acquired and analyzed on-the-fly, using dual-count mode and noise-reduction on. All other settings were either default settings or optimized with tuning solution, as instructed by Fluidigm Sciences. After data acquisition, the mass bead signal was used to normalize the short term signal fluctuations with the reference EQ passport P13H2302 during the course of each experiment and the bead events were removed [26].

Data analysis

SPADE analyses were performed as described [15] with 500 target number of nodes and 10% of target down-sampled events using the implementation in Cytobank [27]. Data from exported FCS files of major immune lineages as delineated by SPADE (**Figure 1)** were transformed using hyperbolic arcsin with a cofactor of

5. Because the number of cell events varied greatly between PBMC and intestinal biopsies, the contribution of intestinal and PBMC cells were normalized to a 1:1 ratio for each immune lineage and up to 10,000 events per sample was used. After down-sampling, the cumulative dataset per immune lineage was subjected to t-SNE dimensionality reduction. t-SNE was carried out using the Barnes-Hut implementation of t-SNE, a recent developed, computationally efficient t-SNE optimization algorithm (obtained from L.J.P. van der Maaten) [17] to accommodate the large volumes of our clinical data. t-SNE was run with a default perplexity of 30. Cellular 'finger-print' signatures of immune cells and marker expression color overlays of t-SNE maps were generated with Cyt [12]. We used the Jensen-Shannon (JS) divergence to quantify the similarity between t-SNE maps. After converting t-SNE maps into 2-dimensional probability density functions, the similarity between two maps is quantified as the JS divergence between their corresponding probability density functions. We used the base 2 logarithm in the JS divergence computation, which results in a continuous range of JS divergence values between 0 (for identical distributions) and 1 (for fully disjoint distributions). The density-peak detection algorithm to identify phenotypically distinct subsets was carried out with ACCENSE [14], using the two coordinates of the t-SNE map for each cell as input. The density-based clustering first searches for the optimal bandwidth, followed by estimating the kernel density that allows the detection of density peaks. The respective amount of subpopulations identified per major lineage was based on the calculated optimal kernel bandwidth. Two-dimensional gating analysis was done using Cytobank [27]. Median intensity values of markers were calculated and visualized via plotting heatmaps. Hierarchical clustering dendrograms of heatmaps were produced using Pearson Correlation and average linkage clustering with MultiExperiment Viewer (www.tm4.org). Numbers of cells in different immune subsets were counted for each sample and percentages of each subset were calculated. t-SNE coordinates, ACCENSE subset number, and sample coding tags were added to FCS files as additional parameters to allow aggregate data analysis and visualization. The sample t-SNE map (**Figure 6**) was computed with the fractions of the total cell count per subtype (as % of $CD45^+$ cells) as input variables. Standard t-SNE pre-processing was applied: the data matrix was normalized by centering each variable to zero mean, and scaling to unit vector length. In the sample map, a reprojection of the data on a reduced set of high-variance principal components (PCs) was performed. The component scores of the 10 highest variance PCs were used as input to the t-SNE. To reduce sensitivity to local optima, map construction was repeated 100 times with different randomly generated initial maps and the map with the minimal t-SNE error metric (Kullback Leibler divergence per data point) was selected. The subset t-SNE map

in **Figures 6H and 6I** was computed by transposing the normalized datamatrix. This switches the role of samples and variables, hence subsets with similar profiles across the population end up close together in the map. The t-SNE perplexity parameter was set to 10% of the number of data points in each map, i.e. 10 for the sample maps in **Figures 6A-6F**, and 15 for the subset map in **Figure 6H and 6I**. Average subset values were computed per disease subgroup as identified in the sample t-SNE map in **Figure 6G**, and displayed per subgroup.

Author contributions

V.v.U. and F.K. conceived the study and wrote the manuscript. V.v.U. performed most experiments with the help of N.L. Moreover, V.v.U., N.L., M.T., T.H., B.L. and J.B. analyzed the data and provided conceptual input. I.M., A.M.J., H.V., L.M. and C.M. provided clinical material. All authors discussed the results and commented on the manuscript.

Acknowledgements

We thank the patients and their families. The research leading to these results has received funding from the Leiden University Medical Center, The Netherlands Organization for Scientific Research (ZonMW grant 91112008), a World Wide Cancer Research grant (14-0245), the European Union Seventh Framework Programme (FP7/2007-2013, grant agreement no. 604102), the Dutch Technology Foundation STW (project 12720: "VANPIRE" under the STW IMAGENE perspective program), and DigestScience Foundation. We thank Drs. F.J.T. Staal, F.A. Ossendorp, M.W. Schilham, M. Yazdanbakhsh and W.E. Fibbe for critical review of the manuscript.

The authors declare no competing financial interests.

References

1. Kappelman, M.D., Moore, K.R., Allen, J.K. & Cook, S.F. Recent trends in the prevalence of Crohn's disease and ulcerative colitis in a commercially insured US population. *Dig Dis Sci* **58**, 519-525 (2013).

2. Rubio-Tapia, A., Ludvigsson, J.F., Brantner, T.L., Murray, J.A. & Everhart, J.E. The prevalence of celiac disease in the United States. *Am J Gastroenterol* **107**, 1538-1544; quiz 1537, 1545 (2012).

3. Al-Toma, A., Verbeek, W.H., Hadithi, M., von Blomberg, B.M. & Mulder, C.J. Survival in refractory coeliac disease and enteropathy-associated T-cell lymphoma: retrospective evaluation of single-centre experience. *Gut* **56**, 1373-1378 (2007).

4. Pascual, V. et al. Inflammatory bowel disease and celiac disease: overlaps and differences. *World J Gastroenterol* **20**, 4846-4856 (2014).

5. Randall, C.W. et al. From historical perspectives to modern therapy: a review of current and future biological treatments for Crohn's disease. *Therap Adv Gastroenterol* **8**, 143-159 (2015).

6. Schwartz, D.A. et al. The natural history of fistulizing Crohn's disease in Olmsted County, Minnesota. *Gastroenterology* **122**, 875-880 (2002).

7. Molendijk, I., Nuij, V.J., van der Meulen-de Jong, A.E. & van der Woude, C.J. Disappointing durable remission rates in complex Crohn's disease fistula. *Inflamm Bowel Dis* **20**, 2022-2028 (2014).

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

9. Verbeek, W.H. et al. Flow cytometric determination of aberrant intra-epithelial lymphocytes predicts T-cell lymphoma development more accurately than T-cell clonality analysis in Refractory Celiac Disease. *Clin Immunol* **126**, 48-56 (2008).

10. Geremia, A., Biancheri, P., Allan, P., Corazza, G.R. & Di Sabatino, A. Innate and adaptive immunity in inflammatory bowel disease. *Autoimmun Rev* **13**, 3-10 (2014).

11. Bandura, D.R. et al. Mass cytometry: technique for real time single cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry. *Anal Chem* **81**, 6813-6822 (2009).

12. Amir el, A.D. et al. viSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. *Nat Biotechnol* **31**, 545-552 (2013).

13. Bendall, S.C. et al. Single-cell mass cytometry of differential immune and drug responses across a human hematopoietic continuum. *Science* **332**, 687-696 (2011).

14. Shekhar, K., Brodin, P., Davis, M.M. & Chakraborty, A.K. Automatic Classification of Cellular Expression by Nonlinear Stochastic Embedding (ACCENSE). *Proceedings of the National Academy of Sciences of the United States of America* **111**, 202-207 (2014).

15. Qiu, P. et al. Extracting a cellular hierarchy from high-dimensional cytometry data with SPADE. *Nature biotechnology* **29**, 886-891 (2011).

16. van der Maaten, L. & Hinton, G. Visualizing Data using t-SNE. *Journal of Machine Learning Research* **9**, 2579-2605 (2008).

17. van der Maaten, L. Accelerating t-SNE using Tree-Based Algorithms. *Journal of Machine Learning Research* **15**, 3221-3245 (2014).

18. Spits, H. & Cupedo, T. Innate lymphoid cells: emerging insights in development, lineage relationships, and function. *Annu Rev Immunol* **30**, 647-675 (2012).

19. Geremia, A. et al. IL-23-responsive innate lymphoid cells are increased in inflammatory bowel disease. *J Exp Med* **208**, 1127-1133 (2011).

20. Becher, B. et al. High-dimensional analysis of the murine myeloid cell system. *Nat Immunol* **15**, 1181-1189 (2014).

21. Spencer, J., Isaacson, P.G., Diss, T.C. & MacDonald, T.T. Expression of disulfide-linked and non-disulfide-linked forms of the T cell receptor gamma/delta heterodimer in human intestinal intraepithelial lymphocytes. *Eur J Immunol* **19**, 1335-1338 (1989).

22. Hazenberg, M.D. & Spits, H. Human innate lymphoid cells. *Blood* **124**, 700-709 (2014).

23. Schmitz, F. et al. Identification of a potential physiological precursor of aberrant cells in refractory coeliac disease type II. *Gut* **62**, 509-519 (2013).

24. Kamm, M.A. & Ng, S.C. Perianal fistulizing Crohn's disease: a call to action. *Clin Gastroenterol Hepatol* **6**, 7-10 (2008).

25. Ornatsky, O.I. et al. Development of analytical methods for multiplex bio-assay with inductively coupled plasma mass spectrometry. *J Anal At Spectrom* **23**, 463-469 (2008).

26. Finck, R. et al. Normalization of mass cytometry data with bead standards. *Cytometry A* **83**, 483-494 (2013).

27. Kotecha, N., Krutzik, P.O. & Irish, J.M. Web-based analysis and publication of flow cytometry experiments. *Curr Protoc Cytom* **Chapter 10**, Unit10 17 (2010).

Supplemental information

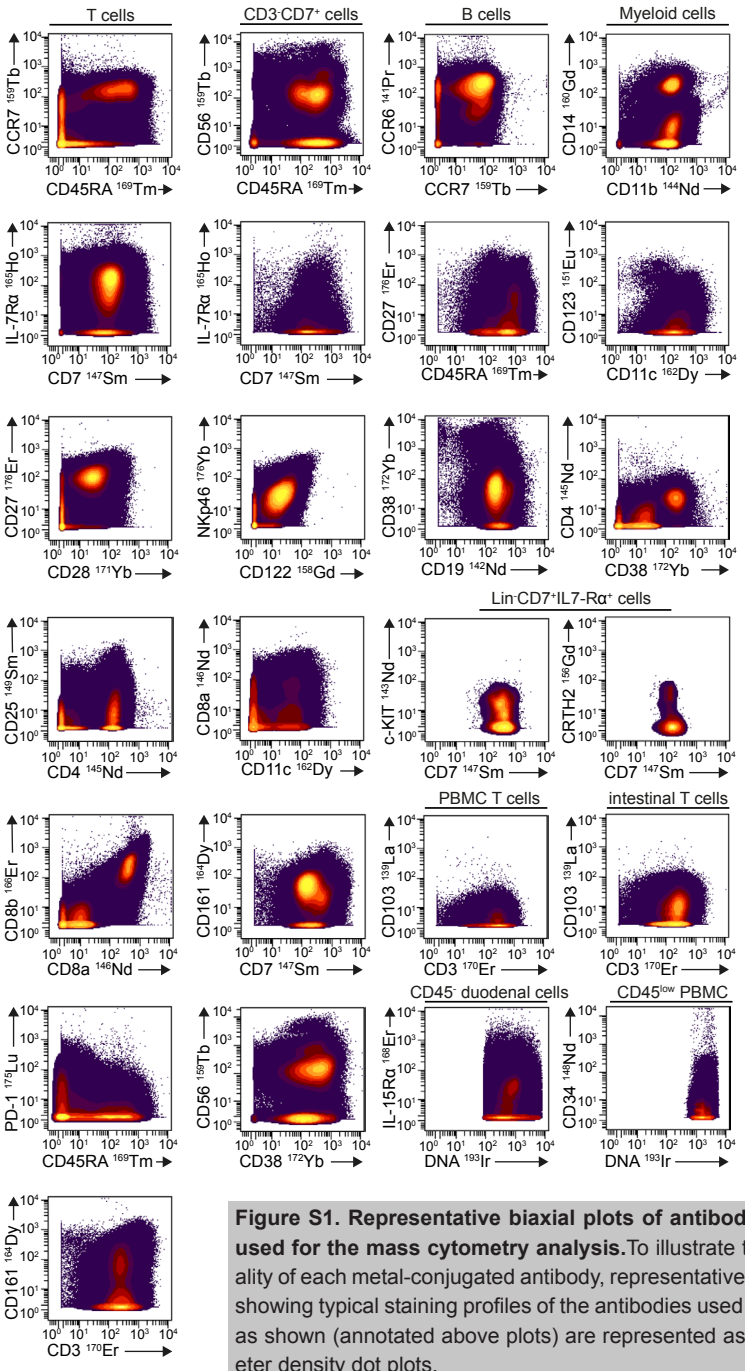


Figure S1. Representative biaxial plots of antibody stainings used for the mass cytometry analysis. To illustrate the functionality of each metal-conjugated antibody, representative biaxial plots showing typical staining profiles of the antibodies used. Cells gated as shown (annotated above plots) are represented as two-parameter density dot plots.

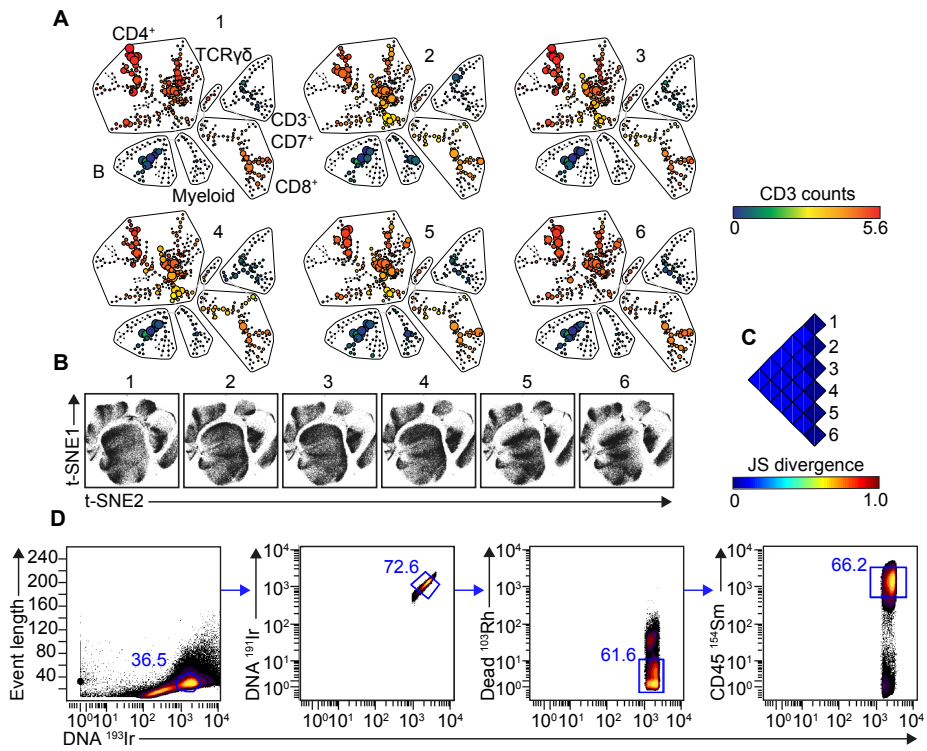


Figure S2. SPADE and t-SNE analysis of the internal peripheral blood control. (**A**) Six PBMC samples from the same individual taken at different time points during the entire study period are shown after collective SPADE analysis of live single CD45+ cells. Size of the nodes is proportional to the respective number of clustered cells. Color bars represent ArcSinh5 transformed values for CD3 marker expression. The identities of major immune lineages are annotated on the basis of lineage marker expression (CD4+ = CD4+ T cells, CD8+ = CD8+ T cells and B = B cells) (for details SPADE analysis see Figure 1) (**B**) Collective t-SNE dimensionality reduced single-cell data containing live single CD45+ cells from 6 PBMC samples as described in panel a analyzed are plotted. Each sample was down-sampled to 50,000 events prior to the t-SNE analysis. (for details t-SNE analysis see Figures 2 and 4) (**C**) Pairwise Jensen-Shannon (JS) divergence plot of the collective t-SNE maps from the 6 PBMC samples analyzed. A lower JS divergence value indicates more similarity between a pair of t-SNE maps as shown in panel b. (**D**) Live, single CD45+ cells gating strategy. Representative mass cytometry plots of a duodenum biopsy showing sequential gates with percentages. Event length is a mass cytometric measurement for the amount of scans it took to acquire a given ion cloud.

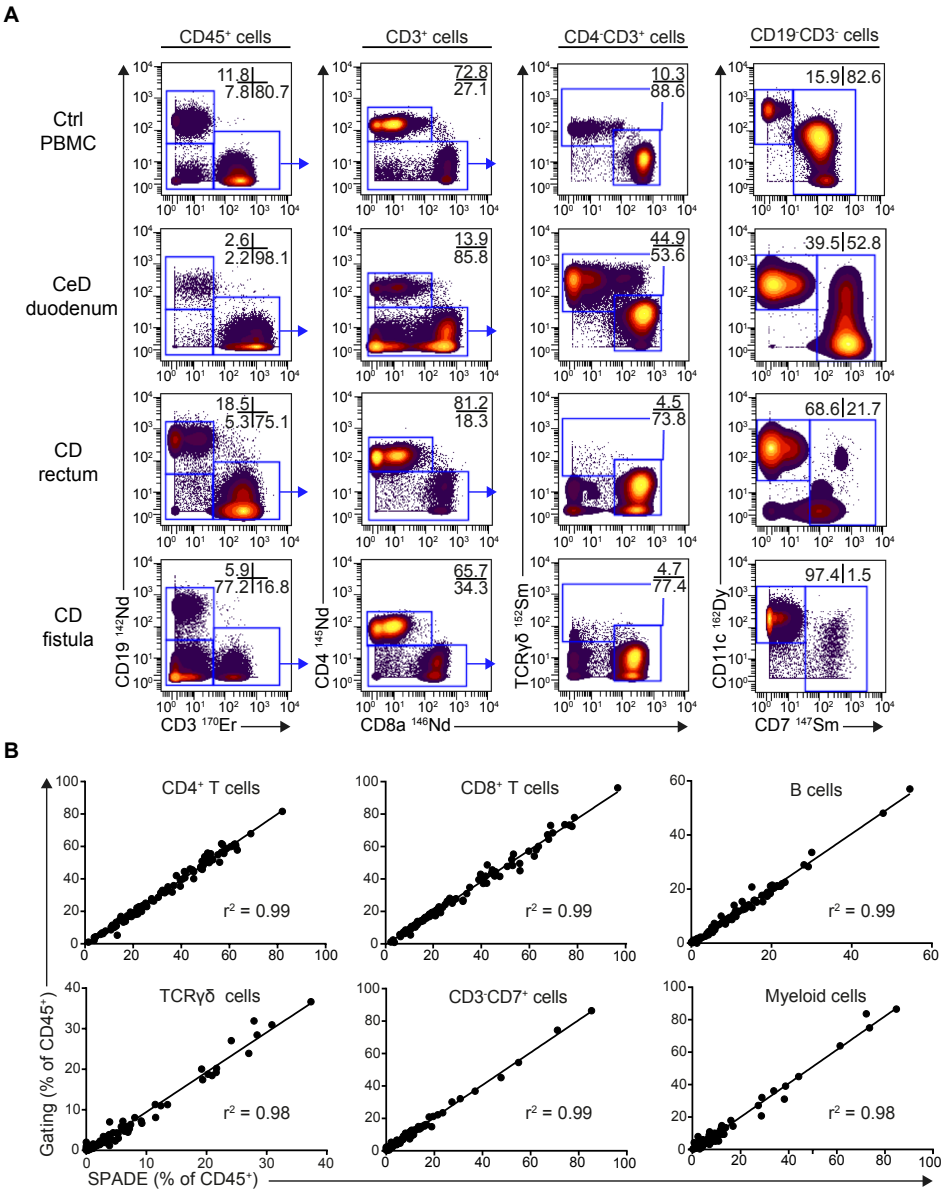


Figure S3. Traditional gating yields similar frequencies of major immune lineage cells as with SPADE analysis. (A) Representative samples are shown to demonstrate the gating strategy applied to identify the major immune lineages: $CD4^+$ T cells, $CD8^+$ T cells, B cells, TCRγδ cells, $CD3^-CD7^+$ cells and myeloid cells. Cells gated as shown (annotated above plots) are represented as two parameter density dot plots. (B) Correlation of cell frequencies obtained through traditional gating and with SPADE analysis are shown for the 102 samples as analyzed with linear regression. A dot represents a single sample.

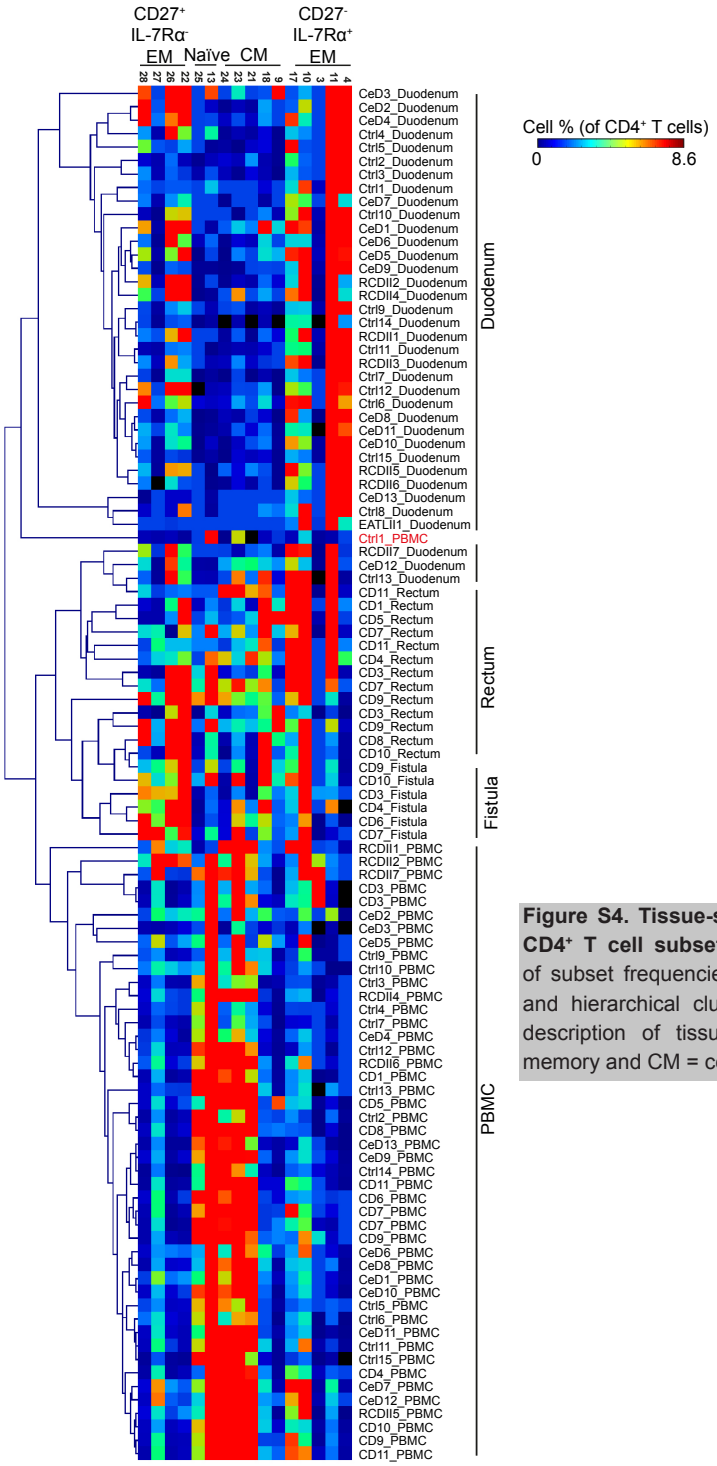


Figure S4. Tissue-specific composition of CD4⁺ T cell subsets. A heatmap summary of subset frequencies showing 102 samples and hierarchical clustering of samples with description of tissue type. EM = effector memory and CM = central memory

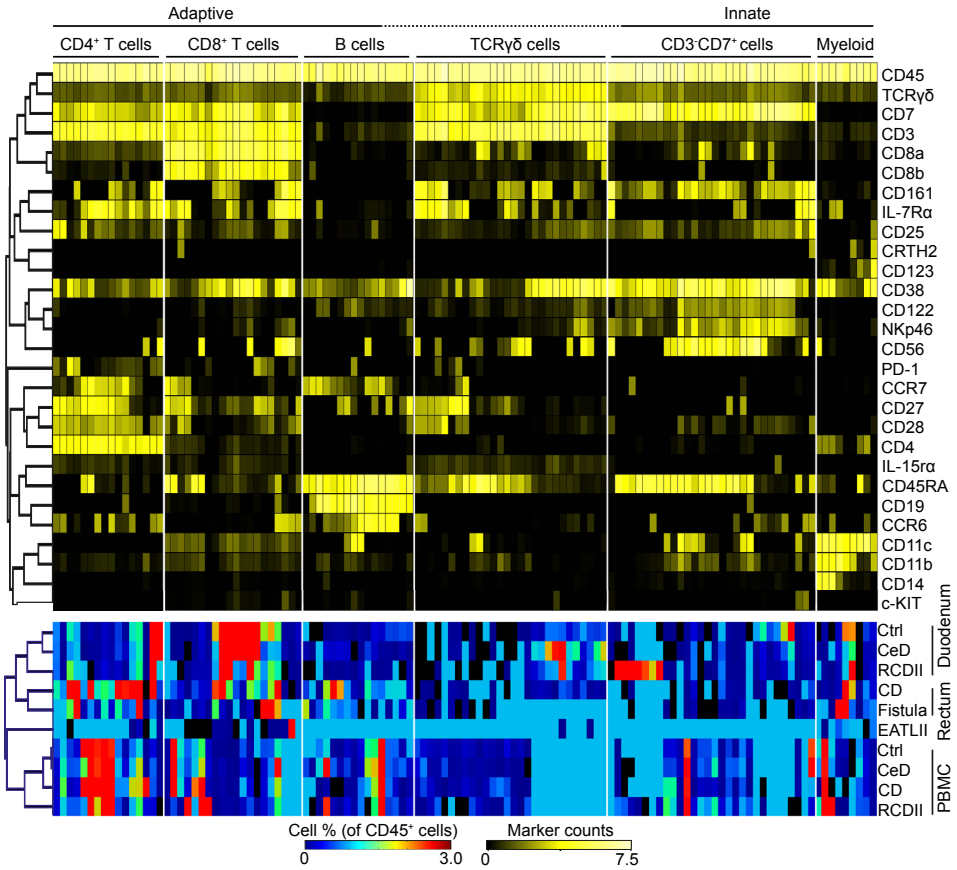


Figure S5. Marker expression profiles and composition of immune subsets in peripheral blood and intestine. Heatmaps showing characterization of cell populations (median ArcSinh5-transformed values of marker expression; black-to-yellow scale), composition (average cell percentages; rainbow scale) and hierarchical clustering of markers and samples.

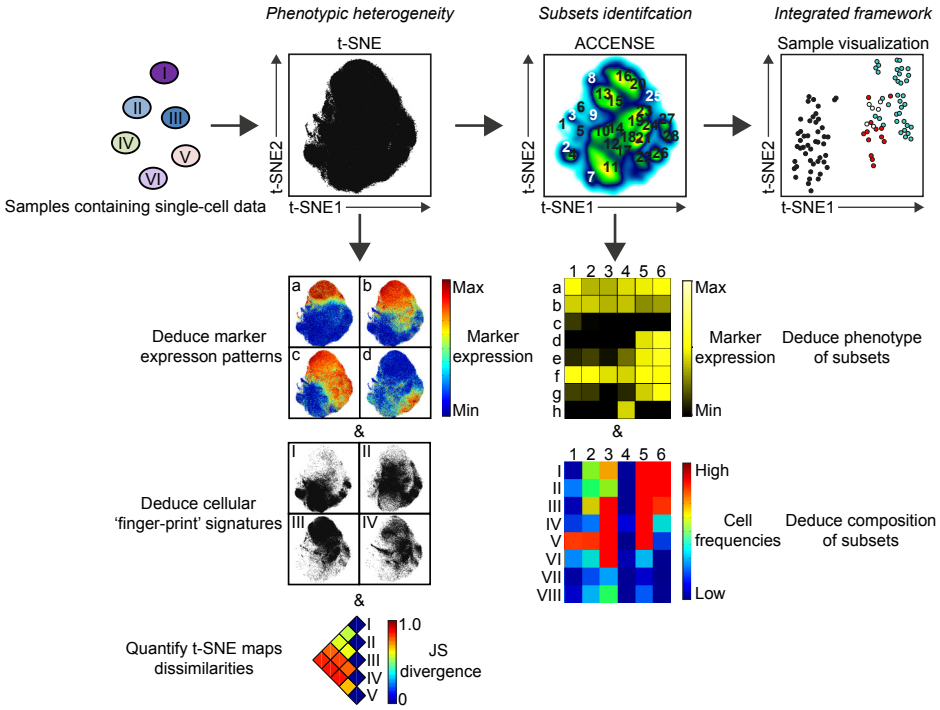


Figure S6. High-dimensional mass cytometry analysis pipeline. Samples containing single-cell data of cell populations of interest (e.g. based on SPADE analysis or gated subsets) are tagged with integer sample identification values, marker expression values are transformed using hyperbolic arcsin with a cofactor of 5, and cells events are downsampled to match sample cell number. Collective t-SNE dimensionality reduction was performed on the single-cell data derived from all the samples. A dot represents a single cell and color overlays illustrate marker expression values to deduce marker expression patterns. t-SNE maps are stratified for tissues and disease states to illustrate cellular 'finger-print' signatures. Pairwise Jensen-Shannon (JS) divergence analysis quantifies dissimilarity between a pair of t-SNE maps. On the basis of the map created by t-SNE, a machine-learning cell cluster detection approach (ACCENSE) delineates cell subsets. Heatmaps illustrate the phenotypic signature of the identified immune subsets and illustrate the composition of subsets within the samples. The cell frequency data of all the identified immune subsets are used to visualize samples in an integrated framework using the t-SNE algorithm, where a dot represents a single sample and the color overlay illustrates clinical information. Markers are indicated in letters, samples in roman letters and cell subsets in numbers.

Table S1. CyTOF antibody panel

Antigen	Tag	Clone	Com.	Cat#	Conc.
CD56	^{176}Yb	NCAM16.2	DVS	3176008B	1:100
CD4	^{145}Nd	RPA-T4	DVS	3145001B	1:100
CD8a	^{146}Nd	RPA-T8	DVS	3146001B	1:100
CD3	^{170}Er	UCHT1	DVS	3170001B	1:100
CD14	^{160}Gd	M5E2	DVS	3160001B	1:100
CD19	^{142}Nd	HIB19	DVS	3142001B	1:100
TCRgd	^{152}Sm	11F2	DVS	3152008B	1:100
CD45	^{154}Sm	HI30	DVS	3154001B	1:100
CD45RA	^{169}Tm	HI100	DVS	3169008B	1:100
CD27	^{167}Er	O323	DVS	3167002B	1:100
CD38	^{172}Yb	HIT2	DVS	3172007B	1:100
CD127	^{165}Ho	AO19D5	DVS	3165008B	1:100
CD11b	^{144}Nd	ICRF44	DVS	3144001B	1:100
CD7	^{147}Sm	CD7-6B7	DVS	3147006B	1:100
CD34	^{148}Nd	581	DVS	3148001B	1:100
C-Kit	^{143}Nd	104D2	DVS	3143001B	1:100
CD161	^{164}Dy	HP-3G10	DVS	3164009B	1:100
CD123	^{151}Eu	6H6	DVS	3151001B	1:100
CCR6	^{141}Pr	G034E3	DVS	3141003A	1:100
CD25	^{149}Sm	2A3	DVS	3149010B	1:100
CCR7	^{159}Tb	G043H7	DVS	3159003A	1:100
PD-1	^{175}Lu	EH 12.2H7	DVS	3175008B	1:100
CD11c	^{162}Dy	Bu15	DVS	3162005B	1:100
CD8b	^{166}Er	SIDI8BEE	eBio	14-5273	1:50
CRTH2	^{156}Gd	BM16	BioL	350102	1:50
TCRab	^{150}Nd	1P26	BioL	306702	1:50
IL-21R	^{153}Eu	2G1-K12	BioL	347802	1:50
IL-15Ra	^{168}Er	eBioJM7A4	eBio	14-7159-82	1:50
CD103	^{139}La	Ber-ACT8	BioL	350202	1:100
CD28	^{171}Yb	CD28.2	BioL	302902	1:100
NKp46	^{174}Yb	9E2	BioL	331902	1:100
CD122	^{158}Gd	TU27	BioL	339002	1:200

DVS Sciences (DVS), eBioscience (eBio), and Biolegend (BioL).

The antibody panel was designed to obtain a global overview of the heterogeneity of the immune system. For this purpose, the panel included markers that distinguish the major immune cell lineages, i.e. $CD4^+$ T cells, $CD8^+$ T cells, TCRγδ cells, B cells, innate lymphocytes ($CD3^-CD7^+$) and myeloid cells. In addition, markers were included that allow the identification of phenotypically distinct subsets within those immune cell lineages, like naïve and memory cells, resting and activated cells. In addition, markers were included that provide information on additional differentiation stages of cells, homing properties of cells, and their potential responsiveness to humoral factors like cytokines and chemokines. Our choice in markers was biased towards phenotyping CD7-expressing lymphoid cells (T cells and $CD3^-CD7^+$ cells) rather than B cells and myeloid cells.

Table S2. Characteristics of control and patients with inflammatory intestinal diseases

Characteristics	Ctrl (N=15)	CeD (N=13)	RCDII[a] (N=5)	EATLII (N=1)	CD[b] (N=10)
AgeBiop, (mean, ±SEM)	33.1±5.6	34.5±6.5	74.8±1.6	66	37.3±2.9
Gender, N (%)					
Male	5 (33.3)	6 (46.2)	5 (100.0)	0 (0.0)	3 (30.0)
Female	10 (66.6)	7 (53.8)	0 (0.0)	1(100.0)	7 (70.0)
GFD, N (%)					NA
No	13 (86.7)	5 (38.5)	0 (0.0)	0 (0.0)	-
Yes	2 (13.3)	8 (61.5)	5 (100.0)	1 (100.0)	-
Type of Biopsy					
Duodenum	15	13	7	1	-
Rectum	-	-	-	-	13
Fistula	-	-	-	-	6
PBMC[d]	14	13	6	-	14
Inflamed gut biopsy[c], N (%)					
No	14 (93.3)	6 (46.2)	2 (28.6)	0 (0.0)	12 (66.7)
Yes	1 (6.7)	7 (53.8)	5 (71.4)	1 (100.0)	6 (33.3)
Marsh score biopsy N (%)				NA	NA
M0	14 (93.3)	6 (46.2)	2 (28.6)	-	-
M1	0 (0.0)	1 (7.7)	0 (0.0)	-	-
M2	0 (0.0)	0 (0.0)	0 (0.0)	-	-
M3a	1 (6.7)	1 (7.7)	2 (28.6)	-	-
M3b	0 (0.0)	4 (30.8)	1 (14.3)	-	-
M3c	0 (0.0)	1 (7.7)	2 (28.6)	-	-

[a]Two RCDII patients were biopsied twice at different time points.
[b]Four CD patients were rectally biopsied twice at different time points.
[c]All 13 rectum biopsies of CD patients were non-inflamed, whereas all 6 fistula samples were inflamed.
[d]PBMC of CD patients were analyzed cryopreserved, whereas all other PBMC samples were analyzed fresh.
Celiac disease (CeD), refractory celiac disease type II (RCDII), enteropathy-associated T cell lymphoma type II (EATLII), Crohn's Disease (CD), age at biopsy time point (AgeBiop), glutenfree diet (GFD), and peripheral blood mononuclear cells (PBMC)

Chapter 6

Summarising discussion

It has long been thought that the human fetal immune system develops in a sterile womb without exposure to foreign antigens. Therefore, the developing immune system is considered to be immature. However, due to the challenge to obtain fetal tissues, much of our understanding of the developing fetal immune system is based on studies performed with cord blood collected from preterm or term babies. Moreover, in recent years, several lines of evidence have challenged the dogmas about both the sterile environment in utero and the immature status of the developing fetal immune system. With the introduction of advanced single cell analysis methods, novel opportunities have become available to characterize the heterogeneity and composition of the immune system with unprecedented resolution. In this thesis, I have taken advantage of this opportunity to analyze the human fetal innate and adaptive immune compartments in the spleen, liver and intestine during the second trimester. The results demonstrate an early-life immune compartmentalization across the tissues before birth. By applying advanced computational and visualization tools I was able to identify putative differentiation trajectories in the developing intestine, results that were corroborated by functional analysis in vitro.

Diversity in the innate immune compartment and identification of a novel ILC subset

To explore the heterogeneity and development of the immune system in the human fetal intestine, we used mass cytometry with a 35-antibody panel that was designed to capture heterogeneity in both the innate and adaptive compartments. In **chapter 2** we focused on the innate lymphoid cell (ILC) compartment. Here, 34 phenotypically distinct innate lymphoid cell clusters were distinguished, including previously identified NK and $CD127^+$ ILC subsets as well as several previously unrecognized clusters, providing evidence for extensive heterogeneity in the innate compartment. In particular, we identified a $Lin^-CD7^+CD127^-CD45RO^+CD56^+$ subset that by unbiased hierarchical clustering was positioned in between the known NK cell and $CD127^+$ ILC subsets, suggestive of potential developmental relationships with both the NK cells and $CD127^+$ ILCs.

The differentiation of ILCs has been well studied in mice whereas there are fewer studies on human ILC development[1, 2]. Also, a recent report demonstrated that human ILC development was distinct from the established murine model[3], which highlights the need to conduct research directly on human samples. The differentiation of cells or transition from one subset to another usually exhibits stepwise changes in the transcriptional program and protein expression profiles. t-SNE is exceptionally well suited to integrate such gradual changes for many

markers simultaneously and in Cytosplore we were able to visualize the t-SNE computation over time. Therefore, we made use of this approach to probe potential cell differentiation trajectories in the fetal intestine (**Chapter 2 and 3**). This provides compelling evidence that in all 7 intestinal samples the int-ILCs clustered in between and connected to both the ILC2s, ILC3s and NK cells. By designing a minimal flow cytometry panel based on the mass cytometry data we were able to identify and isolate the int-ILC by conventional flow cytometry. Furthermore, in functional assays we validated that int-ILCs can generate both NK cells and ILC3s in vitro (**Chapter 2**). Consistent with our findings, Chen et al. identified an intermediate ILC subset in human tonsil that could give rise to NK cells and ILC3s[3]. Similar to our $CD117^{+}CD94^{-}$ int-ILCs, this subset lacked the expression of CD34, CD127 and CD94, but expressed CD56 and CD117. Also, Chen et al. showed that the $CD56^{-}$ counterpart of this intermediate ILC could give rise to ILC2s. As ILC2s were prominent in the fetal intestine (about 5% of total ILCs) and were found to be connected with int-ILCs, we investigated whether the int-ILCs could differentiate into ILC2s as well. As we observed a sub-population of $CD8a^{-}$ int-ILCs expressing high level of GATA3, a masker transcriptional factor for ILC2 function and differentiation, we flow purified these $CD8a^{-}$ int-ILCs and stimulated them with ILC2 stimuli (IL-7, SCF, IL-2, IL-4, IL-25 and IL-33) in an OP9-DL1 co-culture system. However, under these conditions the cell surface phenotype of the int-ILCs remained unchanged. Further research is thus needed to clarify the putative relationship between int-ILC and ILC2.

ILC plasticity has been extensively addressed in both mouse and human, where certain ILC subsets will convert into other subsets after the appropriate cytokine stimuli[4, 5, 6]. As shown in **chapter 2**, all ILC subsets except ILC1s were identified in the human fetal intestine. Moreover, we observed a clear trajectory between ILC2s and ILC3s, suggesting that ILC2s may convert into ILC3s in utero as there is a virtual absence of ILC2s in the intestine after birth[7]. Furthermore, we noticed that the $CD8a^{-}$ int-ILCs-derived ILC3s and NK cells could partly revert their phenotype upon exposure to appropriate cytokine stimuli, suggesting that the int-ILCs may represent an intermediate subset between two plastic lineages. Importantly, the majority of the ILC subsets defined in the intestine including int-ILCs were not found in the spleen and liver as shown in **chapter 4**, indicative of specific functions in the mucosa. The factors in the tissue microenvironment that promote the differentiation of int-ILC into NK cells and ILC3s need to be further investigated in the future.

To extend the understanding of ILCs, the composition of ILCs in the fetal spleen and liver were determined in **chapter 4** and additional heterogeneity was

observed in several NK clusters compared with the fetal intestine, likely reflecting distinct functions in each tissue microenvironment. Furthermore, several new clusters were identified with this unbiased data-driven analysis approach, such as the CD8a^{+} ILC3s. This raises the issue of the biological significance of these and other clusters identified by high-dimensional mass cytometry. Therefore, it is important to validate these clusters and distinguish them functionally from previously identified subsets in future studies. However, our systemic and detailed studies of ILCs in fetal tissues have already advanced the understanding of ILC biology in human.

Memory formation in the human fetal intestine

One of the main features of adaptive immunity is the formation of immunological memory, where T cells play a critical role. Naive T cells recirculate through lymph nodes where they encounter peptide epitopes presented by major histocompatibility complex (MHC) class molecules on professional APCs such as DCs. After recognition of a given peptide-MHC complex, naive T cells proliferate and differentiate into effector cells, which migrate to the infection site and clear the pathogen. Thereafter the majority of the effector cells die, however, a small proportion survive as memory cells, which mediate the anamnestic immune responses upon encounter with the same pathogen. In the current thesis, we dissected the CD4^{+} T cell compartment in the human fetal intestine using an array of advanced single cell technologies, which revealed the presence of three major populations: T_N, T_M and T_{regs} (**Chapter 3**). Furthermore, with an array of computational tools including diffusion map, principal component analysis (PCA), Vortex and developing t-SNE, a linear differentiation trajectory was revealed. In this trajectory T_N cells clustered next to CD161$^{-/low}$ T_{CM} and CD161$^{-/low}$ T_{EM} cells, and the latter connected to CD161^{+} T_{EM} cells in both single cell mass cytometry and single cell RNA-seq datasets, consistent with CD4^{+} T cell memory formation in the fetal intestine. Moreover, pseudotime analysis revealed an up-regulation of TCR signaling associated transcripts at the end of T_N-T_M trajectory, which further underpinned the formation of memory T cells in the fetal intestine. These results are in line with a previous study that revealed a similar linear differentiation model of memory formation in circulating CD4^{+} T cells based on genome-wide profiles of DNA methylation, histone modifications and DNA accessibility[8]. A linear differentiation model was also observed in the human CD8^{+} T cell compartment[9, 10].

CD45RO is considered a marker for memory T cells, however, the majority of the CD4CD8 double-positive and 80% of the single-positive thymocytes are CD45RO^{+}

as well. Consequently, the occurrence of the isoform switching from CD45RO to CD45RA is considered one of the final steps in the maturation of T cells in the thymus[11]. This brings the possibility that the occurrence of CD45RO cells in the human fetal intestine might be due to thymic emigrants that have not switched from CD45RO to CD45RA. However, our mass cytometry analysis demonstrated that the T_M cells in the fetal intestine expressed CD161 and CD127, while these markers were not expressed by single-positive thymocytes in the fetus (Data not shown). Together with the abundance of CD45RA$^+$ T_N cells in cord blood and the observation that the fetal intestinal T_M cells readily produced cytokines, it is highly unlikely that the fetal intestinal T_M cells are derived from CD45RO thymocytes. However, the possibility that (a proportion) of the intestinal memory CD4$^+$ T cells originate from the mother still needs to be investigated through HLA typing or genotyping.

In the absence of foreign antigen exposure, the generation of memory-like T cells has been reported in mice. These antigen-inexperienced memory-like T cells harbored two main subsets: "innate memory" T cells and "virtual memory" T cells. The development of the former subset depends on IL-4 and thymic PLZF$^+$ cells, whereas the latter need IL-15 and CD8a$^+$ DCs[12]. However, these antigen-inexperienced memory-like T cells were found predominantly in the CD8$^+$ T cell compartment[12] and only few studies reported the formation of such T cells in the CD4$^+$ T compartment[13, 14]. While in man CD45RO$^+$ memory-like T cells have been identified in the fetal spleen[15, 16], these cells were able to proliferate upon stimulation with IL-2 but not after CD2 or CD3 cross-linking, indicative of an anergic state[16]. Therefore, these cells are distinct from the CD45RO$^+$ memory-like cells that we identified in the human fetal intestine (**Chapter 3 and 4**), as these readily produced IL-2, IFN-γ, IL-4, granzyme B and large amounts of TNF after stimulation.

Migratory DCs were found to be virtually absent before week 16 of gestation in the human fetus[17]. Consistent with this previous report[17], our single-cell RNA-seq analysis of fetal intestinal cells identified a cluster of APCs, which displayed high expression levels of genes encoding HLA-DR, CD74 (HLA-class II invariant chain), CD80, CD86 and chemokine receptor CCR7. As CCR7 mediates DCs migration to lymph nodes in adults[18], this would be compatible with priming of naive T cells in the mesenteric lymph nodes followed by homing of the resulting memory T cells to the lamina propria after week 16 of gestation. Also, we observed clonal expansion of intestinal T_M cells. Together, our data in **chapter 3** provide strong evidence for the formation of the bona fide memory CD4$^+$ T cells in the developing human fetal intestine in utero.

Nature of the antigens

The formation of memory T cells in the human fetal intestine (**Chapter 3**) suggests that this occurs due to exposure to (foreign) antigens. Strikingly, a marked clonal overlap in CDR3 amino acid repertoires was observed between memory $CD4^+$ T cells isolated from different fetal intestinal samples, suggesting exposure to similar antigen(s).

It is well established that maternal cells can cross the placental barrier and reside in the human fetal tissues from the second trimester onward during pregnancy, a phenomenon defined as "microchimerism" that can persist throughout adulthood[19]. In addition, it has been reported that non-inherited maternal antigens (NIMAs) promote the differentiation of human fetal $CD4^+$ T cells[19], presumably due to the direct contact between the fetal immune system and the semi-allogeneic maternal cells. Thus, some of the memory $CD4^+$ T cell responses in the fetus may result from this. However, as the memory formation was most pronounced in the fetal intestine and not in the fetal spleen and liver, it is highly unlikely that this is solely due to exposure to NIMAs.

It is now well established that the microbiota plays an important role in the development and shaping of the function of the immune system. However, in fetal life the infant has been thought to be protected from exposure to foreign antigens[20], a dogma that has been challenged by the detection of microbiota in the placenta[21], amniotic fluid[22] and meconium[23]. Although it is highly unlikely that there will be an abundance of viable microbes that pass the placental barrier, it is conceivable that microbial-derived antigens derived from the maternal intestinal microbiota will reach the amniotic fluid and as such reach the lumen of the fetal intestine where they could prime $CD4^+$ T cell responses. In addition, it has been reported that infection with cytomegalovirus is one of the most common causes of congenital infection and that cytomegalovirus-specific T cell responses can be detected in early life[24, 25], which might also lead to the priming of T cell responses in fetal life. In addition, allergen-specific T cells have been detected in cord blood[26, 27] and antigen-specific fetal T cells have also been detected in children whose mothers were infected with human immunodeficiency virus (HIV)[28], hepatitis C virus (HCV)[29] and malaria[30]. However, in our current studies, we used materials from fetuses obtained from mothers with a healthy pregnancy making it unlikely that the formation of memory T cells is driven by HIV, HCV or parasites. Rather we speculate that the memory $CD4^+$ T cells in the fetal intestine are specific for microbial antigens derived from the mother. Therefore, we are currently investigating if intestinal T cell lines can be generated from fetal samples that are

specific for such microbial antigens. For this purpose, HLA-typed fetal intestinal cells are cultured with lysates of relevant intestinal bacterial strains and propagated with cytokines and repeated antigen stimulation in the presence of HLA-matched antigen presenting cells. The specificity of such primary cultures and T cell clones derived thereof will subsequently be determined in cell proliferation and cytokine secretion assays. It is tempting to speculate that the composition of the microbiota of the mother may determine the imprinting of the mucosal immune system in the fetus already at this early age and that this may have consequences later in life.

Although the frequency of Tregs identified in our studies (**Chapter 3 and 4**) is lower than expected, we did observe clear clonal expansions in the CD45RO$^+$ T_{regs} compartment and a separate differentiation trajectory of the T_{regs} compared with T_M cells in the intestine (**Chapter 3**). Thus, the relationship between these cell subsets and the nature of the antigens that drive the expansion of the CD45RO$^+$ T_{regs} warrants further investigation.

In this respect it is interesting that Glanville et al. developed an elegant algorithm termed GLIPH (group of lymphocyte interactions by paratope hotspots), which can cluster TCRs that have a high probability of sharing specificity, taking both conserved motifs and global similarity of complementarity-determining region 3 (CDR3) sequences into account[31]. As fetal intestinal T cells will be exposed to a large repertoire of microbes upon birth, it will be of interest to apply this algorithm to our TCR-sequencing dataset as a means to identify commonalities in the TCR repertoire, which may point to the specific antigens that drive the formation of the T_M and CD45RO$^+$ T_{regs}.

Immune compartmentalization across and within tissues

It is well established that the anatomical tissue location of T cells is of great importance for their T cell function in both mice and man[32, 33]. Recently, Farber et al. demonstrated early-life compartmentalization of human T cells with abundant naive T cells in the pediatric blood and almost all tissues and abundant effector memory T cells in pediatric lung and small intestinal tissue only[34]. Here, we utilized high-dimensional mass cytometry to profile the immune landscape in fetal intestine, spleen and liver (**Chapter 4**) with a 35-antibody panel. In line with the previous study, fetal intestines harbored memory-like T cells, whereas most T cells in fetal spleens and livers displayed a naive-like phenotype, indicative of T cell compartmentalization as early as the second trimester. Moreover, mass cytometry analysis of fetal DCs derived from spleen, thymus, lung and intestine revealed substantial heterogeneity between these tissues, suggestive of different tissue

imprinting[17]. Very recently, Yudanin et al. have reported that the tissue location differentially impacted the composition of ILCs in human as well[35], supporting our finding of site-specific ILC signatures across fetal tissues. Our study in chapter 4 was one of the first to show the existence of the early-life compartmentalization of the immune cells across tissues within almost each immune lineage in utero, which suggests different immune responses in situ and further highlights the importance of tissue environment on the development of immune cells. Moving on from the fetus to the adult, our study in **chapter 5** also provides evidence for the regional specialization within the intestinal immune system in adults. Therefore, the integrated high-dimensional analysis provides a global and deeper understanding of the entire immune system and is in line with a local and tissue-specific immune signature.

Functionality of the clusters

The number of phenotypically distinct immune clusters has increased significantly due to the introduction of high-dimensional single-cell techniques such as mass cytometry and RNA-sequencing. This raises the question to what extent such clusters represent functionally distinct entities. In this thesis, clusters were distinguished by differential expression of at least one protein or gene. In **chapter 2 and 3**, 34 and 22 phenotypically distinct clusters were revealed in the intestinal ILC and $CD4^+$ T cell compartment based on the marker expression profiles, respectively. To allow further functional studies with the identified clusters, a minimal marker gating strategy was designed to identify and isolate clusters of interest by conventional flow cytometry. By applying this approach, we demonstrated that the int-ILCs can differentiate into NK cells and ILC3s in vitro and the differentiation trajectory and cytokine production profiles supported the observation of the generation of intestinal memory T cells in the fetus. Therefore, this proved to be an useful approach where mass cytometry is used as an exploratory tool leading to novel hypotheses that can be further tested in functional assays (**Figure 1**).

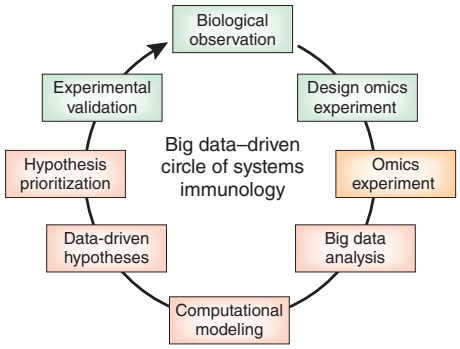


Figure 1 Big data–driven circle of systems immunology. Big dataset can be used to generate novel hypotheses in an unbiased data-driven fashion for further experimental validation using the classical approaches such as loss- and gain-of-function experiments, different murine models and other functional assays with flow cytometry. Adapted from Schultze (2015).

Interestingly, the int-ILCs were a main contributor to the separation of non-affected inflammatory bowel disease (IBD) (control biopsies) and affected IBD biopsies that contained NK cells and ILC3s (**Chapter 5**). Thus, the int-ILC may give rise to NK cells and ILC3s in the intestine in the inflammatory context. Compared with the phenotype of intestinal ILCs in adult, fetal intestinal ILCs mainly expressed CD45RO, a marker expressed by memory T cells. As it has been reported that $CD45RO^+$ ILC3s produced higher amounts of IL-17, IL-22, GM-CSF, TNF and IL-8 than the $CD45RA^+$ ILC3s[36], the function of these $CD45RO^+$ ILCs need to be investigated in the future. Consistent with previously finding[37], we did not find IL-17A-producing $CD4^+$ T cells in the fetal intestine, indicating that the generation of such cells takes place at a later time point and is associated with the post-natal exposure to the microbiota. Of note, a subgroup of the memory-like $CD4^+$ T cells expressed CCR6 and CD117. As the former marker is associated with Th17 cell differentiation, these cells may be the precursor for the Th17 cells. Alternatively, the expression of CD117 may indicate that these cell represent a reservoir of memory cells that can rapidly expand when stem cell factor is released. Thus, the function of these cells warrants future investigation. In addition, the intestinal $CD4^+$ T cells produced not only Th1- and Th2-type cell cytokines, but also IL-2, granzyme B and TNF. Recently, it has been shown that low numbers of TNF^+CD4^+ T_{EM} cells promoted intestinal epithelial development, whereas high numbers impaired the epithelial development[38], pointing to a non-immune function of these cells. However, these cells can also mediate an intestinal inflammatory response in preterm babies that may contribute to the necrotizing enterocolitis[38]. Thus, tight regulation of these memory $CD4^+$ T cells is crucial already at a very early age, which may be reflected by the prominent T_{reg} compartment in the developing fetal intestine. Together, these data illustrate that studies on the early-life immune system can potentially contribute to the understanding of derailed immune-associated diseases later in life. With the identification of the 177 and 142 immune clusters in different fetal tissues and in the intestinal biopsy and PBMCs through the unbiased data-driven approach, respectively (**Chapter 4 and 5**), there is a clear need to further determine the functional significance of these clusters in future studies.

Spatial distribution of the immune subsets in tissue microenvironment

Next to cell types, the cellular organization and cell-cell interactions play an important role in the maintenance of the appropriate tissue function. The tissue architecture and relationships between cells are lost when using mass and flow cytometry, however, imaging-mass cytometry breaks this limitation, as it can measure up to 40 markers in a single tissue section at 1 µm resolution, which thus

allows for an in-depth analysis of tissue structure[39]. We were one of the first to use imaging-mass cytometry in the Netherlands. Using this promising technique, the co-localization of the $CD4^+$ T cells with HLA-DR^+CD163^+ APCs and $CD127^+$ ILCs was revealed in the fetal intestine, which suggests functional interactions between these types of cells in the developing fetal immune system (**Chapter 3 and 4**). With the identification of a large number of phenotypically distinct cell clusters in both the innate and adaptive immune compartment within and across tissues, we were able to visualize the early-life immune compartmentalization in the fetal tissues in situ by using imaging-mass cytometry (**Chapter 4 and Figure 2**). Therefore, imaging-mass cytometry provides an unique opportunity to determine immune heterogeneity and cell-cell interactions in situ, providing another layer of understanding of the immune system. Furthermore, imaging-mass cytometry makes it possible to investigate issues that were raised in observations made in the current thesis. For example, we observed that int-ILCs can differentiate into NK cells and ILC3s during co-culture with stromal cells (OP9-DL1) in vitro (**Chapter 2**). Consequently, better understanding of the neighborhood of int-ILCs may help to put this observation in the tissue context. Also, we expect that the determination of spatial location of the distinct $CD4^+$ T_M cell clusters and the interaction of these memory T cells with stromal cells and other innate or adaptive immune cells in the human fetal intestine will shed light on the cellular interactions underlying the generation of various $CD4^+$ memory T cells populations in utero. In this respect, it will also be very interesting to study the development of fetal lymphoid structures like Peyer's patches. Finally, the intestine has been considered as "the second brain" in our body and it has been reported that ILCs are adjacent to mucosal neurons[40, 41], so application of the imaging-mass cytometry on the human fetal intestine can provide more insights into the development of the mucosal neuronal network and its interactions with the developing immune system.

Imaging-mass cytometry data can be visualized by an overlay of individual makers as in traditional immune fluorescence staining. However, the limited number of markers that can be visualized simultaneously is not compatible with the complex multiplexed imaging-mass cytometry data. Up to recently, only few tools have been developed to analyze imaging-mass cytometry data, such as CellProfiles[42] and histoCAT[43], which mainly work at the cellular level. Therefore, the analysis of imaging-mass cytometry data is still a big challenge in this field. Besides the development of the Imacyte toolkit (accepted), which facilitates the analysis of imaging-mass cytometry data at the cellular level, in collaboration with the bioinformaticians in the LUMC and TU Delft we have extended this analysis to the pixel level (**Figure 2**) using another in-house developed software tool (Cytosplore Imaging, unpublished), which allows an in-depth exploration of the data at the

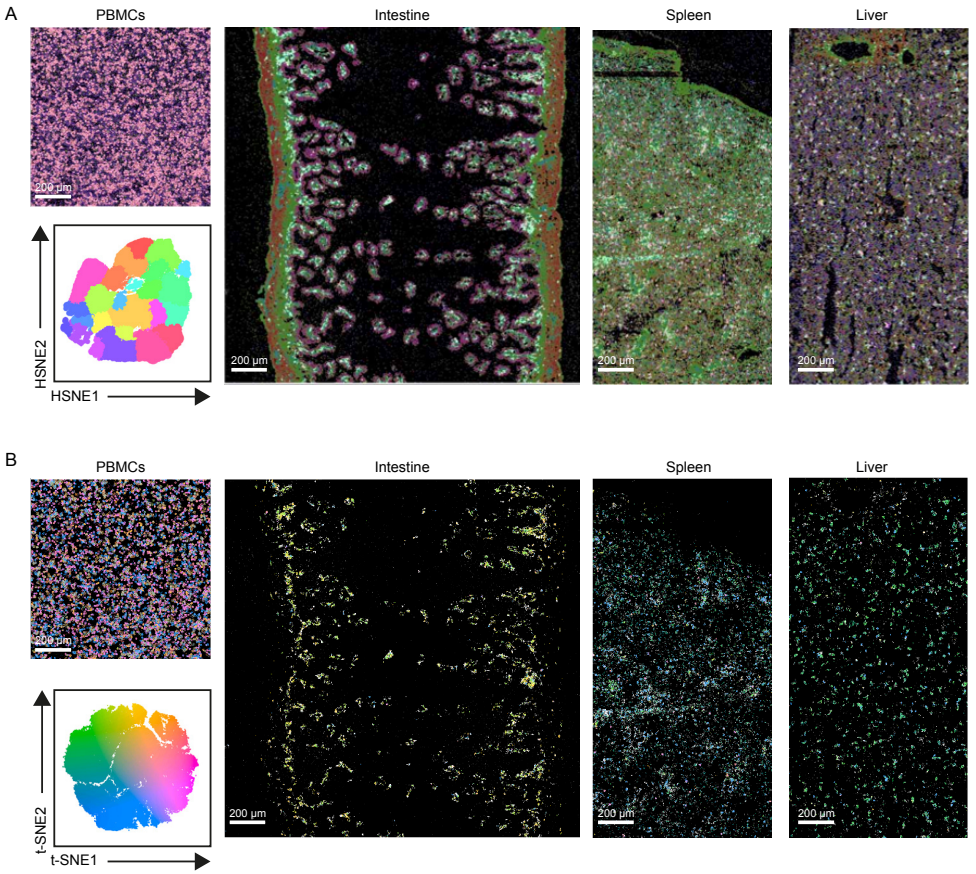


Figure 2 Spatial distribution of the entire immune cells reveals the early immune compartmentalization across tissues *in situ*. (A) Pixels derived from PBMCs (1ROI), fetal intestine (4ROIs), fetal spleen (2ROIs) and fetal livers (2ROIs) based on the expression of E-cadherin, Vimentin, SMA, D2-40, collagen I and CD45 Cytosplore Imaging software (unpublished). This yielded distinct color-coded clusters of pixels that were projected back onto the original images to obtain a global overview of the tissue architecture, reflecting the overall morphological differences between the tissues. Representative mass cytometry images of the PBMCs control, a fetal intestine, spleen and liver, showing the spatial distribution of the immune and stromal clusters. HSNE embedding of all the pixels derived from the PBMCs control, a fetal intestine, spleen and liver. Colors represent the different cell clusters. Scale bar: 200 µm. (B) The similar analysis were performed on the $CD45^+$ pixels based on the 26 immune makers. Representative mass cytometry images of the PBMCs control, a fetal intestine, spleen and liver, showing the spatial distribution of only the immune cell pixels. t-SNE embedding of the $CD45^+$ pixels derived from the PBMCs control, a fetal intestine, spleen and liver. Colors indicate the XY-coordinates of the t-SNE plot. Scale bar: 200 µm.

subcellular level. Currently, we are integrating the analysis at the pixel level with that at the cellular level in the tissue context. In the future, we would like to integrate the imaging-mass cytometry data with mass cytometry and RNA-seq

data. Eventually, a 3D image of the fetal tissues could also be reconstructed by using a series of tissue sections, in which the entire network between immune cells and surrounding stromal cells could be visualized in the tissue microenvironment.

Although there are many advantages to imaging-mass cytometry, a fundamental limitation is the speed of acquisition: 1 mm^2 in 2 hours. This hampers the identification of rare cell populations such as the int-ILCs and $CD117^+$ T cells in the tissue samples. To overcome this limitation, we are currently developing 7-color Vectra antibody panels to detect int-ILCs in sections to allow preselection of regions of interest (ROIs) in a serial section for analysis with imaging-mass cytometry. Such an approach can also be applied to tissue samples from patients where there is substantial heterogeneity in the cell distribution patterns, such as in tumors or inflamed intestine. Finally, imaging-mass cytometry can also be used as an explorative tool to identify disease-specific clusters in the tissue context. This can form the basis for the development of a targeted Vectra or immunochemistry antibody panel to detect such disease-specific features for diagnostic and/or prognostic purposes.

System immunology

System immunology is a new approach, where the immune system is investigated by integrating the results of a variety of approaches including single-cell mass cytometry, (single-cell) transcriptomics, TCR and BCR repertoire sequencing, metabolomics, proteomics, multiplexed imaging and epidemiology[44]. In **chapter 3**, we used this approach by combining results obtained from single-cell mass cytometry, with single-cell RNA sequencing, TCR sequencing, flow cytometry and imaging-mass cytometry, to investigate the $CD4^+$ T cell compartment in the human fetal intestine. Here, all techniques used reinforced the conclusion that memory-like $CD4^+$ T cells are generated in the human fetal intestines in utero. Such a multidisciplinary systems immunology approach is particularly useful for studies on the human immune system as there are very limited possibilities for experiments to determine causal correlations in vivo. Importantly, for such a cooperative approach, a multi-disciplinary team of clinicians, technicians, bioinformaticians and immunologists is crucial to the success. Continuous integration of novel technologies will further enhance the power of this approach. For example, at the time we were conducting our studies, Cite-seq to quantify the expression of proteins and transcripts simultaneously at the single-cell level was not yet available so we had to rely on indirect approaches to distinguish between $CD45RA^+$ and $CD45RO^+$ T cells. Thus, with the development of more advanced techniques such as the third generation RNA sequencing, imaging-mass cytometry

and the scalable, interactive and user-friendly computational tools, the system-immunology approach will provide us an unique opportunity to determine how the complex immune system operates in the tissue niche in health and disease.

A prenatal window of opportunity?

The early postnatal period is considered a period in which environmental factors strongly affect the development of the immune system, which is likely linked to protection or susceptibility to a variety of immune-mediated diseases later in life[45]. In several studies, the impact of environmental factors on the fetal immune system have been investigated. Van de Pavert et al. reported that there was a significantly reduction of the lymph size and a decrease of the immune response efficiency in offspring that was fed a vitamin A-deficient diet[46]. Furthermore, an elegant study has shown that the maternal microbiome reprogramed the intestinal transcriptional expression profiles and shaped the immune composition of the offspring[47]. Also, allergen-specific T cell reactivity has been found in cord blood as early as week 23 of gestation[26, 27] and, as mentioned above, antigen-specific T cells have been identified in fetuses whose mothers were suffering from human cytomegalovirus (HCMV)[25, 48, 49], HIV[28], hepatitis C virus[29], malaria[30], Cruzi[50] and toxoplasmosis[51] infection during pregnancy. Consistent with these findings, in the present study, tissue-resident memory-like $CD4^+$ T cells have been identified in the fetal intestine whose mothers were with the normal pregnancy as early as the second trimester (**Chapter 3**). Moreover, DCs were identified in the human fetal intestine (**Chapter 3 and 4**), which may mediate T cell priming[17]. Also, intestinal NK cells readily produce granzyme B and perforin, suggesting functional maturation of NK cells in the developing fetus as well (**Chapter 2**). The formation of memory $CD4^+$ T cell responses is accompanied by the appearance of $Foxp3^+$ Tregs in several fetal tissues, the fetal intestine in particular (**Chapter 3 and 4**). In addition, it has been shown that fetal DCs strongly induced the production of Tregs[17]. Finally, several B cell clusters were $CD20^+CD27^{int}CD38^+$, indicative of a regulatory B cell phenotype, in the fetal tissues. As such it is likely that the fetal immune system can be "trained" by the maternal-derived environmental factors before birth, which may provide an opportunity for the development of preventive strategies for the development of immune-mediated diseases such as celiac disease, inflammatory bowel diseases and allergy by influencing the balance between the development of memory $CD4^+$ T cells and $Foxp3^+$ Tregs in utero.

Conclusions and future perspectives

Our studies provide a global, comprehensive and detailed description of the fetal

immune system during healthy pregnancy by integrating an array of advanced high-parameter single-cell techniques. We further determined the function of several identified cell clusters and identified a novel intestinal ILC subset that can give rise to NK cells and ILC3s in vitro (**Figure 3**), which adds another layer of understanding of ILC differentiation and plasticity. The full differentiation potential of this new subset needs to be explored in the future. What's more, we revealed the generation of memory-like $CD4^+$ T cells in the developing human fetal intestine (**Figure 4**), indicating the exposure to antigens in utero. Determination of the nature of the antigens, as well as the function of these memory-like T cells would be crucial for understanding the transition from the relatively "clean" womb to the relatively "dirty" external world. Additionally, the observation of site-specific immunity highlighted the importance to investigate the immune system in the tissue niche. The crosstalk between immune cells and surrounding stromal cells can now be determined using imaging-mass cytometry, which can be integrated with other omics data as well. Next to the immune landscape of the fetal spleen, liver and intestine that we have described, it will be relevant to investigate the composition of the developing immune system in other fetal lymphoid and non-lymphoid tissues. In particular, comparisons of the fetal immune system at several time points in gestation and the neonatal period would be worthwhile as this will give a global picture of the development of human immune system and fill the gaps between prenatal and neonatal immunity. Overall, the results presented in the current thesis deepens our understanding of prenatal immunity and may ultimately be useful for the development of "early" intervention strategies to prevent the development of immune mediated diseases later in life.

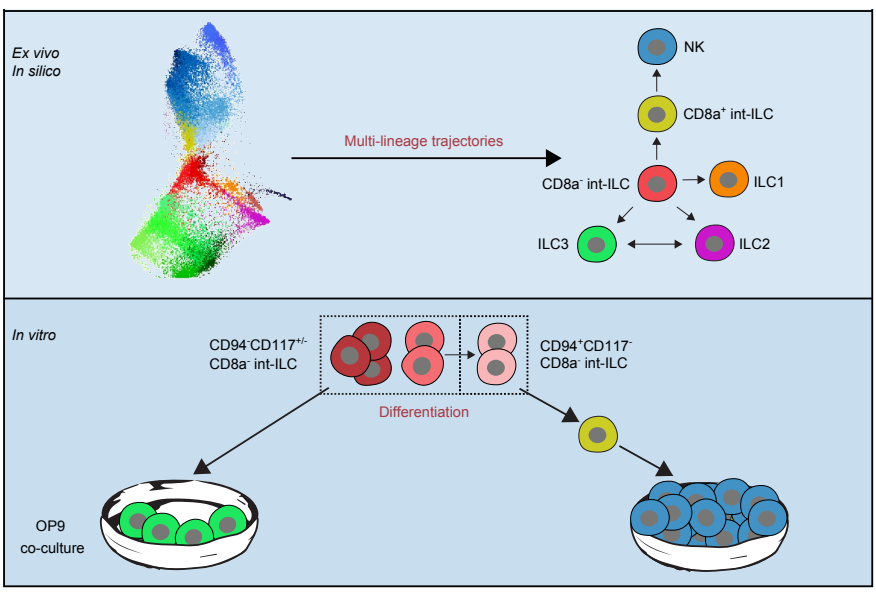


Figure 3. Differentiation of int-ILCs. Monitoring t-SNE computation dynamics predicts the potential differential trajectories *in silico* (upper panel). Upon co-culturing with OP-DL1, the int-ILCs can give rise to NK cells (blue) and ILC3s (green) (bottom panel).

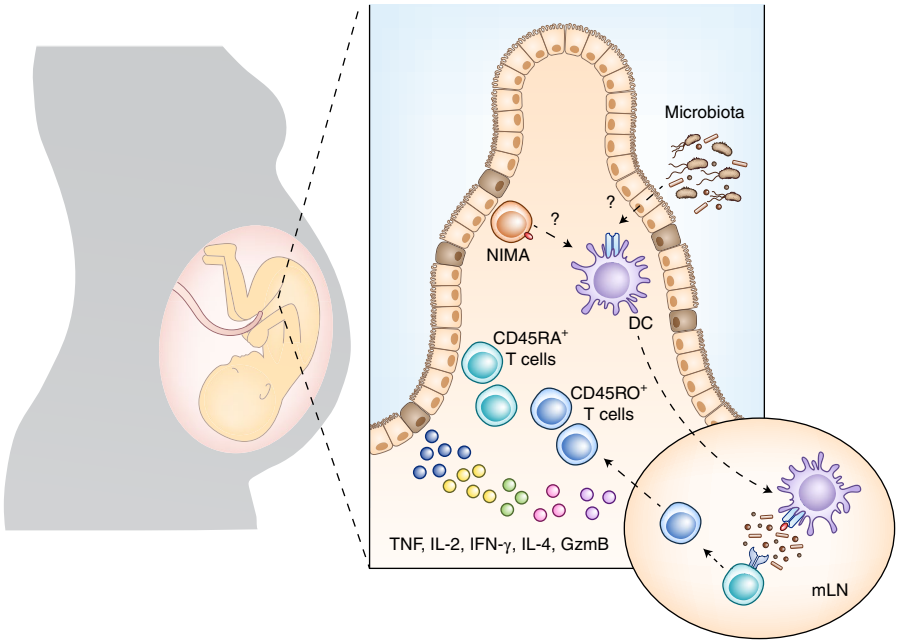


Figure 4. Memory generation of CD4⁺ T cells in the human fetal intestine. Phenotypically distinct naive (CD45RA⁺) and memory-like (CD45RO⁺) CD4⁺ T cells reside in the human fetal intestine at the second trimester during healthy pregnancy. The naive CD4⁺ T cells can produce TNF and IL-2, whereas memory-like T cells can produce not only TNF and IL-2, but also IFN-γ, IL-4 and granzyme B (GzmB), indicating functionality of the CD4⁺ T cell compartment. The identification of resident DCs and migratory DCs in the intestinal lamina propria and mesenteric lymph nodes (mLNs) by the second trimester prompted us to speculate their role in the generation of functional fetal memory-like CD4⁺ T cells. Although commensal microbe-derived antigens or NIMAs have been thought to be important in the memory generation of fetal intestinal T cells, the antigen specificity of these fetal memory-like CD4⁺ T cells remains unidentified. Adapted from Lim et al. (2019).

References

1. Vivier, E. et al. Innate Lymphoid Cells: 10 Years On. Cell 174, 1054-1066 (2018).
2. Juelke, K. & Romagnani, C. Differentiation of human innate lymphoid cells (ILCs). Curr Opin Immunol 38, 75-85 (2016).
3. Chen, L. et al. CD56 Expression Marks Human Group 2 Innate Lymphoid Cell Divergence from a Shared NK Cell and Group 3 Innate Lymphoid Cell Developmental Pathway. Immunity 49, 464-476 e464 (2018).
4. Bernink, J.H. et al. Interleukin-12 and -23 Control Plasticity of CD127(+) Group 1 and Group 3 Innate Lymphoid Cells in the Intestinal Lamina Propria. Immunity 43, 146-160 (2015).
5. Bal, S.M. et al. IL-1 beta, IL-4 and IL-12 control the fate of group 2 innate lymphoid cells in human airway inflammation in the lungs. Nat Immunol 17, 636-+ (2016).
6. Cella, M., Otero, K. & Colonna, M. Expansion of human NK-22 cells with IL-7, IL-2, and IL-1beta reveals intrinsic functional plasticity. Proc Natl Acad Sci U S A 107, 10961-10966 (2010).
7. Mjosberg, J.M. et al. Human IL-25- and IL-33-responsive type 2 innate lymphoid cells are defined by expression of CRTH2 and CD161. Nat Immunol 12, 1055-1062 (2011).
8. Durek, P. et al. Epigenomic Profiling of Human CD4(+) T Cells Supports a Linear Differentiation Model and Highlights Molecular Regulators of Memory Development. Immunity 45, 1148-1161 (2016).
9. Abdelsamed, H.A. et al. Human memory CD8 T cell effector potential is epigenetically preserved during in vivo homeostasis. J Exp Med 214, 1593-1606 (2017).
10. Moskowitz, D.M. et al. Epigenomics of human CD8 T cell differentiation and aging. Sci Immunol 2 (2017).
11. Fujii, Y., Okumura, M., Inada, K., Nakahara, K. & Matsuda, H. CD45 isoform expression during T cell development in the thymus. Eur J Immunol 22, 1843-1850 (1992).
12. White, J.T., Cross, E.W. & Kedl, R.M. Antigen-inexperienced memory CD8(+) T cells: where they come from and why we need them. Nat Rev Immunol 17, 391-400 (2017).
13. Marusina, A.I. et al. CD4(+) virtual memory: Antigen-inexperienced T cells reside in the naive, regulatory, and memory T cell compartments at similar frequencies, implications for autoimmunity. J Autoimmun 77, 76-88 (2017).
14. Younes, S.A. et al. Memory Phenotype CD4 T Cells Undergoing Rapid, Nonburst-Like, Cytokine-Driven Proliferation Can Be Distinguished from Antigen-Experienced Memory Cells. Plos Biol 9 (2011).
15. 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).
16. Byrne, J.A., Stankovic, A.K. & Cooper, M.D. A novel subpopulation of primed T cells in the human fetus. J Immunol 152, 3098-3106 (1994).
17. McGovern, N. et al. Human fetal dendritic cells promote prenatal T-cell immune suppression through arginase-2. Nature 546, 662-+ (2017).
18. Ohl, L. et al. CCR7 governs skin dendritic cell migration under inflammatory and steady-state conditions. Immunity 21, 279-288 (2004).
19. Mold, J.E. et al. Maternal alloantigens promote the development of tolerogenic fetal regulatory T cells in utero. Science 322, 1562-1565 (2008).
20. Hornef, M. & Penders, J. Does a prenatal bacterial microbiota exist? Mucosal Immunology 10, 598-601 (2017).
21. Aagaard, K. et al. The placenta harbors a unique microbiome. Sci Transl Med 6, 237ra265 (2014).
22. DiGiulio, D.B. Diversity of microbes in amniotic fluid. Semin Fetal Neonatal Med 17, 2-11 (2012).
23. Ardissone, A.N. et al. Meconium microbiome analysis identifies bacteria correlated with premature birth. PLoS One 9, e90784 (2014).
24. Carlson, A., Norwitz, E.R. & Stiller, R.J. Cytomegalovirus infection in pregnancy: should all women be screened? Rev Obstet Gynecol 3, 172-179 (2010).
25. Miles, D.J. et al. CD4(+) T cell responses to cytomegalovirus in early life: a prospective birth cohort study. J Infect Dis 197, 658-662 (2008).

26. Jones, A.C. et al. Fetal peripheral blood mononuclear cell proliferative responses to mitogenic and allergenic stimuli during gestation. Pediatric Allergy and Immunology 7, 109-116 (1996).

27. Szepfalusi, Z. et al. Transplacental priming of the human immune system with environmental allergens can occur early in gestation. Journal of Allergy and Clinical Immunology 106, 530-536 (2000).

28. Rowland-Jones, S.L. et al. HIV-specific cytotoxic T-cell activity in an HIV-exposed but uninfected infant. Lancet 341, 860-861 (1993).

29. Babik, J.M., Cohan, D., Monto, A., Hartigan-O'Connor, D.J. & McCune, J.M. The human fetal immune response to hepatitis C virus exposure in utero. J Infect Dis 203, 196-206 (2011).

30. Odorizzi, P.M. et al. In utero priming of highly functional effector T cell responses to human malaria. Sci Transl Med 10 (2018).

31. Glanville, J. et al. Identifying specificity groups in the T cell receptor repertoire. Nature 547, 94-98 (2017).

32. Mueller, S.N., Gebhardt, T., Carbone, F.R. & Heath, W.R. Memory T cell subsets, migration patterns, and tissue residence. Annu Rev Immunol 31, 137-161 (2013).

33. Sathaliyawala, T. et al. Distribution and Compartmentalization of Human Circulating and Tissue-Resident Memory T Cell Subsets. Immunity 38, 187-197 (2013).

34. Thome, J.J. et al. Early-life compartmentalization of human T cell differentiation and regulatory function in mucosal and lymphoid tissues. Nat Med 22, 72-77 (2016).

35. Yudanin, N.A. et al. Spatial and Temporal Mapping of Human Innate Lymphoid Cells Reveals Elements of Tissue Specificity. Immunity 50, 505-519 e504 (2019).

36. Shikhagaie, M.M. et al. Neuropilin-1 Is Expressed on Lymphoid Tissue Residing LTi-like Group 3 Innate Lymphoid Cells and Associated with Ectopic Lymphoid Aggregates. Cell Reports 18, 1761-1773 (2017).

37. 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, 238ra272 (2014).

38. Schreurs, R. et al. Human Fetal TNF-alpha-Cytokine-Producing CD4(+) Effector Memory T Cells Promote Intestinal Development and Mediate Inflammation Early in Life. Immunity 50, 462-476 e468 (2019).

39. Giesen, C. et al. Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. Nature Methods 11, 417-+ (2014).

40. Cardoso, V. et al. Neuronal regulation of type 2 innate lymphoid cells via neuromedin U. Nature 549, 277-281 (2017).

41. Jowett, G.M. & Neves, J.F. Commentary: Neuronal regulation of type 2 innate lymphoid cells via neuromedin U. Front Pharmacol 9, 230 (2018).

42. Carpenter, A.E. et al. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biol 7, R100 (2006).

43. Schapiro, D. et al. histoCAT: analysis of cell phenotypes and interactions in multiplex image cytometry data. Nat Methods 14, 873-876 (2017).

44. Davis, M.M., Tato, C.M. & Furman, D. Systems immunology: just getting started. Nat Immunol 18, 725-732 (2017).

45. Renz, H. et al. The neonatal window of opportunity-early priming for life. J Allergy Clin Immunol 141, 1212-1214 (2018).

46. van de Pavert, S.A. et al. Maternal retinoids control type 3 innate lymphoid cells and set the offspring immunity. Nature 508, 123-127 (2014).

47. Gomez de Aguero, M. et al. The maternal microbiota drives early postnatal innate immune development. Science 351, 1296-1302 (2016).

48. Vermijlen, D. et al. Human cytomegalovirus elicits fetal gammadelta T cell responses in utero. J Exp Med 207, 807-821 (2010).

49. Marchant, A. et al. Mature CD8(+) T lymphocyte response to viral infection during fetal life. J Clin Invest 111, 1747-1755 (2003).

50. Hermann, E. et al. Human fetuses are able to mount an adultlike CD8 T-cell response. Blood 100, 2153-2158 (2002).

51. Hara, T. et al. Human V delta 2+ gamma delta T-cell tolerance to foreign antigens of Toxoplasma gondii. Proc Natl Acad Sci U S A 93, 5136-5140 (1996).

Chapter 7

Appendices

English summary

Fetal immune system

The immune system provides protection against pathogens such as bacteria, viruses, parasites and fungi. It comprises a cellular and humoral compartment. The former is classically divided into the innate and adaptive immune compartment. Innate immunity provides a crucial first line of defense mediated by a swift, general and non-specific response to the invader. The innate compartment consists of various types of phagocytic granulocytes, myeloid cells and innate lymphoid cells (ILCs). Simultaneously, the adaptive immune system will mount a more tailored and specific response through antigen-specific receptors-expressing B and T lymphocytes, the latter including helper $CD4^+$ T cells and cytotoxic $CD8^+$ T cells. Ultimately, the combined innate and adaptive will lead to suppression or elimination of the invading pathogen and the generation of immunological memory that provides long-lasting immunity to the pathogen.

The fetal immune system develops during pregnancy and it has to remain tolerant to the semi-allogeneic environment while being prepared for the immense exposure to foreign antigens directly upon birth. As the largest immune compartment in our body, the intestine is continuously exposed to antigens derived from both harmless dietary constituents and potential harmful (commensal) bacteria. For this, an elaborate intestinal immune system has evolved to distinguish the harmful from beneficial antigens. Although a few studies have reported the identification of several immune cell subsets in the fetal tissues such as memory T cells in spleen and intestine, due to the scarcity of the fetal tissues and technique limitations, most investigation of the fetal immune system are based on cord blood collected at birth. Therefore, a system-wide and detailed phenotypical characterization of the composition and development of the human fetal immune system, especially the intestine was lacking. In the current study, we have approached this issue by making use of recently introduced single cell technologies, including (imaging-) mass cytometry and single-cell RNA-sequencing.

(Imaging-) Mass cytometry

Mass cytometry is a new platform that couples flow cytometry with mass spectrometry. This technique uses rare earth metals as reporters instead of fluorophores, which removes the limitation on the number of antibodies that can simultaneously be used due to the spectral overlap of the fluorochromes, and eliminates the problem of autofluorescence. Through mass cytometry up

to 42 markers can now be used simultaneously at single-cell resolution with minimal overlap between channels. Therefore, mass cytometry now provides the opportunity to study the diversity and heterogeneity of the immune system with an unprecedent high resolution.

To gain spatial information *in situ*, imaging-mass cytometry has been developed recently as a new technology, which couples a laser ablation system with a mass cytometer. In this way, it enables the visualization of dozens of markers on the same tissue section simultaneously with a sub-cellular resolution of 1 μm. Therefore, imaging-mass cytometry further offers an opportunity to gain unprecedented insight into the organization of the immune system *in situ*.

Single-cell RNA-sequencing

As a complementary approach to mass cytometry, single-cell RNA sequencing dissects the gene expression profiles on tens of thousands of individual cells at the single-cell level. As such, it provides an opportunity to determine the cell transcriptomic heterogeneity and the discovery of rare cell populations. Moreover, it facilitates the identification of potential cellular differentiation trajectories and regulatory networks among cells at the gene level.

Heterogeneity and differentiation of innate lymphoid cells

In **Chapter 2** we explored the heterogeneity and development of the innate lymphoid cell (ILC) compartment in the fetal intestine using mass cytometry with a 35-antibody panel. Here, 34 phenotypically distinct innate lymphoid cell clusters were distinguished, including previously identified NK and $CD127^+$ ILC subsets as well as several previously unrecognized clusters, providing evidence for extensive heterogeneity in the innate compartment. Moreover, by visualizing the dynamics of the t-SNE computation, we identified smooth phenotypic transitions from cells within the $Lin^-CD7^+CD127^-CD45RO^+CD56^+$ (int-ILCs) cluster to both NK cells and $CD127^+$ ILCs, revealing potential differentiation trajectories. Finally, in functional assays, we validated that the int-ILCs can indeed give rise to both NK cells and ILC3s.

Memory formation in the human fetal intestine

In **Chapter 3** we combined mass cytometry with single-cell RNA-sequencing, TCR-sequencing, imaging-mass cytometry and flow cytometry to dissect the $CD4^+$ T cell compartment in the human fetal intestine. Here, mass cytometry identified 22 phenotypically distinct clusters, including naive-like, memory-like and regulatory-

like subpopulations, which were confirmed and further characterized at the transcriptional level. Moreover, we observed that the memory-like $CD4^+$ T cells readily produced the cytokines TNF, IFN-γ and IL-2. Furthermore, the integrated single-cell analysis revealed a robust memory differentiation trajectory. Finally, clonal expansion and co-localization of memory-like $CD4^+$ T cells with APCs further supported the concept of the generation of memory-like $CD4^+$ T cells in the human fetal intestine, suggesting exposure to foreign antigens *in utero*.

Site-specific immune signatures across and within tissues

To further extend the understanding of the fetal immune system, in **Chapter 4** we applied mass cytometry to profile the innate and adaptive immune compartment in the fetal intestine, spleen and liver. Here, we identified 177 phenotypically distinct clusters including both previously identified and novel cell clusters. PCA analysis indicated substantial differences between the composition of the immune system in the different tissues. Moreover, imaging-mass cytometry further underpinned the distinctness of the immune system in the tissues *in situ.* Thus, our results provide evidence for early-life compartmentalization of the immune system in fetal spleen, liver and intestine.

In **Chapter 5**, we applied the mass cytometry to analyze the composition of the mucosal immune system of patients with inflammatory intestinal diseases and controls. Here, we revealed previously unrecognized heterogeneity in the mucosal immune system and tissue- and disease-specific immune subsets.

Conclusion

Our studies provide a global, comprehensive and detailed description of the fetal immune system during healthy pregnancy by integrating an array of advanced high-parameter single-cell techniques. We further determined the function of several identified cell clusters and identified a novel intestinal ILC subset that can give rise to NK cells and ILC3s *in vitro*, which adds another layer of understanding of ILC differentiation and plasticity. Moreover, we revealed the generation of memory-like $CD4^+$ T cells in the developing human fetal intestine, indicating the exposure to antigens *in utero*. Additionally, our results provided evidence for the early-life immune compartmentalization across tissues as early as second trimester, which highlights the importance to investigate the immune system in the tissue niche. Together, our studies deepens our understanding of prenatal immunity and may provide a prenatal window of opportunity for the development of preventive strategies for the immune-mediated diseases later in life.

Nederlandse samenvatting

Het foetale afweersysteem

Het afweersysteem beschermt ons tegen pathogenen zoals bacteriën, virussen en schimmels. De cellulaire component van het afweersysteem bevat twee compartimenten, het innate en het adaptieve compartiment. Het innate afweersyteem bevat diverse phagocyterende cellen, antigeen presenterende cellen en lymphocyten die zorgen voor een snelle en aspecifieke reactie tegen binnendringende pathogenen. Tegelijkertijd reageert ook het adaptieve afweersysteem dat een veel specifiekere reactie opbouwt waarbij het gebruik maakt van antigeen-specifieke receptoren op B- en T-lymphocyten. Deze gecombineerde reactie zal uiteindelijk tot onderdrukking of eliminatie van het pathogeen leiden waarbij tegelijk geheugen wordt opgebouwd zodat hetzelfde pathogeen bij een volgende infectie sneller onschadelijk kan worden gemaakt.

Het foetale afweersysteem ontwikkelt zich tijdens de zwangerschap waarin het tolerant moet zijn tegen de deels allogene omgeving waarin de foetus zich ontwikkelt terwijl het ook moet worden voorbereid op de ontwikkeling van de bacteriële microbiota in de darm direct na de geboorte. Het afweersysteem in de darm is het grootste in ons lichaam en is continue blootgesteld aan zowel onschuldige voedselcomponenten als potentieel schadelijke bacteriën die dit afweersysteem van elkaar moet kunnen onderscheiden. Hoewel enkele studies al hebben aangetoond dat het foetale afweersysteem diverse componenten bevat, is vanwege de beperkte beschikbaarheid van foetaal weefsel de meeste kennis over het foetale afweersysteem gebaseerd op studies met navelstrengbloed. Een gedegen en brede studie naar de compositie van het foetale afweersysteem, met name in de darm, ontbrak tot op heden. In het huidige proefschrift heb ik dit onderzocht waarbij ik gebruik gemaakt heb van de nieuwste ontwikkelingen op het gebied van cellulaire analyse, waaronder massa cytometrie en RNA-sequentie bepalingen.

Massa cytometrie

Massa cytometrie is een nieuwe methode waar flow cytometrie is gekoppeld aan massa spectrometrie. In massa cytometrie worden metalen als reporters gebruikt waardoor de beperkingen van het traditionele gebruik van fluorochomen opgeheven wordt en tot 42 markers gelijktijdig geanalyseerd kunnen worden. Door deze revolutionaire innovatie kan de diversiteit en compositie van het afweersysteem tot in detail worden bestudeerd.

Het is ook mogelijk om deze techniek op weefselcoupes toe te passen, imaging massa cytometrie genaamd, waarmee het mogelijk is om de diversiteit en organisatie van het afweersysteem in de context van het weefsel te bestuderen.

Bepalen van genexpressie op cel niveau

Complementair aan massa cytometrie kan de expressie van genen van tienduizenden cellen op single-cel niveau worden bepaald waarmee de heterogeniteit van cellen en de ontdekking van zeldzame celtypes mogelijk wordt. Ook kan deze aanpak leiden tot de identificatie van cel ontwikkelingstrajecten en regulatoire netwerken binnen het afweersysteem.

Heterogeniteit en differentiatie van innate lymphocyten

In **hoofdstuk 2** is de heterogeniteit en ontwikkeling van het innate lymphoide cel compartiment in de foetale darm bestudeerd m.b.v. massa cytometrie. Hierbij werden 34 fenotypisch verschillende innate cel clusters gevonden, waaronder reeds eerder geïdentificeerde cel clusters zoals NK cellen en $CD127^+$ ILC subsets, maar ook een aantal niet eerder beschreven cel clusters. Gebruikmakend van de eigenschappen van de computationele analyse methoden konden mogelijke ontwikkelingstrajecten binnen het innate lymphoide compartiment worden geïdentificeerd die erop duiden dat de NK cellen en ILC3 kunnen ontstaan uit $Lin^-CD7^+CD127^-CD45RO^+CD56^+$ (int-ILCs) cellen. In functionele assays werd deze hypothese verder onderbouwd.

De opbouw van geheugen in de foetale darm

in **hoofdstuk 3** Is massa cytometrie gecombineerd met single-cel RNA-sequencen, T cel receptor sequencen, imaging massa cytometrie en flow cytometrie om het CD4 positieve T cel compartiment in de humane foetale darm te ontrafelen. Met massa cytometrie konden 22 CD4 T cel subsets worden onderscheiden waaronder naïeve, geheugen en regulatoire subsets. Middels RNA-sequencen werd dit verder onderbouwd. Opmerkelijk genoeg produceerden de geheugen cellen ontstekings mediatoren zoals TNF-alpha en Interferon-γ. Ook legde een gedetailleerde en geïntegreerde analyse van de data een differentiatie traject bloot dat de vorming van geheugen ondersteunde. Tenslotte wezen zowel de co-lokalisatie van de geheugen T cellen met antigeen presenterende cellen in het darmweefsel zelf, en aanwijzingen voor klonale cel expansie in de darm erop dat de geheugenvorming in de foetale darm het gevolg was van blootstelling van de foetus aan omgevingsfactoren in de baarmoeder.

Weefselspecifieke samenstelling van het afweersysteem

Om verder inzicht in het foetale afweersysteem te verkrijgen wordt in **hoofdstuk 4** beschreven hoe massa cytometrie is gebruikt om het innate and adaptieve afweersysteem in de foetale darm, milt en lever met elkaar te vergelijken. Deze analyse leidde tot de identificatie van 117 fenotypisch verschillende cel clusters waaronder zowel eerder beschreven alsmede nieuw ontdekte clusters. PCA analyse liet zien dat er grote verschillen zijn in de samenstelling van het afweersysteem in de drie geanalyseerde organen. Met imaging massa cytometrie werd deze weefsel-specifieke samenstelling bevestigd. Deze resultaten laten zien dat heel vroeg in het leven weefsel-specifieke compartimentalisatie van het afweersysteem ontstaat.

In **hoofdstuk 5** is massa cytometrie gebruikt om het afweersysteem te analyseren in darmbiopten van patiënten die leiden aan chronische darmontsteking en dat te vergelijken met dat in gezonde controles. Deze analyse liet zien dat zowel weefsel- als ziekte-specifieke immuun subsets konden worden gedetecteerd.

Conclusies

Het in dit proefschrift beschreven werk geeft een breed en gedetailleerd inzicht in de samenstelling van het foetale afweersysteem gedurende de zwangerschap. Hiervoor is gebruik gemaakt van een serie geavanceerde multiparameter technieken om cellen op single-cel niveau te bestuderen. Daarnaast is de functie van een aantal nieuw geïdentificeerde cel subsets bepaald. Hierbij is een nieuwe ILC subset gevonden die *in vitro* kan differentiëren tot zowel NK cellen als ILC3's. Hiermee is nieuw inzicht verkregen in de plasticiteit van het ILC compartiment. Verder werd gevonden dat gedurende de zwangerschap geheugen cellen worden gevormd in de foetale darm, wat er op wijst dat de foetus in de baarmoeder is blootgesteld aan omgevingsfactoren. Tenslotte is gevonden dat reeds heel vroeg in het leven het afweersysteem in verschillende weefsels een eigen signatuur heeft en het afweersysteem zich dus aanpast aan het weefsel waarin het moet functioneren. Om de functie van het afweersysteem goed te doorgronden is het dus noodzakelijk om het afweersysteem in de weefselcontext te bestuderen. Al met al biedt dit werk nieuw inzicht in de samenstelling en functie van het foetale afweersysteem, kennis die mogelijk gebruikt kan gaan worden om de ontwikkeling van ziekten die door ontregelde reacties van het afweersysteem worden veroorzaakt te voorkomen.

Curriculum Vitae

Na Li was born on May 28th 1987 in Tangshan, Hebei, China, where she grew up and attended primary, middle and high school. In 2007, she continued her studies on Veterinary Medicine at Human Agriculture University and obtained her bachelor degree in 2011 with the thesis titled "Compound of melamine and cyanuric acid on liver damage in mice". This thesis was honored as an 'Outstanding undergraduate dissertation'. During her master, she started to study on microbiology and immunology in Veterinary Medicine at Jilin University under the supervision of Prof. Liancheng Lei. In 2014, she obtained her master degree with the research project on "Screening of virulence genes contributing to *Streptococcus Suis Type 2* translate across the blood brain barrier and the mechanism of meningitis pathogenesis". During her bachelor's and master's degree from 2008 to 2013, she obtained National scholarship for Encouragement or first-grade scholarship. After obtaining funding for a four-year PhD abroad from the Chinese Scholarship, she started her PhD trajectory in October 2014 at the department of Immunohematology and Blood Transfusion (IHB) at Leiden University Medical Center (LUMC) under the supervision of Prof. Frits Koning. During her PhD, she investigated the development of the human fetal immune system using high-parameter advanced techniques, which have been described in this thesis. She was one of the pioneers using the mass cytometry and imaging-mass cytometry. This also resulted in covers for Journal of Experimental Medicine (Vol 215, No 5, May 2018) and Nature Immunology (Vol 20, No 3, March 2019) and a prize "DIRK BARTZ PRIZE FOR VISUAL COMPUTING IN MEDICINE" at Eurographics 2019. After her PhD graduation in 2019, she will move back to China and continue her research in the field of immunology.

List of publications

Li N, van Unen V, Guo N, Abdelaal T, Somarakis A, Eggermont J, Mahfouz A, Sousa Lopes SM, Lelieveldt BPF, Koning F. Early compartmentalization of immune cells in human fetal tissues revealed by high-dimensional mass cytometry. Accepted by *Frontiers in Immunology* (2019).

Li N, van Unen V, Abdelaal T, Guo N, Kasatskaya SA, Ladell K, McLaren JE, Egorov ES, Izraelson M, Chuva de Sousa Lopes SM, Höllt T, Britanova OV, Eggermont J, de Miranda NFCC, Chudakov DM, Price DA, Lelieveldt BPF, Koning F. Memory $CD4^+$ T cells are generated in the human fetal intestine. *Nature Immunology* 20, 301-312 (2019).

Li N, van Unen V, Höllt T, Thompson A, van Bergen J, Pezzotti N, Eisemann E, Vilanova A, Chuva de Sousa Lopes SM, Lelieveldt BPF, Koning F. Mass Cytometry Reveals Innate Lymphoid Cell Differentiation Pathways in the Human Fetal intestine. *Journal of Experimental Medicine* 215: 1383-96 (2018).

van Unen V, Höllt T, Pezzotti N, **Li N**, Reinders MJT, Eisemann E, Koning F, Vilanova A, Lelieveldt BPF. Visual Analysis of Mass Cytometry Data by Hierarchical Stochastic Neighbour Embedding Reveals Rare Cell Types. *Nature Communications* 8: 1740 (2017).

van Unen V, **Li N**, Molendijk I, Temurhan M, Höllt T, van der Meulen-de Jong AE, Verspaget HW, Mearin ML, Mulder CJ, van Bergen J, Lelieveldt BP, Koning F. Mass Cytometry of the Human Mucosal Immune System Identifies Tissue- and Disease-Associated Immune Subsets. *Immunity* 44: 1227-39 (2016).

van Unen V, **Li N**, Abdelaal T, Kooy-Winkelaar Y, Ouboter LF, Höllt T, Mearin ML, Witte AMC, Clemens C, Abraham S, Escher HC, Lelieveldt BPF, van der Meulen-de Jong AE, Koning F. Stratification of immune cell infiltrates in inflammatory bowel disease by high-dimensional mass cytometry. *Submitted*.

Acknowledgements

During my PhD, I met lots of difficulties in languages, culture differences and experiments, fortunately, there are people who are always willing to help, encourage me. Therefore, I would like to thank all the people who direct and indirectly contributed to my PhD studies.

First and foremost, I would like to express my sincere gratitude to my promoter/ supervisor Frits, who gave me the opportunity to study in such a fantastic research group. Many thanks to your trust, encouragement, insightful instructions and guidance in my research through these years. I will always be grateful to you. Welcome to China.

Vincent, my best Dutch friend and lab partner. Many thanks for your contributions to our joint projects, helping in adjusting myself into Dutch culture and the four-year accompany in science and life.

Moreover, I would like to express my appreciation to the warmhearted colleagues in our group, Yvonne (grandma of cells), Allan, Jeroen, Lin Zhou, Tessa (my paranymph), Natasja, Laura, Sanne, Fernanda and Nannan (my paranymph) thanks for all the valuable suggestions and sharing your knowledge and your life story with me. I really like to work with you.

I would also like to thank my fantastic colleagues in our IHB department. Thank you for all your support in my research including the valuable discussions, advice, sharing your knowledge and all the activities. A special gratitude to the Frank, Koen, Tanja, Anita, Gonca, Cynhia, Laura, Els, Amber and Anouk. I really enjoy the time in the IHB.

I would like to thank for Edwin, Simone, Guido for helping me sorting the cells in the 12-color panel; all the people in the CyTOF team especially Rene, Marjolijn, Sandra, Sanne, Juliette.

Thanks to all the fantastic Collaborators: Russian Academy of Sciences, Sofya, Evgeny, Mark, Olga, Dmitriy; Cardiff University School of Medicine, Kristin, James, Monika, David; Embryology, Maaike, fang, Maria, Kristin and Susana; Pathology, Marieke and Noel; LUMC and TU Delft computational team, Nicola, Ahmed, Jeroen, Tamim, Antonis, Thomas, Elmar, Anna and Boudewijn; Parasitology, Maria; Pulmonology, Padmini, Jan and Pieter.

Thank to my lovely Chinese friends. I feel so lucky to have so many good friends.

Thanks for your company and enriching my life. I really enjoyed the time to eat, discuss and play games together. Without you, I will be homesick.

Last but not least, I would like to express my appreciation to my beloved family. My parents, thanks for your encouragement and unconditional support. My little sisters, thanks for taking care of our parents. Dear my husband, Dr. Yuxiang Song, thank for your dedication. You always take care of me. Your love and encouragement inspire me to become better. Thanks to all the family for the greatest support.

Na Li
李娜
August 2019