

CURRICULUM VITAE
David Thomas Rach
Department of Microbiology and Immunology
Baltimore, MD 21201

Research scientist, computational biologist, and cytometry enthusiast. I have extensive experience with spectral and conventional flow cytometry (both analyzers and sorters), from high-parameter panel design through unsupervised data analysis. Passionate about training others to further best practices. I am the author and maintainer of 3 cytometry-focused R packages, that enable more reproducible analyses and further our fundamental understanding of human immunology.

Education:

University of Maryland, Baltimore	Baltimore, MD
<i>Ph.D.</i> , Molecular Microbiology and Immunology	August 2018 – June 2026

University of Wyoming	Laramie, WY
<i>B.S.</i> , Molecular Biology and Microbiology	August 2014 - May 2016

Northwest College	Powell, WY
<i>A.S.</i> , Natural Resources Biology	August 2011 - May 2013

Research Experience

Molecular Microbiology and Immunology PhD Student	Baltimore, MD
University of Maryland School of Medicine	August 2018 – Present

Thesis title: Characterization of Cord Blood Innate-like T cell responses in neonates born to healthy women and women living with HIV

Designed and validated high-parameter spectral flow cytometry panels (29-32 fluorophores) characterizing cytokine responses of innate-like and adaptive T cells from rare clinical specimens on a 5-laser Cytex Aurora. Created and implemented supervised and unsupervised analytical pipelines in R, facilitating troubleshooting of unmixing errors, revealing hidden functional heterogeneity and enabling more thorough and reproducible analysis in the acquired datasets. Trained 10+ individuals (from undergraduates to assistant professors) to carry out their own spectral flow cytometry experiments, including staining, unmixing and analysis.

Pre-Doctoral Staff	Baltimore, MD
UMGCCC Flow Core Shared Resource	April 2024 – Present

Performed all core operations encompassing user training, panel design, data analysis, cell sorting, instrument maintenance and quality control, procurement, as well as general troubleshooting and correspondence. Proposed, organized and co-led the yearly Baltimore Introductory Spectral Cytometry Course (BISCC), an intensive 5-day workshop, delivering lectures on unmixing and autofluorescence and led hands-on training and troubleshooting sessions. Created and taught the weekly Cytometry in R course, aimed at those without prior coding experience, both in-person and online (<https://umgcccfcsl.github.io/CytometryInR/>), reducing barriers to implementation of

unsupervised analysis. Developed automated quality control data processing for the core's Cytex and BD instruments, enabling real-time and historical instrument performance monitoring via a website (<https://umgccfcss.github.io/InstrumentQC/>).

INBRE Undergraduate Research Fellow
University of Wyoming

Laramie, WY
August 2014 - May 2016

Investigated Natural Killer cell responses to secondary and chronic *Toxoplasma gondii* infection, utilizing conventional flow cytometry, genetic and molecular biology approaches. Attained additional expertise in cell culture, animal handling and infection, chemical ablation of bone marrow and adoptive transfers.

Technical Skills

Flow Cytometry: Experienced operator of both spectral (Cytex Aurora) and conventional (BD LSRII, Guava easyCyte) analyzers. Proficient with the operation of cell sorters (Cytex Aurora CS and BD Aria II). Extensive experience with supervised analysis and troubleshooting of acquired data using R, FlowJo, FCSExpress, Diva and SpectroFlo software.

Bioinformatics: Advanced knowledge of R programming language, with functional knowledge of Rust and Python. Experience creating and validating unsupervised analytical pipelines to handle larger-than-memory spectral flow and mass cytometry datasets. Author and maintainer of 3 open-source cytometry-focused R packages (Luciernaga, Coereba, CytometryQC), with additional experience building Quarto websites and Shiny apps. Familiarity with Windows, Mac, and Linux (Debian) operating systems.

Laboratory Skills: Routinely performed blood processing, cell cryopreservation, cell culture, all performed under BSL-2 conditions with required blood-borne pathogen certification. Previously extensive experience with animal handling, IP and IG infections, chemical bone marrow ablation and adoptive transfers. Additional experience with ELISA, RT-PCR and fluorescence microscopy.

Teaching Experience

Instructor – Cytometry in R
UMaryland Greenebaum Comprehensive Cancer Center

Baltimore, MD
February 2026 – Present

Developed and taught a free weekly “Cytometry in R” course, targeting participants with no-to-little prior coding experience, offered both in-person and online, reaching over 600 active participants worldwide. Delivered 5 hours of instruction per week; additional responsibilities encompassed lesson planning and preparation, course website development and maintenance, and providing targeted participant feedback.

Science Education Volunteer

U.S. Peace Corps

Klikor-Agbozume, Ghana

May 2016 - July 2018

Taught high school biology and chemistry at the Klikor Secondary Technical School for 650 1st and 2nd year students. Lectured 20 hours/week, average class size of 50 students, with diverse ages and existing knowledge base. Prepared and graded lessons, assignments, practical demonstrations and exams. Additionally served on the exam and library committees.

Languages

Spanish (Fluent), Ewe (Advanced)

Publications

1. **David Rach**, Hao-Ting Hsu, Ngina Nampota, Godfrey Mvula, Felix A. Mkandawire, Oswald M. Nyirenda, Bernadette Hritzo, Francesca Boldrin, Giulia Degiacomi, Laura Cioetto Mazzabò, Riccardo Manganelli, Andrea G. Buchwald, Franklin R. Toapanta, Marcelo B. Sztein, Miriam Laufer, Kirsten E. Lyke, Cristiana Cairo. Cord Blood Innate-like T cell responses in neonates born to healthy women and women living with HIV. *Frontiers in Immunology* 2025 21 August, 16 <https://doi.org/10.3389/fimmu.2025.1628145>
2. Hao-Ting Hsu, Claudio Zanettini, Modupe Coker, Sarah Boudova, **David Rach**, Godfrey Mvula, Titus H Divala, Randy G Mungwira, Francesca Boldrin, Giulia Degiacomi, Laura Cioetto Mazzabò, Riccardo Manganelli, Miriam K Laufer, Yuji Zhang, Luigi Marchionni, Cristiana Cairo. Concomitant assessment of PD-1 and CD56 expression identifies subsets of resting cord blood V δ 2 T cells with disparate cytotoxic potential. *Cellular Immunology* 2024 Jan 01, 395-396, 104797 <https://doi.org/10.1016/j.cellimm.2023.104797>
3. Hao-Ting Hsu, Sarah Boudova, Godfrey Mvula, Titus H Divala, **David Rach**, Randy G Mungwira, Francesca Boldrin, Giulia Degiacomi, Riccardo Manganelli, Miriam K Laufer, Cristiana Cairo. Age-related changes in PD-1 expression coincide with increased cytotoxic potential in V δ 2 T cells during infancy. *Cellular Immunology* 2021 Jan 01, 359, 104244 <https://doi.org/10.1016/j.cellimm.2020.104244>
4. Daria L. Ivanova, Ryan Krempels, Stephen L. Denton, Kevin D. Fettel, Giandor M. Saltz, **David Rach**, Rida Fatima, Tiffany Mundhenke, Joshua Materi, Ildiko R. Dunay, Jason P. Gigley. NK cells negatively regulate CD8 T cells to promote immune exhaustion and chronic *Toxoplasma gondii* infection. *Frontiers in Cellular and Infection Microbiology* 2020 Jul 07, 10, 313 <https://doi.org/10.3389/fcimb.2020.00313>

Invited Talks

1. **David Rach**. Complex Data Analysis – “No cells left behind – Wrangling the full biological discover potential out of spectral flow cytometry (SFC) datasets”. Association for Biomolecular Resource Facilities (ABRF) Annual Meeting. Pittsburgh, USA. March 30, 2026.

Presentations

1. **David Rach**, Natarajan Ayithan, Xiaoxuan Fan. Cytometry in R: A free weekly course for flow cytometrist with no-to-little coding experience. Bioc2026. Seattle, USA. *Accepted - Oral*.
2. **David Rach**, Nginache Nampota-Nkomba, Godfrey Mvula, Felix A. Mkandawire, Oswald M. Nyirenda, Winter A. Okoth, Daniela Franco, Andrea G. Buchwald, Franklin R. Toapanta, Marcelo B. Sztein, Miriam K. Laufer, Kirsten E. Lyke, Cristiana Cairo. A semi-supervised pipeline for a comprehensive and scalable analysis of immune heterogeneity in human samples. Cyto 2026. West Palm Beach, USA. *Oral*.
3. **David Rach**, Natarajan Ayithan, Xiaoxuan Fan. Being Everything, Everywhere, All at Once: Open-Source Automation for Situational Awareness in SRLs. Cyto 2026. West Palm Beach, USA. *Oral*.
4. **David Rach**, Natarajan Ayithan, Xiaoxuan Fan. Cytometry in R: A free weekly course for coding beginners. Cyto 2026. West Palm Beach, USA. *Poster*.
5. **David Rach**, Kirsten E. Lyke, Cristiana Cairo. “Well, how bright does it need to be?” Investigating the interplay of fluorescent signature and brightness in single-color unmixing controls. Cyto 2025. Denver, USA. Exceptional Student Award finalist – *Oral*.
6. **David Rach**, Kirsten E. Lyke, Cristiana Cairo. “Are these autofluorescences in the room with us right now?” Quantifying impact of autofluorescence variation on unmixing. Cyto 2025, Denver, USA. *Poster*.
7. **David Rach**, Kirsten E. Lyke, Cristiana Cairo. “Well, how bright does it need to be?” Investigating the interplay of fluorescent signature and brightness in single-color unmixing controls. Cyto 2025. Denver, USA. *Poster*.
8. **David Rach**, Mikayla Trainor, Natarajan Ayithan, Xiaoxuan Fan. “Wait, when was QC last run???” Evaluating MFI drift after morning QC and its impact on unmixing. Cyto 2025. Denver, USA. *Poster*.
9. **David Rach**, Hao-Ting Hsu, Nginache Nampota, Godfrey Mvula, Felix A. Mkandawire, Oswald M. Nyirenda, Bernadette Hritzo, Ingrid Peterson, Franklin R Toapanta, Marcelo B Sztein, Miriam Laufer, Kirsten E. Lyke, Cristiana Cairo. V γ 9V δ 2 T cell responses in HIV-exposed Uninfected (HEU) Infants. 10th International $\gamma\delta$ T cell Conference 2023. Lisbon, Portugal. *Poster*.
10. **David Rach**, Hao-Ting Hsu, Nginache Nampota, Godfrey Mvula, Felix A. Mkandawire, Oswald M. Nyirenda, Ingrid Peterson, Franklin R. Toapanta, Marcelo B. Sztein, Miriam Laufer, Kirsten E. Lyke, Cristiana Cairo. Spectral flow cytometry analysis of Innate-like T cell responses in Malawian HIV-exposed Uninfected (HEU) Infants. Cyto 2023. Montreal, Canada. *Poster*.
11. **David Rach**, Hao-Ting Hsu, Nginache Nampota, Godfrey Mvula, Franklin R. Toapanta, Marcelo B. Sztein, Miriam Laufer, Kirsten E. Lyke, Cristiana Cairo. Innate-like T cell responses in HIV exposed uninfected Malawian infants. ASTMH 2021. Virtual. *Poster*.

References available upon request

Abstract

Title of Dissertation: Characterization of Cord Blood Innate-Like T cell responses in neonates born to healthy women and women living with HIV

Dissertation Directed By:

Cristiana Cairo, Ph.D., Assistant Professor

Department of Microbiology and Immunology

University of Maryland School of Medicine

Kirsten E. Lyke, M.D., Professor

Center for Vaccine Development

University of Maryland School of Medicine

Innate-like T cells (ILT), including $\gamma\delta$ T cells (particularly V δ 2 cells), Natural Killer T (NKT) cells, and Mucosal-associated Invariant T (MAIT) cells, serve as critical immunologic intermediaries that integrate features of both innate and adaptive immunity. These cells contribute to tissue homeostasis and play essential roles in host defense during infection and inflammation. Although ILT subsets are present at the maternal–fetal interface, their phenotypic and functional characteristics in early life remain insufficiently defined. In this study, we performed a comprehensive characterization of ILT populations in umbilical cord blood from neonates born to healthy women and women living with HIV. Leveraging high-parameter spectral flow cytometry, we defined the phenotypic diversity,

activation states, and functional potential of these rare cell subsets. We identified substantial heterogeneity within V δ 2 T cells at baseline and following in vitro stimulation, revealing distinct subsets with differential cytokine production and activation profiles. Comparisons between HIV-exposed and HIV-unexposed neonates demonstrated modest but meaningful perturbations in ILT composition and function among infants born to women with HIV. Specifically, we observed altered frequencies of ILT subsets ex vivo and shifts in the proportions of V δ 2 T cell subsets producing IFN- γ or TNF- α both at baseline and after expansion. Consistent with prior observations, maternal initiation of antiretroviral therapy prior to pregnancy was associated with attenuated immune dysregulation in exposed infants, underscoring the protective effects of early maternal treatment. To support these immunologic analyses, we developed new computational and programming tools optimized for the interpretation of high-dimensional spectral flow cytometry datasets. These tools enabled enhanced quality control assessments, advanced clustering, visualization, and subset identification, providing a framework for more robust and reproducible characterization of complex immune populations. Collectively, this work advances our understanding of innate-like T cell development and immune imprinting at birth, particularly in the context of maternal HIV infection. By integrating high-parameter cytometric profiling with novel analytical methodologies, this dissertation provides both biological insights and technical innovations that will enhance future studies of early-life immunity and maternal–fetal immunology.

Characterization of Cord blood Innate-Like T cell
responses in neonates born to healthy women
and women living with HIV.

by
David Thomas Rach

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, Baltimore in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2026

© 2026 by David Thomas Rach
CC-BY-SA

Dedication

Para vivir es demasiado el tiempo;
Para saber no es nada.
¿A qué vinimos, noche, corazón de la noche?
No es posible sino soñar, morir,
soñar que no morimos
y, a veces, un instante, despertar.

Rosario Castellanos
Nocturno

Acknowledgements

Growing up, the complementary close to the family newsletter was "Joy in the Journey". This PhD has been a long journey, and while there have been joyous times, it would have ended far differently without the support of countless individuals. While I can't name everyone, please know that I appreciate where our paths in this life crossed, and I wish you all the best in your own journeys.

My first and foremost thanks is to both the mothers and their infants, and the Malawian communities of which they are part, for without their participant and trust this study would have not been possible. I extend my thanks to the Malawian clinical team (including Godfrey Mvula, Ngina Nampota-Nkomba, Felix Mkandawire, Oswald Nyirenda and Kingsley Zuze) for their efforts in ensuring the cord blood was collected, cataloged, cryopreserved, and shipped.

Similarly, it was the decades long collaboration between the Blantyre Malaria project and Center for Vaccine Development that kept the lights on, providing the research infrastructure and institutional know-how that this study and many other global health projects like it relied on. I gratefully acknowledge Miriam Laufer, Sudhaunshu Joshi, Rashida Koudjra, Andrea Buchwald, Marcelo Szein for their contributions.

I owe my sincere thanks to the Flow Core Shared Resource at the University of Maryland Greenebaum Comprehensive Cancer Center (Xiaoxuan Fan, Brian Hahn, Karen Underwood, Mikayla Trainor and Natarajan Ayinathan), for the acquisition and maintenance of the cutting-edge instrumentation, as well as their initial training and support that sparked the evolution of my own flow cytometry interest.

I would like to thank my co-mentors, Drs. Cristiana Cairo and Kirsten Lyke for sticking with me throughout this journey, for their valiant attempts at extracting me from rabbit holes, and attempting to translate the hieroglyphics I brought back and rambled about. And for going above and beyond in editing manuscripts where my lack of formal English grammar class was apparent. I am indebted to my thesis committee, Drs. Tonya Webb, Franklin Toapanta, Shannon Takala-Harrison, Yutaka Tagaya and Margaret Feeney, and formerly Arlene Dent for their support and guidance through the winding mountain paths it took to get to this point.

Special thanks go to Hao-Ting Hsu for her flow cytometry and cell culture training during the first few years, as well as to all current, former and rotating members of the Cairo Lab (Lena King, Bernadette Hritz, Esosa Omorogbe, Mehnaz Falguni, Juliana Marshall, Winter Okoth) who tolerated my over-caffeinated explanations and haphazard experiment schedule. I am deeply thankful to my MMI cohort, Liron Marnin, Alexandra Soare and Rainer Butler for the support network throughout my PhD. I similarly would like to thank everyone at the Molecular Microbiology and Immunology (MMI) and broader Graduate Program in Life Sciences (GPILS) programs, namely Brett Hassel, Eileen Barry, Heather Ezelle, Bess Tracey, and Elice Garcia-Barca, for their continued support through the ups and downs. I wish to express my appreciation for Nevil Singh and Martin Flajnik for

endless thought-provoking immunology questions that prompted thinking beyond immunological borders. I extend my appreciation to the Departmental Seminar and Immunology Journal Club speakers for the inspiration and discourse early in my PhD before I burnt out.

I would like to thank Lucia Pierson, Geoffrey Parnell, and Christopher McCune for their IT assistance throughout the PhD, and apologize for the sheer number of "oh no, not him again!" my emails likely elicited given all the university firewall head-aches.

I would like to thank my parents (Tom and Nancy) and brother (Joe), for their care and support throughout. I am sorry I vanished for so many months on end and look forward to being able to return to weekly calls. I am also immensely grateful to my fiancé Daniela Franco, my world became brighter after I met you, and I wouldn't have made it this far without you in my corner. Tu!

Beyond graduate school, I want to thank the Canton Kayak Club, including the sunrise kayak and coffee crew (Rob Bennett, Stephanie and Mark Brick, et al.) for sharing what few in Baltimore experience. I am deeply grateful for the Returned Peace Corps Volunteers Baltimore group (Sandi Giver, Brittany Goetsch, et al.) for an outlet to talk about anything other than experiments. My sincere thanks to the Patapsco Valley 50K organizers and volunteers for their amazing work, Halloween costumes, and pocket-potatoes that make it my favorite ultramarathon. And my heartfelt appreciation to the countless unnamed individuals that collectively preserve and maintain the paths, trails, and waterways of both Baltimore city and the surrounding state parks, providing refuge in the stillness.

I would like to thank Christopher Hall, Laura Johnston, Hadley Wickham, Susan Holmes and Richard McElreath for their efforts at educational outreach, that were formative for much of my work. Additionally, many thanks to the online Cytometry and Bioconductor communities (Cytometry Discord, r/flowcytometry, Bioconductor Zulip) for both the assistance and nerd-snipes that opened unexpected doors to exploration.

I am thankful for my undergraduate mentors Eric Atkinson and Jason Gigley, that gave me a chance and shaped my interest in science, research and reproducibility regardless whether it was in the field or at the bench.

I would like to thank my former students and colleagues at the Klikor Secondary Technical School in Klikor-Agbozume, Ghana, for having inspired me to tackle graduate school in the first place. I apologize for not having kept in touch these past few years as life went off the rails.

My deepest thanks to the goats and observant monks that first discovered coffee, and my heartfelt thanks to the staff at Koba, Daily Grind, and THB that kept me overly-caFFEinated-but-content throughout both writing my paper and thesis.

And finally, I would like to acknowledge that three chapters in this thesis wouldn't have occurred if Alexis Perez-Gonzales hadn't suggested I just use GraphPad Prism to compare fluorescent signatures, which I took as a challenge, further cementing my descent into R madness.

Table of Contents

Acknowledgements.....	iv
List of Figures	viii
List of Tables.....	ix
List of Abbreviations	x
Chapter 1: Introduction	1
Chapter 2: Materials and Methods	7
Participant recruitment and study design	7
Sample collection and processing.....	8
Spectral flow cytometry analysis	8
Conventional flow cytometry analysis	10
Cord Blood V δ 2 T lymphocytes function after in vitro expansion	11
Supervised analysis	12
Semi-supervised analysis	13
Statistical analyses.....	15
Chapter 3: Luciernaga.....	16
Introduction	16
Results.....	23
Discussion	31
Chapter 4: Coereba	35
Introduction	35
Results.....	42
Discussion	48
Chapter 5: CytometryQC.....	52
Introduction	52
Results.....	54
Discussion	60
Chapter 6: Cord Blood Innate-Like T cells	63
Introduction	63
Results.....	65
Discussion	76

Chapter 7: Activated Cord Blood Innate-Like T cells	79
Introduction	79
Results.....	81
Discussion	89
Conclusions and Future Directions	92
Appendix	97
Observation: Vd2 After Expansion	97
Discussion	101
Observation: Sex-specific Differences	102
Discussion	104
References	116

List of Figures

Figure 1: Conventional and Spectral Flow Cytometry	17
Figure 2: Impact of unaccounted autofluorescent signatures on unmixing	20
Figure 3: Raw and Normalized Signatures	24
Figure 5: Leftover autofluorescence contamination has greater impact on cells with dimmer signal.....	29
Figure 6: Defining a Cluster.....	37
Figure 7: Coereba Visualizations	45
Figure 8: Levey-Jennings Plots	55
Figure 9: CytometryQC-generated dashboard landing page	59
Figure 10 Baseline heterogeneity of cord blood Innate-like T (ILT) cells.....	68
Figure 11 Semi-supervised analysis highlights cord blood Vd2 cell phenotypic heterogeneity between clusters	71
Figure 12 Altered marker expression of cord blood ILT subsets in HEU infants.....	75
Figure 13 Functional heterogeneity of cord blood ILT in response to polyclonal stimulation ex-vivo with altered cytokine expression in HEU infants.....	82
Figure 14 Semi-supervised analysis highlights cord blood Vd2 T cell functional heterogeneity between clusters after polyclonal stimulation.	86
Figure 15: Lower cytokine production in BCG-expanded Vd2 cells following restimulation in HEU infants.....	99
Figure 16 Differences in ILT global marker expression by infant sex.	103
 Supplementary Figure 1: Gating strategy for spectral flow cytometry (SFC) panel. ..	105
Supplementary Figure 2: Gating strategy for cultured day 17 conventional flow cytometry (CFC) panel.....	106
Supplementary Figure 3 Differentiation of Cord Blood ILTs.....	107
Supplementary Figure 4: Different visualization algorithms identify comparable ILT clusters.	109

Tables

Supplementary Table 1 SFC Antibodies	110
Supplementary Table 2 CFC Antibodies	112
Supplementary Table 3. Generalized Linear Model Results	115

List of Abbreviations

• AGPL3	Affero GNU Public License 3
• ART	Anti-Retroviral Therapy
• BCG	Bacillus Calmette-Guerin
• CBMC	Cord Blood Mononuclear Cell
• CFC	Conventional Flow Cytometry
• CHKV	Chikungunya Virus
• CMV	Cytomegalovirus
• DENV	Dengue Virus
• FBS	Fetal Bovine Serum
• FC	Flow Cytometry
• FCS	Flow Cytometry Standard
• FOSS	Free and Open Source Software
• FSC	Forward Scatter
• GPL3	GNU Public License 3
• GUI	Graphical User Interface
• HEU	HIV-exposed Uninfected
• HIV	Human Immunodeficiency Virus
• HU	HIV-unexposed
• ILT	Innate-Like T cells
• mAb	Monoclonal Antibody
• MAITs	Mucosal-associated Invariant T cells
• MIT	Massachusetts Institute Technology
• MFI	Median Fluorescence Intensity
• MC	Mass Cytometry
• NKTs	Natural Killer T cells
• OOP	Object-Oriented programming
• PBMC	Peripheral Blood Mononuclear Cells
• PBS	Phosphate Buffered Saline
• PFA	Paraformaldehyde
• PMA	Phorbol 12-myristate 13-acetate
• QC	Quality Control
• RAM	Rapid Active Memory
• RBC	Red Blood Cell
• rCV	Robust Coefficient Variation
• RT	Room Temperature
• SSC	Side Scatter
• SFC	Spectral Flow Cytometry
• SRL	Shared Resource Laboratory

- TCR T cell Receptors
- UDS Unmixing-Dependent Spread
- V δ 2s V γ 9V δ 2 T cells
- VL Viral Load
- ZOL Zoledronic acid monohydrate

Chapter 1: Introduction

Innate-like T cells (ILT), form a bridge between classical innate immune cells and adaptive T cells, playing an important role in infection and inflammation. After T cell Receptor (TCR)-dependent or TCR-independent activation (mediated by cell-subset specific ligands and cytokines released by innate cells respectively) (1-3), ILT rapidly mount an inflammatory response in the form of Th1-like cytokines and cytotoxic granule release (4, 5). This enables them to effectively participate in the early immune response before adaptive immune responses ensue. Consequently, they have been shown to play key roles in anti-microbial defense, as well as tumor immunity and immunoregulation(4).

The three known human ILT subsets in humans include the $\gamma\delta$ T cells (V δ 2s), Natural Killer T cells (NKTs) and Mucosal-associated Invariant T cells (MAITs). While extensive research has been carried out on these ILT populations in adults, less is known about them during the neonatal period and infancy. Among the reasons ILT are understudied in neonates is their low abundance in cord blood, and the multiple cell lineage markers required for their identification. Combined with reduced biospecimen availability for clinical samples, this has hindered immunologic study relying on conventional flow cytometry (CFC) (due to limited markers) or mass cytometry (MC)(due to cell quantity limitations). Consequently, prior studies required either specimen allocation for narrowly defined immunological panels (for SFC) or pre-sorting (for MC) to enrich for the target population.

ILT are present on both sides of the fetal-maternal interface during uncomplicated pregnancies (6-8). Maternal NKTs and MAITs increase in the placenta during the course

of pregnancy and produce Th1-like cytokines upon stimulation. Cord blood ILT predominantly express Th1-like cytokines, while cord V δ 2s and NKTs can also produce IL-17A in response to in vitro antigen stimulation (8-11). Both V δ 2s (specifically, V γ 9V δ 2s) and MAITs expand and mature during early life in response to microbial exposure. In contrast, invariant NKTs do not undergo significant expansion and show a central memory phenotype at the time of birth (4, 12).

In uncomplicated full-term pregnancies, several mechanisms converge to create and maintain fetal-maternal tolerance (13). However, maternal infections during pregnancy have the potential to perturb the unique immunologic microenvironment at the fetal-maternal interface through production pathogen-derived proteins and elicitation of inflammatory innate cytokines (14). This can disrupt immunologic homeostasis, as infections may cause activation of immune cell subsets (including ILT) triggering inflammation. Additionally, pre-natal exposure to maternal infections and resultant inflammation may impact the developing fetal immune system. In mice, IL-6 production in response to prenatal exposure to maternal *Yersinia pseudotuberculosis* led to amplified Th17 responses to infection later in life (15). Studies of human infants with prenatal exposure to pathogens including *Plasmodium spp.* and Hepatitis C reported increased Th1-biased cytokine responses to infection and immunization in early life (16-18).

In particular, several studies observed a higher infectious morbidity (due to enteric and respiratory infections) in infants exposed to human immunodeficiency virus (HIV) in utero but born uninfected (HEU) compared to their HIV unexposed (HU) counterparts from the same communities, especially during the first six month of life (19, 20). Multiple investigations have described underlying immunological defects in HEU infants, including

altered frequencies, memory differentiation and cytokine responses in T cells (21-25). However, results differed across studies likely due to differences in maternal clinical characteristics, including anti-retroviral therapy (ART) regimens, magnitude of viral load, and infant feeding modality (21-30).

Our knowledge of fetal ILT subsets, during either a healthy, uncomplicated pregnancy or following immune perturbation during gestation, is limited. The placental environment of pregnant women living with HIV displays inflammatory features (including production of $\text{TNF}\alpha$, IL-6 and IL-1 β), influenced by the peak and duration of viral load (VL) and/or maternal immune activation (31-34). Moreover, the use of ART is widespread, and recent studies suggest that certain antiretrovirals may impact the mitochondrial metabolism in a way that alters the immune system and contributes to the inflammatory placental microenvironment (35-37). Given these numerous considerations, we hypothesize cord blood ILT subsets may also be perturbed in HEU infants as a result of in utero HIV/ART exposure.

Prior technical limitations (38) related to the use of conventional flow cytometry (CFC) include the limited number of markers that can be assessed simultaneously. Mass cytometry (MC) can utilize >45 markers per panel but the slow acquisition rate, cellular destruction during processing, and costs, often limit the depth of analysis as well as analysis of cellular populations with low abundance, and have prevented accurate profiling of rare neonatal ILT subsets. These challenges are further exacerbated by the difficulty of discriminating differences from prenatal exposure to pathogens with the underlying heterogeneity of individuals within clinical cohorts.

Spectral flow cytometry (SFC) is a platform increasingly used within immunology (38). While the first commercially available spectral flow cytometer was introduced by Sony in 2012 (39), it was the launch of the Cytex Aurora in 2018 that popularized the approach. The rapid adoption of SFC is due to occupying a distinct niche compared to cytometry-focused technologies that preceded it, namely CFC and MC.

Mass cytometry (MC), unlike CFC and SFC, relies on mAbs conjugated to rare heavy metal isotopes (typically lanthanide series elements) normally not present in biological systems. While the panel size can be scaled up to 50-70 markers, the use of rare earth metals makes them subject to trade and political disagreements, and have driven the costs up considerably.

Both CFC and SFC rely on fluorophores, molecules that absorb light at a lower wavelengths/higher energy (i.e. excitation), and then release photons at longer wavelengths/lower energy (i.e. emission) (40). Every fluorophore has a characteristic excitation and emission pattern, which distinguishes them from each other (i.e. fluorescent signature). These photons are captured by the detectors and are subsequently amplified electronically on the basis of the instruments gain (i.e. voltage settings), with the total signal originating from every cell being recorded (41). Through various methods, using single-stained reference (or unmixing) controls, this total signal can be deciphered/deconvoluted back into individual fluorophore contributions (42-45). This allows inference for the individual cell regarding presence, absence, and relative abundance of each fluorophore on otherwise morphologically identical cells. Since flow cytometers can acquire tens of thousands of events per second, this approach permits quantitative assessment of immune cell heterogeneity at the system level.

CFC and SFC differ in how single-color reference controls are applied when deconvolving total signal from the full-stained cell back into its individual fluorophore contributions. In CFC, the number of detectors is limited, resulting in events being recorded for only narrow segments of the fluorescent spectrum, typically where the signal is highest (brightest) (41, 44, 46). Measurements from the detector are consequently interpreted as originating from that fluorophore. This results in a fluorophore being assigned to a designated detector. The signal contributions from different fluorophores that also emit photons captured by each specific detector (as characterized with the single-stained reference controls), are systematically/mathematically subtracted (i.e. compensated). This approach, while mathematically exact (42, 44, 45), limits fluorophore combinations that can be used in the same panel. CFC panels consequently select fluorophores based on their peak emission wavelength, and avoid fluorophores that share significant overlaps for these peaks.

In SFC, the detectors are set up in arrays that sequentially segment the spectrum by wavelength, and enable recording full emission spectrum for all fluorophores (42, 45, 47). These fluorescent fingerprints called signatures, are defined after subtracting out the contribution of background cellular autofluorescence (typically median fluorescent intensity values sourced from the unstained control) and incorporated together into a reference matrix. This reference matrix is used to decompose the total signal recorded from an individual stained cell into fractional abundances of individual fluorophore components using linear algebra approaches (typically ordinary least square approximation). Thus, even fluorophores with highly similar emissions for the peak detector region can be discriminated against by differences elsewhere in their spectra. This allows combination fluorophores in a similar panel compared to CFC, where fluorophores with near similar

peak emission spectra are incompatible. As a result, well-designed CFC panels are usually between 10-20 fluorophores, while SFC panels can contain 35-45 fluorophores, which approaches the marker coverage for MC panels.

Spectral flow cytometry, with its capacity to resolve more fluorophores than CFC and the ability to rapidly acquire a far larger number of cellular events than MC, may enable a more comprehensive phenotypic and functional analyses of rare cell populations and thus provides a platform to deepen our understanding of ILT cell subsets at birth. However, as a newly emergent technology, underappreciated technological hurdles may require novel solutions in order to rigorously profile rare subsets.

In this dissertation, I aimed to characterize cord blood Innate-like T cells. We developed a 29-color SFC panel which includes lineage, activation, and differentiation markers as well as cytokines, to analyze ILT cells in a cohort of Malawian infants, comparing neonates born to mothers with and without HIV infection. We characterized the marker expression profile of cord blood ILT cell subsets and assessed the production of cytokines following short-term polyclonal stimulation. In the process, I developed novel R packages to address underappreciated, yet key aspects related to the performance of unmixing, assessing instrumental variation, and methods for semi-supervised analysis targeting rare cell subsets. Utilizing the power of SFC at a level of detection not previously possible, we observed modest differences in frequencies of rare immune cell subsets and cytokine production between unexposed and HEU infants, as well as delineated general features of ILT in neonates.

Chapter 2: Materials and Methods

Participant recruitment and study design

As part of an observational longitudinal cohort study designed to examine clinical and immunologic impact of exposure to maternal HIV, pregnant women were recruited during their first antenatal care visit at two health centers in Malawi, East Africa; the Ndirande research clinic in Blantyre, an urban center, and the Bvumbwe Health center in Thyolo district, a rural area. HIV prevalence during antenatal care (ANC) was estimated to be 12-14% for both sites in 2023 (48). The study was approved by the Institutional Review Board of the University of Maryland School of Medicine and the Research Ethics Committee of the Kamuzu University of Health Sciences. Written informed consent was obtained from participants before screening and study procedures.

Our cohorts were defined on the basis of maternal clinical ART history and HIV viral load (VL), and stratified as follows: a) newborns of HIV uninfected women (HU); b) newborns of women living with HIV who began ART before conception and had an undetectable VL through pregnancy (HEU-lo); and c) newborns of women diagnosed with HIV infection and high VL ($>1 \times 10^4$ copies/ml) at >20 weeks of gestation (HEU-hi). Viral load was determined for pregnant women at enrollment and repeated at delivery along with CD4 counts. Infants who were found to have developed HIV infection by nine months of age were excluded from the study. Additionally, mothers and infants whose samples were used for this study screened negative for potential infection/exposure during pregnancy to malaria including *Plasmodium falciparum*, Syphilis (*Treponema pallidum*), SARS-COV-2, Dengue (DENV), and Chikungunya (CHKV) using dry blood spots and sera acquired during pregnancy and at delivery. Screening for congenital Cytomegalovirus (CMV) was

also carried out retrospectively using neonate urine samples collected during the first two weeks of life.

Sample collection and processing

All infant specimens collected for this study were obtained from uncomplicated term pregnancies. After delivery, cord blood was collected in citrate phosphate dextrose via puncture of the umbilical vein and was processed within 12 hours of acquisition utilizing a Lymphoprep (STEMCELL Technologies, Vancouver, Canada) gradient. Cord blood mononuclear cells (CBMC) were cryopreserved in freezing medium consisting of 90% fetal bovine serum (FBS, Gemini Bio, Sacramento, CA, USA) and 10% Dimethyl sulfoxide (Millipore Sigma, Burlington, MA)(49). Samples were stored in vapor phase of liquid nitrogen until shipment to Baltimore, MD, USA for sample processing.

CBMC were thawed, treated with DNase I (STEMCELL Technologies, 1 mg/mL), and resuspended in complete RPMI 1640 (Thermo-Fisher, Waltham, MA, USA). Cells were stained with anti-CD45 and counted on a Guava easyCyte (Millipore Sigma) and the cell numbers reported as the number of CD45⁺ events. CBMC were resuspended at 3×10^6 cells/mL in complete medium and incubated for 6 hours with protein transport inhibitors (Thermo-Fisher, Waltham, MA) or with Phorbol 12-myristate 13-acetate plus ionomycin (PMA+ionomycin) plus protein transport inhibitors (eBioscience Cell Stimulation Cocktail)(50). After incubation at 37°C, 5% CO₂ the cells were prepared for analysis.

Spectral flow cytometry analysis

After a 6-hour incubation, CBMC were stained with Zombie NIR viability dye (Biolegend) for 15 minutes at RT in the dark, followed by a wash. Cells were sequentially stained with cocktails of surface monoclonal antibodies (mAbs) specific for surface markers at specific

temperatures (see below and Supplementary Table 1) with or without tetramers (hMR1 5-OP-RU (MAIT), hCD1d PBS-57 (NKT) acquired from the NIH tetramer core, or their respective controls (hMR1 6-FP and unloaded hCD1d tetramers)(12, 31, 51) with a wash between steps. The first cocktail of mAbs (primarily against chemokine receptors, Supplementary Table 1) was incubated at 37°C for 30 minutes, the second cocktail (including tetramers) at room temperature (RT) for 40 minutes, the third cocktail (against other surface markers) at 4°C for 30 minutes. Following the last incubation, CBMC were treated with BD Pharm Lyse (BD Biosciences) for 3 minutes to lyse remaining red blood cells (RBCs) and immediately washed after the treatment. The cells were then fixed/permeabilized with Cytofix/Cytoperm (BD Biosciences), washed twice, and incubated at RT for 40 minutes with a mix of antibodies specific for intracellular cytokines. After washing, the samples were resuspended in 0.4% paraformaldehyde (PFA) and immediately stored at 4°C. All experiments were acquired on a 5 laser Cytex Aurora (Cytex Biosciences, Fremont, CA, USA) the following day, after instrumentation passed automated quality control. The median number of viable lymphocytes that were acquired for fully stained specimens was 6.8×10^5 cells [IQR 5.7×10^5 , 8.5×10^5].

For normalization purposes, each experiment included control peripheral blood mononuclear cells (PBMC) obtained from a single healthy adult donor in a single blood draw. Unmixing controls were pre-optimized following best practices (52-55). The choice of CBMC or PBMC for specific unmixing controls was determined in advance, to ensure that the median fluorescent intensity (MFI) of each single-color fluorophore was greater than or equal to that of the fluorophore in the fully stained sample. The same mAbs were used for the unmixing controls as for the respective fully stained panel. For each

experiment, the full set of unstained and single-color control specimens ($3\text{-}5 \times 10^5$ cells per control) were processed alongside the complete panel specimens. A backup set of single color-unmixing controls using beads (eBiosciences UltraComp Beads Plus, ThermoFisher) were processed as the other controls at the start of the experiment series.

Raw .fcs files for unstained and single-color unmixing controls were exported from SpectroFlo (v3.3.0, Cytex Biosciences) to R (R Statistical Software v4.3.1; R Core Team 2023) to evaluate autofluorescence and tandem stability using our R package Luciernaga (see Chapter 3). Unmixing controls exhibiting substantial signature variation across experimental runs were flagged for scrutiny during the unmixing process and replaced as necessary. Unmixing was carried out in SpectroFlo. The universal negative (unstained) controls always matched the respective single-color unmixing controls and fully stained specimens in terms of treatment and cell type. The general autofluorescence extract feature was utilized for all samples. We evaluated unmixing by employing 1xN plots in both SpectroFlo and R. When substantial unmixing errors were observed, we substituted brighter single-color unmixing controls from a different experimental run as the main method to resolve unmixing issues.

Conventional flow cytometry analysis

Cells from every study participant were stained with four separate CFC panels, each consisting of 8–9 markers (Supplementary Table 2). Following fixable viability staining with Zombie Aqua (Biolegend, San Diego, CA, USA) or Horizon 780 (BD Biosciences, Franklin Lakes, NJ, USA), the CBMC were washed and incubated with a cocktail of surface markers (Supplementary Table 2) for 20 minutes at 4°C. After RBC lysis, streptavidin was added for 15 minutes at RT in the dark.

For analysis of intracellular cytokines and cytotoxic effector molecules, the CBMC were fixed/permeabilized using BD Cytofix/Cytoperm (BD Biosciences) and intracellularly stained with an optimized panel of mAbs (Supplementary Table 2). Cells were resuspended in 0.4% PFA fixation buffer and analyzed within 24 hours on a 4 laser LSR-II flow cytometer (BD Biosciences). The median number of viable lymphocytes that were acquired for fully stained specimens was 4.3×10^5 cells [IQR 3.0×10^5 , 5.0×10^5].

Cord Blood V δ 2 T lymphocytes function after in vitro expansion

To assess V δ 2 cytokine responses and degranulation after expansion, CBMC were cultured in each of the following conditions: A) zoledronic acid monohydrate (ZOL, 0.5 μ M, Millipore Sigma) and human recombinant interleukin 2 (IL-2, 20 ng/mL, R&D Systems, Minneapolis, Minnesota, USA); B) Bacillus Calmette-Guerin (BCG, Pasteur strain) at a pre-optimized multiplicity of infection of one (kindly provided by Dr. R. Manganelli, University of Padova) in the presence of IL-2 (20 ng/mL); and C) IL-2 alone as a negative control. The samples were incubated for 17 days at 37°C with 5% CO₂ and fresh cytokine was added every three days doubling the cell culture volume with fresh medium on days 7 and 10. On day 14, only 2 ng/mL of cytokine was added to rest the cells (56). On day 17, the CBMC were counted on a Guava easyCyte (Millipore Sigma) and allocated either to analysis of differentiation/activation post-expansion or restimulation to assess cytokine production and granule mobilization via CFC.

Seventeen days after stimulation, the CBMC were restimulated with anti- $\gamma\delta$ T Cell Receptor (TCR)(clone B1.1, Thermo Fisher Scientific) in 96-well plates coated overnight at 4°C with anti- $\gamma\delta$ TCR (Clone B1, diluted 1:100 in PBS, 50 μ l/well). The cells were plated in triplicate (2×10^5 per well) in the presence of anti-CD107a AlexaFluor488 (clone

H4A3), GolgiPlug (brefeldin A, 1:1000), and GolgiStop (monensin, 1:1500) (BD Biosciences). Following a 6-hour incubation, the cells were collected and washed with cold PBS, before being processed for CFC as detailed above. At least 1×10^5 lymphocytes were acquired on a LSR II (BD Biosciences) analyzer the following day, as described above.

Supervised analysis

Following acquisition (for CFC) or unmixing (for SFC), fcs files were imported in FlowJo (v10.10.0, Beckton-Dickinson subsidiary, Ashland, OR, USA). Samples were visualized using the bi-exponential transformation, and a gating template was applied, with manual adjustments to exclude doublets, cell debris, and dead cells. For SFC, events in the live cell gate of each fully stained sample were exported as an individual file to a separate workspace where another gating template for ILT and non-ILT cell subsets was applied and adjusted (Supplementary Figure 1). V δ 2 cells were gated as CD7+CD3+V δ 2+ cells, using the clone B6, which is likely to recognize the V γ 9V δ 2 TCR (57). MAITs were gated as either CD7+CD3+CD161hiV α 7.2+ or CD7+CD3+CD161hihMR1 5-OP-RU+ cells for anti-TCR- or tetramer-based identification (2, 12, 51). NKTs were gated as CD7+CD3+V α 24J α 18+ or CD7+CD3+hCD1d PBS-57+ cells for anti-TCR- or tetramer-based identification. For CFC, pre-acquisition compensation was generated with single color controls and fine-tuned post-acquisition for individual specimens. A panel-specific gating template was applied and adjusted as needed (Supplementary Figure 2). For both methods, data including frequencies, MFI and counts was then exported as .csv files for use in statistical analysis.

Semi-supervised analysis

Events in the manually gated ILT sub-populations described above were imported from FlowJo to R using the CytoML and flowWorkspace R packages (58). Samples were bi-exponentially transformed, with quality control checks carried out on individual specimens using the PeacoQC (59) and Luciernaga packages (see Chapter 3). Normalization was conducted using either CytoNorm v2.0 or CyCombine packages (60-62), when required based on visual screening for batch effects (due to instrumental variation or other causes). The effect of normalization was visualized by plotting the data before and after normalization for each specimen and cell subset in R. To profile ILT features ex-vivo and explore the heterogeneity of cord blood ILT subsets, we implemented a semi-supervised analytical approach validating unsupervised analysis by manual gating. Using the interactive Shiny App in our R package Coereba (see Chapter 4), manually gated ILT cells from each specimen were visualized as individual bivariate plots. Individual markers were sequentially visualized across all specimens by switching the x-axis parameter while keeping the y-axis parameter constant. Automated minimum-density gating from the openCyto package (63) was used to estimate the split-point between positive and negative events for each given marker and individual. These automated gates were visualized and manually adjusted as needed through Coereba's RShiny application. The MFI values corresponding to these individualized split-point gates were then exported to a .csv file. The final gate placement for each given marker and individual was validated via their visualization on a 1xN plot using the Luciernaga package.

Following split-point validation, an identity column was generated with Coereba for each cell within an individual specimen, classifying the cells by the expression of their markers.

The frequency of positive cells was defined for each marker in every donor by the location of the relative split point. The identity column, alongside metadata corresponding to the individual specimen and treatment, was then appended as new columns in the .fcs file. After this process, all the cells originally identified as specific ILT subsets by manual gates across all individual samples were concatenated in a single .fcs file for each of the two treatments (PMA+ionomycin or unstimulated control). Each concatenated file was imported into FlowJo for dimensionality visualization using the Pairwise Controlled Manifold Approximation (PaCMAP) plugin (64). Markers that had been used to exclude events in the manual gating strategy, or were not expressed after a specific treatment, were not included as parameters to run the PaCMAP. Additionally, we ran other dimensionality visualization methods (tSNE, UMAP and PHATE (65-67)) to compare the output across algorithms (Supplementary Figure 4).

We manually gated clusters of the resulting PaCMAP on the basis of cell density distribution and marker expression, separating out distinct islands and peninsulas of cells from cells that aggregated on basis of separate markers. An individual .fcs file was generated for each cluster within the visualization and brought into R using the CytoML package. Coereba was then used to retrieve and reassemble the stored metadata for individual cells into a “tidy” format (68), allowing calculation of the frequency of cells derived from each individual across the various clusters. Using the stored manually annotated gating information, we calculated median expression frequency for every marker across all individuals in each manually identified cell subset or PaCMAP cluster. Expression levels were summarized as beeswarm plots or heatmaps using the Coereba and gt packages. The heatmaps use a combination of colors and symbols to display, for each

marker, the median of individual frequencies of positive cells, across ILT clusters. For this study, within the text we defined the analysis of manually gated ILT subsets as “global”, and that of PaCMAP-visualized ILT clusters as “local”. Finally, we brought into the R workspace the manually gated ILT subsets and generated co-expression matrices for the individual markers using the CytoGLMM package (69).

Statistical analyses

All analyses were performed using R Statistical Software (v4.3.1). For null hypothesis significance testing (NHST), a p-value < 0.05 was considered significant. We assessed normality using D’Agostino Omnibus K2 test (implemented in fBasics R package), visually inspecting the distributions with histogram and Quantile-Quantile plots. Where needed, we compared potential differences in the variance of distributions between exposure groups using the Bartlett’s test. Continuous variables in the text are presented as medians for proportions of positive cells in the parent gate, followed by interquartile ranges (e.g., X [Y,Z]). Pairwise comparison of means or median frequencies were conducted using a two-tailed student’s t-test or a Wilcox non-parametric test for normal or non-normal distributions respectively. For comparisons between three groups, we employed either a One-Way Anova (Type II) or a Kruskal-Wallis test for normal and non-normal distributions, respectively. P-values were adjusted for multiple comparisons using Benjamin Hochberg correction. To assess whether sex-specific differences were independent of HIV/ART exposure status, we fitted two Generalized Linear Models with the stats package using the following formulas: 1) $\sim$ Infant sex + HIV/ART exposure group, where HIV/ART exposure group was a factor with 3 variables (HU, HEU-lo, HEU-hi); and 2) $\sim$ Infant sex + HIV/ART exposure group + Infant sex *HIV/ART exposure.

Chapter 3: Luciernaga

Introduction

In cord blood, innate-like T cells (ILT) are found at low frequencies, with a prevalence of <0.5% of T cells (6, 7, 50). After activation, ILT produce Th1-like cytokines. Additional information about these rare cell populations during homeostasis or following infection is limited due to several factors. Identifying ILT populations requires numerous markers which limit the platforms upon which they can be measured (2, 4, 70). The limited availability of biospecimen in clinical settings further hinders the ability to thoroughly characterize these cells via a CFC panel. While MC panels can accommodate numerous markers, the low acquisition rate and the rarity of cord blood ILT requires considerable acquisition time per specimen to accumulate enough events to properly characterize these cell populations.

In contrast, the SFC platform, occupying a distinct niche compared to other cytometry-focused technologies, can be particularly suitable for characterizing rare cell populations (38, 47). Its use of the full emission spectrum allows for the incorporation of additional markers compared to CFC (Figure 1A), while the rapid acquisition rate enables the accrual of enough events in a few minutes, ensuring statistically meaningful analysis (40). This may enable the interrogation of previously unexplored heterogeneity in these rare cell populations. However, to fully leverage the SFC potential for biological discovery, addressing and resolving several technology specific challenges is fundamental.

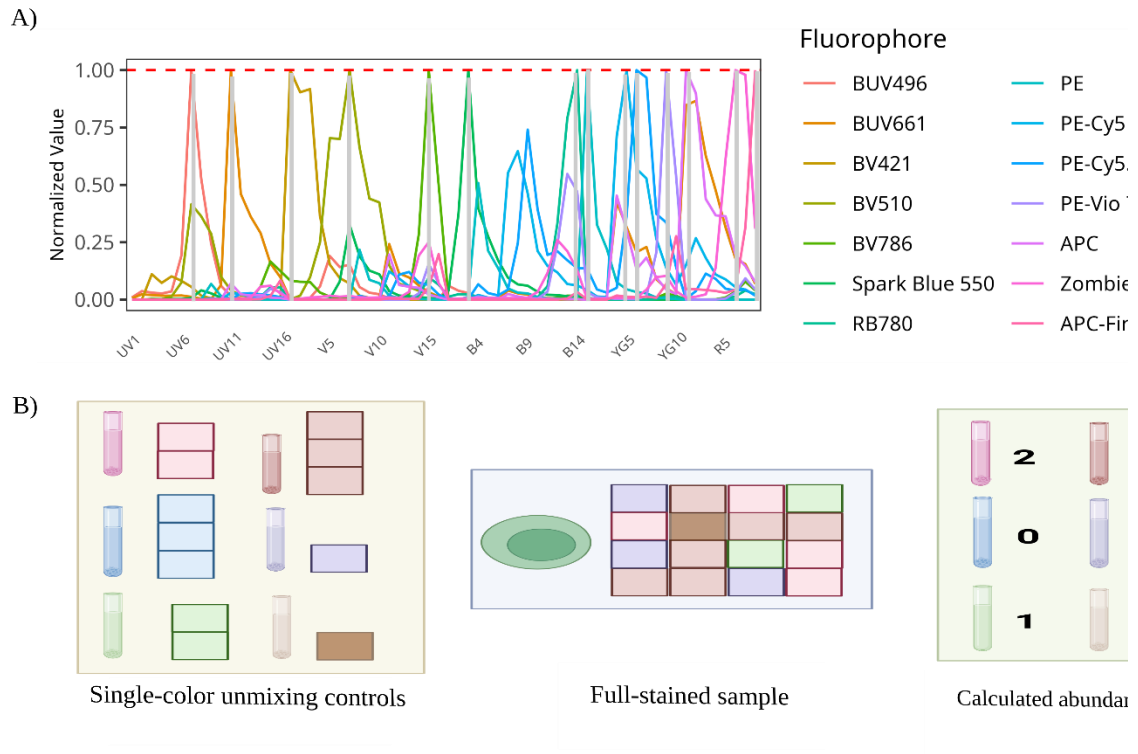


Figure 1: Conventional and Spectral Flow Cytometry. Normalized fluorescent signatures a 14-fluorophore panel. Gray shaded vertical lines indicate emission peak detector. SFC uses entire signature, while CFC uses only the gray-shaded region. B) Simplified example of correct unmixing when single-color normalized signatures are representative of what is found on the full-stained cells.

The quality of the final unmixed SFC data depends on the quality of the relative unmixing controls (both single color and unstained)(42, 44, 71). Unmixing controls are critical for the deconvolution of total SFC signals into the respective fluorophore abundances, as their individual signatures provision the reference matrix (42, 44). For each single-color reference control, following autofluorescence subtraction, the remaining MFI values for each detector are presumed to have originated from the fluorophore. These values are used to distinguish the fluorescent signature of one fluorophore from that of others within the reference matrix (42, 43). As individual fully-stained cells are unmixed, these normalized single-color signatures are referenced to approximate the presence and abundance of respective fluorophores on the full-stained cell (Figure 1B)(42, 43). Consequently, anything that affects the values of the signatures within the matrix, can result in attribution errors and unmixing results not reflective of the true abundance of a fluorophore (Figure 1C). Despite their critical importance, limited tools exist to profile the fluorescent signatures used in the matrix. This is particularly problematic given that there are several common causes known to lead to inaccurate fluorescent signatures.

Fluorescent-conjugated mAbs are subject to physical and biological forces that may affect their fluorescence (40). Reactive oxygen species (ROS) produced by cells under different circumstances can degrade the fluorophores (72). This can decrease the overall signal contribution from particular segments of the light spectrum, altering the final normalized signature. Additionally, tandem fluorophores have potential to degrade back into their respective components, resulting in an underestimation of their own abundance, while inflating that of the parent fluorophore (73). This can amplify the spectral similarity to the other fluorophores, and results in increased uncertainty, and decreased resolution (42, 44).

Low-level fluorophore signal due to reasons other than antigen-specific binding of mAbs (e.g. Fc receptor capture, phagocytosis of PE and APC, or general stickiness) can impart a dim “positive” signal to cells that should have be interpreted as “negative” for a marker (40). This increased spreading of the negative population signal, and loss of resolution between the truly positive and negative population, further complicates the discrimination of dimly positive cell subsets from negative cells with high background (74). Similarly, it limits our ability to utilize internal unstained autofluorescence negatives for unmixing, creating a dependence on a universal unstained control.

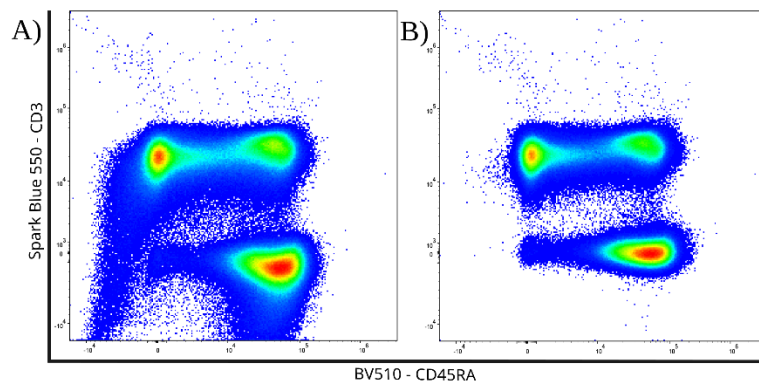


Figure 2: Impact of unaccounted autofluorescent signatures on unmixing. The plots depict the impact of unmixing a 29-color SFC panel, with a reference matrix containing only one (A) or two (B) signatures when two autofluorescent signatures are present. The fluorophores depicted are BV510 CD45RA on the x-axis, Spark Blue 550 CD3 on y-axis.

To identify the real signature of each fluorophore, an averaged autofluorescence signal has to be subtracted from the signal of every single color control (42). The inherent assumption is that the cells share a homogenous autofluorescence matching the signature in the single-color controls.

While often the case for peripheral blood, this differs in tissues, or cells that are activated. Autofluorescence of individual cells is driven by the expression of a mix of biological molecules (ranging from amino acids like tryptophan, to intermediates like NADPH, cyclic molecules, and actin) (75, 76). A shared autofluorescence signature is the background signal of cells that share relative proportions of these intracellular autofluorescent molecules (74). However, an increased expression of one autofluorescent molecule, or the production of a novel autofluorescent molecule with a different emission pattern has the potential to drive the appearance of additional discernable signature(s). Such a signature would be unaccounted for within the unmixing matrix if relying on the “average signature”. The reference matrix is consequently incomplete, with the additional signal being misattributed to other fluorophores, introducing unmixing errors (characteristically defined by distorted appearance and extreme negative values) (43, 77). (Figure 2A, B). Likewise, if an additional autofluorescent signature is included in a matrix when it is not present, it can hamper resolution (75, 76). There is currently no systematic way to decide whether an autofluorescence signature is similar enough to the average to be treated as the average, or whether it should be treated as a distinct autofluorescence signature and included in the reference matrix.

Given the critical role unmixing controls (both single-color and unstained) play in the deconvolution of fluorescence signals from mAb-stained samples, and the myriads of ways

they can wind up distorted by experimental conditions, the limited existence of tools to screen SFC unmixing controls for the aforementioned issues represents a significant technical challenge. Consequently, there is a need for additional tools to prevent these unmixing errors in SFC.

Commercially available unmixing software is slow to adopt best practices on their platforms. Pre-unmixing, spectrum plots display raw fluorescence signals of acquired cells. This enables the observation of major changes to the signature across the entire cell population, although the applied density function often hides less prevalent variant signatures. After unmixing, use of 1xN plots can be used to visually screen for the presence of unmixing errors (53, 71).

While community guidelines exist to identify and mitigate unmixing issues, implementation remains laborious. In medium to large SFC panels, with a greater number of tandem fluorophores, and significant inter- and intra-laser spreading between the fluorophores, deciphering the factors driving the identified unmixing issues can be exceedingly difficult to determine, as misattribution of photons has a domino effect, impacting the unmixing of multiple fluorophores(43).

Existing free and open-source software (FOSS) packages for quality control of flow cytometry files (including the Bioconductor packages PeacoQC, FlowCut, FlowClean, FlowAI) are primarily focused on identifying instrumental issues arising from fluidic or laser fluctuations across instrumental platforms (55, 59, 78-80). They accomplish this by monitoring cytometer event acquisition rate, and change in MFI-brightness for individual fluorophores across time. These tools would not detect altered fluorophore signatures in an individual experiment or an experimental series. Similarly, normalization software

packages (like CytoNorm and CyCombine) are capable of detecting and correcting batch effects by examining variation of marker expression in controls, and applying corrections to normalize to the desired distribution (60-62). These apply a broadly corrective force at the population level. However, is not well characterized whether these broad corrections based on more prevalent cell populations are appropriate when applied to both rare cell populations with unique marker expression patterns; and cells where unmixing-dependent spread (UDS) resulted in less-than-ideal unmixing (43, 77).

I therefore set out to develop software tools suitable for SFC, to profile all fluorescent signatures present in unmixing controls in an automated manner. By enumerating these outputs, I seek to screen for quality issues in unmixing controls (both single color and unstained), to allow for systematic and accurate identification of unmixing issues in full-stained samples, and to generate testable hypotheses to solve them prior to unmixing.

Results

Working in the programming language R, I set out to build code functions to return the averaged raw MFI across all detectors for a given gated population of cells (Figure 3A). Following the identification of the detector with the brightest MFI, values of all the other detectors were divided by it (i.e. normalized), resulting in values ranging from 0 to 1 (Figure 3B). The final results were then plotted, returning raw or normalized average fluorescent signatures of a cell population.

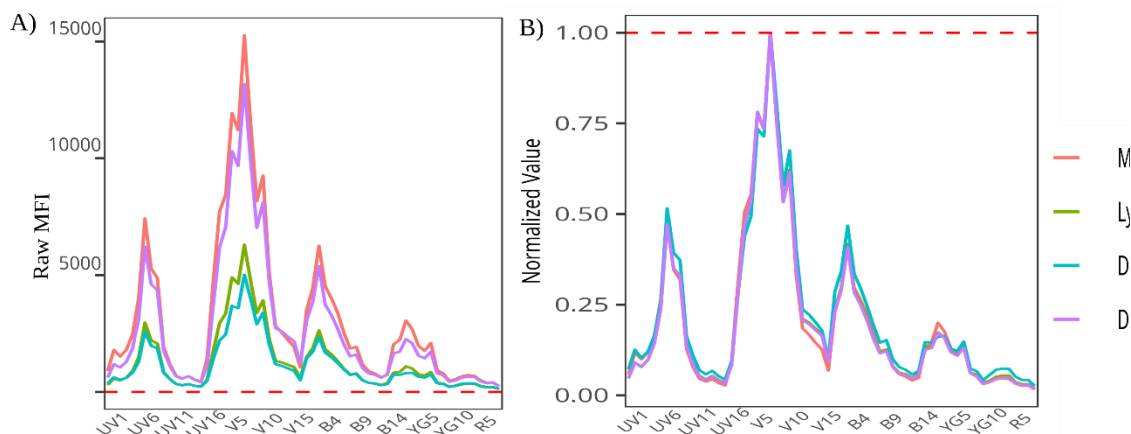


Figure 3: Raw and Normalized Signatures. Line plots depict averaged fluorescent signature of several cell populations defined by FSC and SSC. Depicted are A) the raw fluorescent signatures, that differ in terms of brightness, and B) the normalized fluorescent signature.

For single-color unmixing controls, the MFI values for each detector of both positive and negative gated populations are calculated, the background is then subtracted, normalized, and the fluorophore's fluorescent signature plotted. Averaging fluorescent signature at the population level, however, risks “averaging out” any variant signatures that may be present.

Starting with unstained samples, I modified the existing signature return function to work at the level of individual cells, generating individually normalized values across all detectors for each cell. I then attempted to group the cells based upon shared autofluorescent signatures. However, even with samples of 3 million CBMCs, stochastic noise across 64 detector columns resulted in an enormous quantity of potential combinations. Despite binning individual detector measurements based on percentiles (in

increments of either 20 or 50), few cells ended up clustering together when all 64-detector channels were used. With most fluorophores being defined by characteristic peaks, I created a mechanism to agnostically identify these peaks within an individual fluorescent signature, and use them for clustering signatures of individual cells. This approach prioritizes the peak detectors that form the majority of the area-under-the-curve results, and reduces the impact of noise from regions that contribute little to the overall signature.

I implemented an approach to identify local minima and local maxima within fluorescent signatures, by applying a rolling criterion. For all identified maxima, peaks meeting a certain height were returned and their corresponding detector values were subsequently used for clustering. Normalized signatures of individual cells could therefore be classified based on location and height of their peaks. Working with 10% bins, identifiable groupings for unstained cells, and fluorophores with up to 5 peaks (the vast majority of fluorophores) were returned. Having enabled identification of cells on the basis of shared autofluorescent peaks within a pre-defined height limit, I set up a function to extract these clusters of cells with shared signatures from their original .fcs file and store them as unique secondary .fcs files. By ensuring keywords were retained, the purified signature .fcs files were recognized by commercial software and could be used for unmixing (81-83). In parallel, clusters could be described by counts and ratios for use in exploratory data analysis and reports.

These tools were then adapted for use with single-color controls, starting with an externally provided autofluorescence signature derived from an unstained control for use in background subtraction. Subsequently, I derived a method to safely utilize internal autofluorescence from the same single-color control, using cells with peaks equivalent to primary autofluorescence, and selecting for those with minimal contribution to the

respective fluorophore's primary detector. Upon cross-validation, the signatures generated using the external and internal autofluorescence overwhelmingly matched, with a couple notable exceptions. Namely, cases like CD45 (a uniform positive expression with negative cells being completely mismatched autofluorescence) and viability dyes (with a high degree of positive signal), required the use of an external autofluorescence (71).

I then sought to verify that the returned signatures matched those published by the instrument and antibody vendors, by securing a collection of reference signatures for individually chosen fluorophores. With access to these signatures, I wrote code functions to enable comparison of the fluorophore signatures, utilizing metrics including cosine (similarity matrix), kappa (complexity index), and collinearity (unmixing dependent-spread) (77). I then compared signatures isolated from my individual unmixing controls with those of the vendor to evaluate their similarity. These closely matched when the community-defined proper unmixing guidelines were followed.

After generating automated return of fluorescent signatures for both single-color and unstained unmixing controls, I proceeded to ascertain the effectiveness of detection of variant signatures in an automated fashion. Across the board, for the single-colors, the script quickly identified both mislabeled and accidentally contaminated antibodies (1% APC-Fire750, 99% APC-Fire 810, likely resulting from accidental pipette-tip reuse), and unmixing controls where the tandem fluorophores were beginning to decouple. Notably, I was also able to correlate the unusual overly-corrective normalization patterns carried out by CytoNorm version 2 with the breakdown of the tandem APC-Fire 810 into APC for specific experimental days.

When signatures acquired across different experimental days were plotted together, it allowed for the visualization of changes in autofluorescence arising from experimental processing delays. These plots enabled rapid screening of fluorophore performance across an entire experimental series for unmixing issues, with the processing subsequently automated to return results for all fluorophores as a pdf report. By implementing adjustments for the ratio threshold, we further defined the amount of acceptable signature noise for a given specimen. Strikingly, differences in unstained signatures for individual specimens were noted, allowing for the rapid identification of individual specimens where multiple autofluorescent signatures were present. One discernable pattern was a difference in the height of secondary and tertiary autofluorescence peaks between CBMC and PBMC.

Having established a way to identify specimens containing more than one autofluorescence signature, I sought to understand when autofluorescence signatures are different enough to trigger unmixing issues, particularly those noticeable on 1xN plots against fluorophores that resemble autofluorescence (BUV496, BV510, Spark Blue 550, etc.) (75). For CBMC, within which autofluorescence is homogeneous for most samples, when a primary peak was shared, the presence of secondary, unaccounted for autofluorescences with variation of the second and third peaks heights by up to 30% compared to that represented by the autofluorescence signature in the unmixing matrix resulted did not impact the unmixing within our panel.

By contrast, presence of secondary, unaccounted for autofluorescences with different primary peaks than that represented by the autofluorescence signature in the unmixing matrix resulted in an obvious unmixing error (Figure 2A, B). This occurred even if the secondary autofluorescence was rare (<0.5%). Most secondary autofluorescence causing

unmixing errors were characterized by a primary peak in the UV7 (rather than in the V7 detector).

I then set out to understand the interplay between signal brightness and signature. Utilizing a single-color control, gated on the positive peak, I subdivided the peak into percentile bins (increments of 10), and extracted all the individually normalized signatures identified within each bin (Figure 4 A,B). For antigen-fluorophores combinations that resulted in a bright MFI, minimal difference in signature were present, with one or two signatures being returned from each bin (Figure 4C).

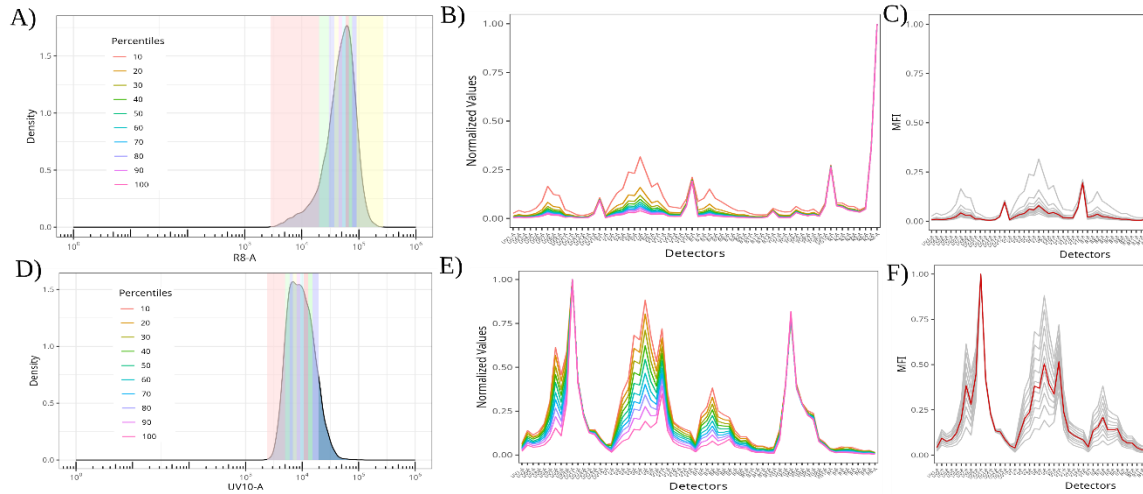


Figure 4: Leftover autofluorescence contamination has greater impact on cells with dimmer signal. Single-color stained unmixing controls for CD38 APC-Fire 810 (top) and CCR4 BUV615 (bottom) were gated for positive staining cells. Cells were then split into 10 percentile bins. Average normalized signatures for cells present in each bin were derived. Depicted are A) Percentile bins within along the positive gated cells. B) Normalized fluorescent signature within each bin, and C) Comparison of the average normalized fluorescent signature for positive cells (red) compared to the percentile bin signatures (gray) (D-F).

In contrast, antigen-fluorophores combinations that resulted in a dim MFI (and were not fully separated from the negatively stained cells) had substantially more normalized signatures retrieved across percentile bins (Figure 4D). These signatures, when plotted, showed consistent fluorescent signatures for the fluorophore, but with a step ladder appearance to additional peaks coinciding with the main autofluorescence peak detectors (Figure 4E, F). We determined that due to the variation in autofluorescence present in individual cells at those detectors, subtraction of an averaged autofluorescence calculated from the unstained control left a discrepancy, with cells that were under or over-extracted, resulting in a step-ladder appearance. Within dimmer antigen-fluorophores combinations, this residual autofluorescence contribution made up substantially more area-under-the-curve than it did for brighter antigen-fluorophores combinations, leading to increased impact on the overall fluorophore signature. When the fluorophore signatures containing autofluorescence contamination were used in unmixing, the increase overlap/collinearity resulted in unmixing issues. By identifying the issue, preparing PE-Cy5 CD25 and BUV615 CCR4 single color controls with beads instead of cells substantially improved unmixing results as the residual autofluorescence contamination was eliminated.

The development and optimization of quality control application for unmixing controls allowed for a pipeline by which any acquired SFC experiment folder in raw format can be screened. This was further expanded into an automated system to screen all acquired reference controls on the flow core's spectral instruments to preemptively identify unmixing issues, forming an early warning system before users identified issues within individual panels. Finally, these tools were incorporated into our Luciernaga R package,

that is now available as a developmental package release via GitHub while undergoing submission to Bioconductor.

Discussion

I set out to develop a collection of automated tools to enable the identification of low-quality unmixing controls (unstained and single color) which result in inaccurate unmixing. This work led to the creation of an R package, *Luciernaga*. Unlike traditional approaches focused on fluorescent signatures at a population level (75, 76), *Luciernaga* operates at the individual cell level. Within unstained controls, normalized signatures are clustered based on the shared peak detectors and height. The majority of cells within our CBMC unstained samples share similar fluorescent peaks. Through iterative unmixing, we showed that cells sharing a common primary peak detector, allow for autofluorescence variance by up to 30% in the second and third peaks with similar unmixing results.

In contrast, unstained cells that possessed different primary peaks resulted in obvious unmixing errors, even when these cells had a low relative abundance. This was driven by unmixing occurring at the level of the individual cell, resulting in the matrix requiring both autofluorescence signatures to properly unmix both populations. While clearly beneficial for the resolution of the autofluorescence-similar fluorophores, the incorporation of these additional autofluorescence signatures did not significantly impact the resolution for any other fluorophore, despite increase in collinearity in our panel.

This was striking and emphasizes the need to acquire sufficient unstained cells to enable identification of rare autofluorescent signatures. It also serves to define the extent that divergent signatures can deviate before unmixing errors arise. As current commercial software approaches focus on manually identifying separate autofluorescence signatures

based on the primary peak-emission detector and a cosine value (75, 76), implementation of automated return of isolated autofluorescence signatures with our selection criteria could render the process more accurate and result in improved data quality. This would especially be useful for the unmixing of cryopreserved longitudinal clinical specimens where processing was suboptimal, and additional autofluorescence signatures that may be present.

Our approach could similarly be applied to single-color controls, where we illustrated why some of the community best-practice guidelines work mechanistically. Pools of bright staining cells contained relatively few individual normalized signatures, however, these signatures become more abundant in dimmer staining cells. These bright staining cells, in terms of MFI are driven by antigen abundance on the surface of a cell, and the photon emission of the respective fluorophore. As a fluorophore-conjugated antibody binds and accumulates on the cell surface, the signal of the fluorophore eventually matches that of the autofluorescence, and both are co-designated as primary detector on the resulting signature. As more mAb binds, the raw MFI height of the primary peak increases, while that of the autofluorescence remains the same. At a certain point, the variation in autofluorescence after subtraction contributes little to the overall normalized fluorescent signature.

By contrast, with dimly staining fluorophore-antigen combinations, while the tipping point where mAb binding to surface antigen emits a brighter signal than autofluorescence is reached, the primary fluorophore peak never becomes bright enough to render the autofluorescence contribution inconsequential. Therefore, inherent variation of autofluorescence at the cellular level that remains following background autofluorescence subtraction distorts the normalized signature. For the dimmer staining cells, residual

autofluorescence after subtraction therefore contaminates the overall fluorescent signature, increasing collinearity and contributing to unmixing errors. By replacing these dimly staining cell signatures with bead controls, the unmixing outcomes significantly improved, as the fluorescence signature was more accurately captured, and free of additional peaks resulting from autofluorescence contamination.

This highlights why researchers taking a binary stance on whether to use beads or cells for the whole panel is not advisable. While our own preference has been to use cells, in this case we encountered a previously unappreciated situation not covered in the best community practice guidelines or peer-reviewed literature that highlights the need to use beads for specific marker/fluorophore combinations.

Additionally, we were able to characterize signatures arising from non-specific staining, hungry monocytes, and tandem degradation. Notably, in the context of the cord blood ILT panel, we were able to pinpoint cases where APC-Fire 810 degraded into respective components, resulting in single-color controls containing both APC and APC-Fire 810. The ensuing “single-color” control showed increased collinearity and cosine similarity with the APC control, sparking unintended consequences and inaccurate data output. One of the first applications of the *Luciernaga* R package was visualization of fluorescent signatures across an entire experimental series. This resulted in rapid identification of isolated contaminated antibodies resulting from pipetting mishaps.

Being able to pre-screen for contaminated or degraded antibody vials is an invaluable and cost-saving tool, given that a single contaminated or degraded vial of antibody added to an antibody-cocktail could severely reduce the overall resolution and accuracy of cell population staining. Being able to proactively spot and replace these vials can therefore

avoid the need to repeat experiments, repurchase antibodies for an entire panel, and repeat time-consuming analysis hours.

Luciernaga provides an additional tool application to the family of flow cytometry R packages. The ability to package cells within secondary, fully compliant .fcs files that consist of single purified signatures has enabled deeper understanding the impact of parameters within the unmixing process. While still in a developmental phase, with incorporation of additional documentation and validation the package will be ready for Bioconductor submission, as a useful tool in the computational cytometry toolbox, allowing for improved unmixing results, and increased confidence in the underlying data interpretation.

Chapter 4: Coereba

Introduction

Due to technological limitations which previously prevented deep interrogation of rare cell populations, little is known about cord blood ILTs in humans. Constraints for ILT detection with CFC or MC include the fact that numerous markers are required in their identification, cord blood ILTs are found at low frequencies ($< 1\%$ of T cells), and limited biospecimen quantity can be drawn from a neonate, reducing the number of cells available for analyses. SFC allows the characterization of these rare cell subsets, however, in doing so, millions of non-target cells are also acquired, generating an enormous quantity of data. Given the rarity of our target cell populations of interest, existing unsupervised algorithmic approaches cannot be directly employed. To address this analytical gap, I created an R package, Coereba, which enables a semi-supervised analysis thoroughly profiling the high-dimensional space, permitting simultaneous characterization of the rare cell subsets of interest, and all cell populations present in each sample. To this end, additional challenges beyond those associated with the unmixing process detailed extensively in Chapter 3 needed to be addressed. Given the novelty of our approach and complexity of the topic, the rationale is provided to contextualize the steps implemented by our algorithm.

Following unmixing, data corresponding to the acquired cells is stored as a Flow Cytometry Standard (FCS) file (81, 82). Within these .fcs files, structured data from each cell is stored within a table as a row, with each column corresponding to the calculated unmixed abundances for each fluorophore (82, 83). Traditionally, supervised (manual) analysis is utilized to explore flow cytometry data, typically mediated through commercial software (commonly FlowJo or FCSEXPRESS) by use of a graphical user interface (GUI).

Once a cell population of interest is identified, a polygon gate is drawn, and cells falling within the gate are selected (filtered) for analysis and can subsequently be gated for additional marker combinations. The end result of the supervised analysis is a series of hierarchical gates that focus from all events down to specific cell populations of interest (55, 84).

SFC can profile more markers than CFC and acquire far more events per minute than MC, thus generating larger .fcs files. From a technical perspective, despite the increase in file size, the supervised analysis of individual SFC .fcs files does not change substantially (52, 55). However, when considering larger data sets, with multiple conditions, individuals, timepoints and experiments, available rapid access memory (RAM) can quickly be exhausted. A common workaround for these size-related problems is to export cells within a specific node to a new .fcs file containing only the population of interest. This permits supervised analysis of that cell population, but restricts the ability to simultaneously characterize other cell subsets (53, 71).

In SFC, a far more substantial barrier for the traditional supervised analytical approach is the impracticality of exploring a high-dimensional space defined by the many markers in a panel with iterations of bidimensional plots (84-86). Given a PBMC sample, when no fluorophores are added, the only distinguishing features on a flow cytometer are morphological parameters, namely size (Forward Scatter, FSC), and granularity (Side Scatter, SSC). While we can identify specific lymphocyte and myeloid subsets within the low-dimensional space defined by FSC and SSC, we are unable to identify additional cell populations that may be present. The two identified cell populations consequently appear as homogenous groups (clusters), and any potential heterogeneity is not exposed.

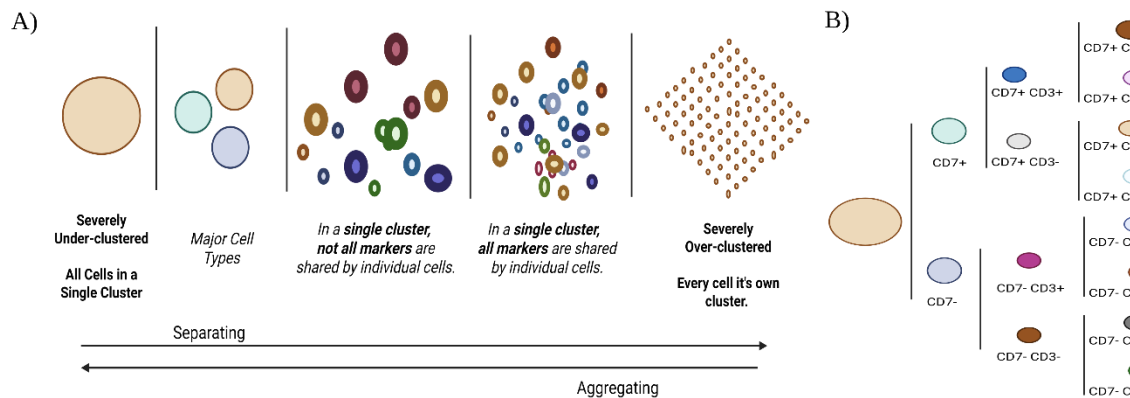


Figure 5: Defining a Cluster Depicted is A) Graphical depiction of how a cell cluster can be defined. B) Graphical illustration of Coereba ClusterID concept.

As fluorophore-conjugated mAbs are added to “stain” the cells, additional heterogeneity in the perceived uniform cell population is revealed as the dimensional space grows, revealing the major immunological cell populations (B cells, T cells, NK cells, etc.) (Figure 5A). Incorporating additional markers may define differentiation and activation states. These cell populations, as defined by the markers within a given panel, are targets of analysis to determine differences in frequency (abundance) or marker expression.

While technological limit or cost may dictate the panel size, and therefore the limit of number of clusters identifiable, the cell subset heterogeneity goes beyond these limits, as incorporating additional markers could further subdivide clusters, creating numerous smaller subsets. The theoretical limit of subdivision is reached when there are as many clusters as cells in the samples, with each cluster containing a single cell (Figure 5A).

Before this point, there exists a stage where all cells in a cluster share marker expression with each other (Figure 5B). If every marker defines cells that are positive or negative, we can estimate the theoretical number of clusters for a given panel as 2^n , with n being the number of markers/parameters present.

For a CFC panel with 9 markers, the theoretical upper number of clusters is 512. Utilizing traditional supervised analysis, also known as manual gating, exploring this dimensional space while time consuming is achievable. By contrast, the theoretical upper number of clusters for SFC (and also MC) panels with 40 or more markers approaches 500 billion. The sheer number of potential subpopulations spread out over the high-dimensional space defined by the markers vastly exceeds the ability to manually gate within an individual file (55, 84, 86). Consequently, if the goal is to characterize every cell population within the sample, SFC (and MC) require leveraging methods beyond manual gating.

While the number of markers may define the expanse of high-dimensional space, they do not inform us how clusters within this space are distributed, which can impact the effectiveness of various unsupervised analytical methods. If cells were to split evenly into clusters for every additional marker included, these cell clusters would be uniformly distributed across the high-dimensional space. However, a 3×10^6 CBMC sample stained with a 40-marker panel would end up with terminal clusters containing only single cells after grouping on the basis of only 12 markers (occupancy threshold). No further clusters would be defined by the remaining 28 markers. The cluster distribution would therefore be theoretically constrained, and the distance between clusters would be mainly dictated by the choice of first selected markers.

The theoretical occupancy threshold does not actually exist in biological reality, as cell clusters do not split on an even distribution as additional markers are added. Many cell markers are co-expressed on the basis of cell type, with networks of shared markers being what helps define a cell population (87). Within these co-segregating clusters, redundancy is present, so clustering on an additional marker that is already uniformly expressed by the clustered cells does not result in another division (69). Similarly, cell density is not uniformly distributed across clusters. Cells divide, giving rise to daughter cells (clones) that mostly share marker expression of the original cell that underwent expansion.

Because of these factors, while the maximum dimensionality of the analytical space is defined by the panel markers, this space is sparsely occupied (84-86). Cell density is distributed unevenly within particular clusters, based on the starting abundance and heterogeneity of the major cell populations in relation to the markers present in the panel. The distance between individual clusters in turn is dictated by differences in marker expression.

While different cell lineages will cluster far apart, clusters of cells with similar marker expressions will be in the same vicinity of high-dimensional space. These separate clusters often arise as the immune system responds to infection, with cells altering their expression of markers resulting in a new cluster that is separate from the original. Similarly, variation from the parent cluster can also arise from distortions introduced by experimental processing, or inaccuracy in the unmixing matrix, when a marker expression ends up misclassified.

Examining clusters at this local level illustrates differences between MC and SFC data, due to the number of acquired cells. While the extent of high-dimensional space and the reality

of the underlying biological heterogeneity within do not change, acquiring 10000 cells, or 10 million cells will impact the resolution of data visualization. MC, with fewer input cells cannot capture the rarer heterogenous clusters, resulting in greater distance between the detectable clusters present, potentially explaining some of the disparate algorithmic results compared to SFC, where these local environs have more adjacent clusters.

Within these contours, unsupervised analytical algorithms attempt to parse the high-dimensional space and delineate separate clusters based on various metrics. The FlowSOM package is emblematic of this approach, and employs self-organizing maps for unsupervised clustering within R (88, 89). This clustering can also be leveraged by the PeacoQC package to detect QC issues, and CytoNorm to correct batch effects by normalization (59, 61, 62). Beyond FlowSOM, other unsupervised algorithms using different approaches exist, typically implemented as free-and-open-source software (FOSS) R packages (90, 91). Bioconductor serves as the main repository, enforcing common standards for documentation, infrastructure interoperability, and continuous testing to ensure the packages remain functional (92).

For major cell types, located far apart in high-dimensional space, in terms of marker expression, the existing unsupervised algorithms often faithfully replicate supervised gating results (55, 84, 93). However, when attempting to delineate cell populations with high degree of co-expression, algorithms often fail to meaningfully represent the underlying cell population, returning too few clusters (i.e. different cell subsets being aggregated together) or too many clusters (i.e. splitting cells into biologically irrelevant populations)(84, 94).

These outcomes can be attributed to metrics used by the algorithms being inaccurate when it comes to discerning separate clusters of cells with similar expression profiles (i.e. those in local proximity). Our ability to refine the metrics to account for this disparity and improve interpretation of the analogous unsupervised clusters is hampered by the fields inability to concisely answer a more fundamentally biological question. Namely, in these local high-dimensional environments, what defines a distinct cell population? (Figure 5A). Do we define two clusters when they differ by one marker? Multiple markers? Or is it marker dependent?

Existing unsupervised algorithms attempt to address this separation by defaulting to various metric cutoffs, including Jaccard distance in the Phenograph algorithm, or meta clustering in FlowSOM (89, 90). The potential reasons for this local mis-clustering, include the fact that not all cell markers display segregating behavior, with some markers being shared across cell types. Likewise, some markers do not clearly delineate a positive or a negative population but display gradient expression levels (40). Additionally, since the metrics rely on the underlying MFI value, they can encounter issues in SFC due to unmixing errors and unmixing-dependent spreading (UDS) which affect both the scale and the range of the signal and muddle the identification of positive and negative populations (43, 77). This results in susceptibility to batch effects, which unless corrected, can lead to technical noise being misinterpreted as biological data (52, 55, 93).

Given the difficulty in fine-tuning algorithm parameters to get unsupervised clusters to reflect the underlying biology, a proposed scalable alternative to a fully unsupervised analytical approach relies on automated gating methods to implement a minimally supervised gating intervention (63, 84, 95, 96). The methodology scales well, but the

maintenance of the FOSS version of the R packages has dropped off substantially after commercialization of the approach (Ozette Technologies, Seattle, WA). While large gating schemas are implementable, there is no efficient way to terminally gate to the point where every cluster only contains cells sharing marker expression.

Given the complexities in the two competing approaches, there is a need for a semi-supervised analytical approach capable of working with the complexity and idiosyncrasies of SFC data, merging the biological interpretability of automated gating, with the scalability of unsupervised clustering. This would enable profiling rare cell populations without discarding information about other cell populations in the sample.

To fill these unmet analytical needs, I developed a computational tool, Coereba, to use SFC data for the assessment of rare cell populations, tying into existing Bioconductor infrastructure, and allowing incorporation of metadata and gating information for statistical interrogation in downstream analysis.

Results

The first step in developing a computational tool to enable semi-supervised analysis was to develop a function in R to determine, for each marker whether the fluorescence signal was above or below a reference splitpoint. A splitpoint is defined by a fluorescence value that enables the classification of cells as negative (for values below the threshold) or positive (for values above the threshold) for a single marker/fluorophore. The dichotomous classifications for each marker was then stored in a single new column (CoerebaID). This identity column could then be used to cluster all cells in the sample on the basis of a shared definition, and enumerate the number of cells present within these clusters.

I next set out to enable individual split-point adjustments for different specimens, to better account for batch effects or differential unmixing accuracy. I modified the function to retrieve splitpoints based on specimen and marker identities from a .csv file. I then altered the existing Luciernaga R package plotting functions to incorporate a vertical line at the designated split point location to visualize whether the location was appropriate.

Next, being able to visualize all the original designated splitpoints for visual inspection, I implemented an openCyto automated gating method to calculate the split point for each marker (and specimen) by the minimum density method (63). I also integrated the ability to set gate ranges to correct the split-point position when the automated results were obviously inaccurate, which constrained the region where the minimum density algorithm could function.

Given the time-consuming nature of adjusting splitpoints for individual specimens and markers by editing a .csv file, I created a RShiny app to visualize the plots in an interactive manner. By adding JSON code, I could manually adjust the location of the splitpoint for individual markers (and specimen) by clicking a location on a plot to save the corresponding coordinate on the x axis as the corrected location of the splitpoint. These coordinates could then be merged with the existing .csv file to update the splitpoints being applied.

Having the ability to calculate, validate and adjust individualized splitpoints, I leveraged CytoML to import all manually-gated ILT cell nodes from FlowJo into R (58). Coereba labeling was then carried out for every cell in a given conditional aliquot and across all specimens. Additional metadata, including infant identification number, HIV-exposure

status, infant sex, and maternal viral load was appended for future use for statistical comparisons.

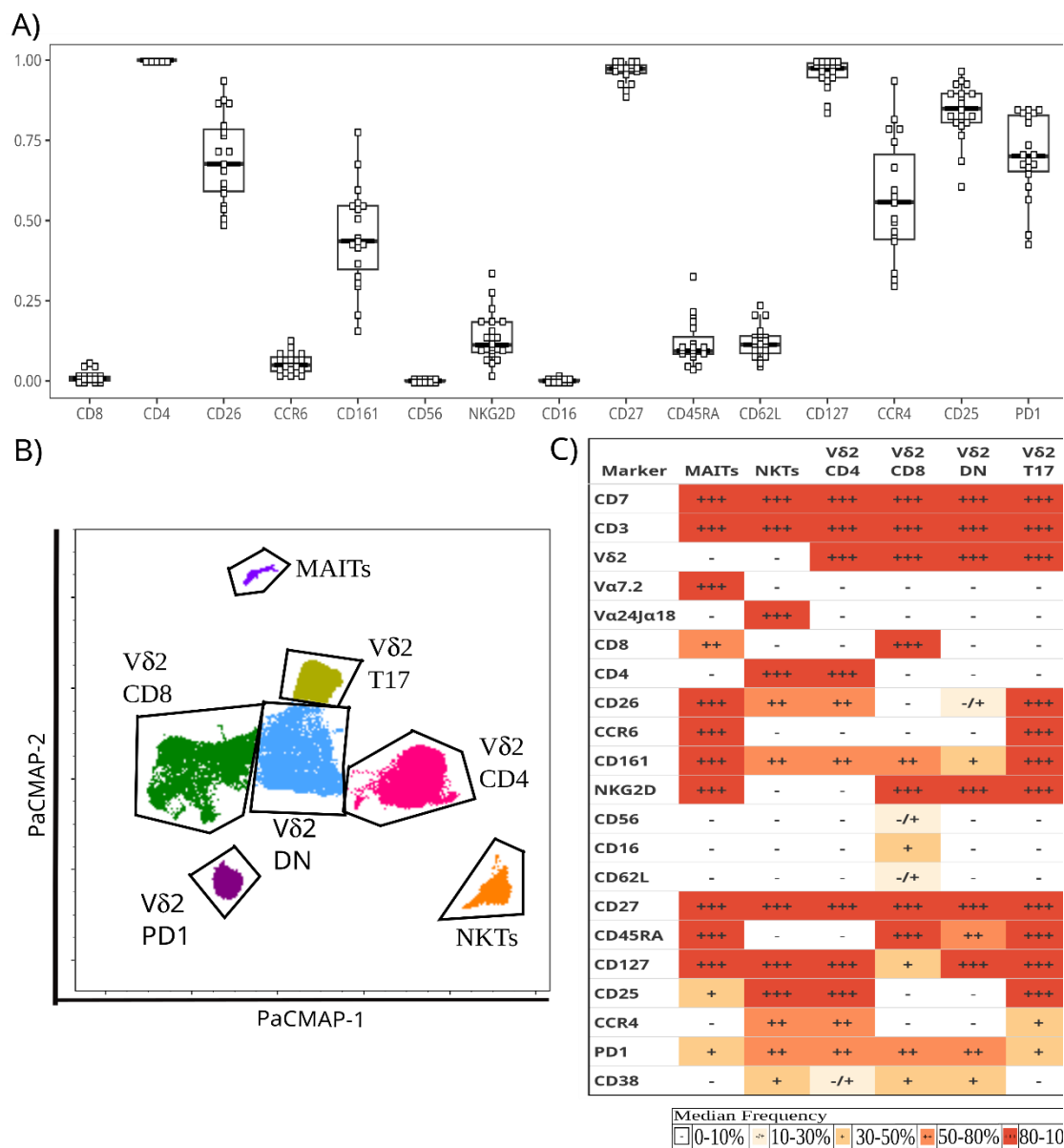


Figure 6: Coereba Visualizations. Example of outputs from a typical Coereba semi-supervised analysis, based on gates drawn on the dimensionality-visualized cells within a concatenated file. Depicted are A) Beeswarm boxplots showing marker expression for the CD4 Vδ2 cluster gated on B) a PaCMap. C) Summary heatmap of marker expression for baseline Cord Blood ILT populations gated in B.

The CoerebaID column, which contained summarized gating information of individual cells, allowed filtering on basis of expressed markers, and enabled the return of clusters in a manner analogous to manual gating. I could create templates to quickly isolate defined cell populations of interest for differential analysis. Due to the processing speed of the code, time from analysis to results was substantially reduced.

Similarly, all CoerebaID clusters containing cells positive for a given marker could be filtered for and aggregated during analysis. This permitted the rapid return of marker expressions profiles for all individuals for every designated cell population. Visualizing the results as beeswarm boxplots (Figure 6A) quickly provided a comprehensive summary of marker expression across the entire cell populations or within individual cell subsets.

While both prior implementations were effective, and efficiently replicated supervised analytical approaches, I also wanted leverage the “unknown population discovery potential” characteristic of unsupervised algorithmic methods. I created a method by which CoerebaID and metadata columns could be integrated into the keyword portion of an .fcs file (82), enabling a dictionary style ability to reference the original values for the equivalent numeric code within the .fcs files expression data. This allowed for .fcs files containing Coereba data to be passed off to unsupervised algorithms, permitting the retrieval of the Coereba gating information and metadata for the individual cells downstream of the respective analysis.

When cells were clustered using FlowSOM, Phenograph or X-Shift algorithms, underlying Coereba gating information for cells that ended up in individual unsupervised clusters

could be retrieved (89-91). This enabled us to compare the effectiveness by which each algorithm could distinguish cells with similar marker expression.

Similarly, cells passed through normalization algorithms like CyCombine and CytoNorm store only adjusted MFI values, but retain their original CoerebaID identities (60-62). Subsequently, by filtering for markers that were mismatched (started as positive, ended up as negative, etc.) we could evaluate whether rare cell populations with unique marker combinations (as well as cells with different unmixing-dependent spread (UDS) were being inaccurately “assigned” to more abundant counterparts by the individual normalization algorithm.

When passed to dimensionality reduction algorithms (64-67) (Supplementary Figure 3), we could verify whether cells that shared CoerebaID ended up clustering close to each other, or were scattered across the dimensionality reduction map. This allowed for the assessment of whether individual algorithms were appropriately segregating cells, in a biologically plausible manner, or were being driven by scaling differences in the underlying datasets. Additionally, by creating manual gates around cell clusters (i.e. islands) of interest (Figure 6B), CoerebaID and metadata of cells within these islands could be re-inputted into R, re-enumerated to create beeswarm boxplots specific to the individual clusters (Figure 6A). This allowed us to quickly assess and identify pre-existing heterogeneity in cell subsets originally perceived as homogeneous. These outputs when repeated for all dimensionality visualization clusters could be converted into heatmap tables, providing an immediate visual summary of differences in marker expression by cell subset (Figure 6C).

In addition to filtering for previously defined populations, I sought to characterize unknown cell populations on the basis of their shared CoerebaID. By enabling a filter option to identify clusters that were present only in isolated individuals or experiments, I could quickly return cluster IDs that matched these criteria. These would theoretically identify populations resulting from poor specimen quality, antibody panel errors, or potentially unique populations within individuals. An expansion of this approach enabled the retrieval of Coereba clusters that were only present for particular treatment conditions or timepoints. By plotting CoerebaIDs for particular cell subsets in the form of tree plots, we were able to visualize how these populations acquired marker expression during the early life period. To facilitate integration with existing Bioconductor R packages, I encapsulated the Coereba outputs data within a Summarized Experiment type object, enabling its future use with more sophisticated statistical analysis packages (92). For differential analysis within the beeswarm plot level, I wrote a consolidated function that would determine and carry out the appropriate statistical test, plot the results, and save the outputs as a pdf, enabling rapid screening. All these functions were gathered and incorporated into an R package, Coereba, which is available as a developmental release via GitHub.

Discussion

The field of SFC is rapidly progressing, allowing for the interrogation of increasingly more complex datasets. This is due to improved fluorophores, increased number of fluorophores within panels, and the ability to rapidly acquire large number of events (38, 47, 52, 55). However, the ability to identify and interpret statistically significant changes within these datasets is limited by the existing analytical tools. While many of these analytical tools are

adopted from other similar fields (such as single-cell RNA-seq and MC), they do not account for many of the unique features of SFC data.

Coereba is our attempt to combine the benefits of supervised manual gating with the scalability of unsupervised analytical workflows. While additional improvements such as improved speed and multi-core utilization (parallelization) would enhance performance, the package has shown significant promise in its current form. It has allowed us to visually summarize heterogeneity identified within high-dimensional datasets as figures containing statistical metrics. Similarly, the capacity to rapidly filter and interpret marker expression for populations of interest allows for a systematic approach to the analysis of flow cytometry data than what had been previously available through out-of-date FOSS packages.

Coereba allows us the opportunity to evaluate the performance of unsupervised tools that are industry standards but the weaknesses of which have not been systematically evaluated, including applications involved in clustering, normalization, and dimensionality visualization. Systematically testing the assumptions of these analytical tools will highlight areas needing improvement and allow us to rigorously question rather than accept an algorithmic output as fact.

Furthermore, for CBMC, we encountered a few thousand clusters that included substantial proportion of cells at the Coereba cluster level. This was both far fewer than the theoretical upper limit based on panel size and cell counts, and far more clusters than current algorithmic clustering and dimensionality reduction algorithms (like FlowSOM, Phenograph, etc.) would have suggested on basis of their metric cutoffs. Defining what

constitutes a cell population or a cluster remains a challenge that is, in part, dictated by the markers contained within a given panel, and our question of interest within the dataset.

The difficulties in statistical interpretation of the Coereba nodes highlight this challenge. When the expression of a single additional marker is sufficient to trigger a switch in cluster, metrics based on counts (and therefore ratios) can be noisy and produce outliers that can drive the inaccurate appearance of statistical significance when it is not present. Likewise, poor specimen quality, or unique individual immunological expansions to stimuli can affect the metrics by which we derive statistical significance. Linking Coereba to the Bioconductor infrastructure should allow for more sophisticated statistical analysis (including those using mixed-models) to be tested and implemented to solve this issue.

Another application of the Coereba package is as a screen for quality control. There is no effective strategy to quantify errors introduced by loss of cells in situations in which poor cryopreservation or errors in experimental processing led to cell death, which subsequently leads to errors in the interpretation of the data. Future work characterizing a greater number of specimens may inform which CoerebaIDs are expected for a given panel, permitting the flagging of experiments when the CoerebaID “fingerprint” of a specimen detects deviations, possible resulting from poor specimen quality.

In conclusion, Coereba provides the infrastructure needed to implement semi-supervised analysis of high-dimensional SFC data, approaching the analysis from a unique angle. This results in outputs that prove informative for many potential downstream applications that previously were difficult to assess or quantify. While answering what defines a “cell population” remains elusive, individual researchers’ ability to thoroughly profile their acquired .fcs files ensures a greater awareness of things that might have been otherwise

missed using either the fully supervised or unsupervised approaches that had been available at the start of this PhD journey.

Chapter 5: CytometryQC

Introduction

Many unsupervised analytical methods, including clustering, dimensionality visualization and normalization algorithms (55), rely on MFI to perform their tasks. As described in the previous chapters, issues related to unmixing controls can carry over into the calculation of fluorophore abundances (45, 75, 77). These in turn can result in altered MFI values, creating batch effects for a particular experiment that can lead to the unsupervised algorithms producing unexpected or inaccurate results (55, 60-62). Consequently, we prioritized reducing variance due to these external factors across the flow cytometry experiment series. With a well optimized panel and proper unmixing controls, differences in abundance should be reflective of the changes in the underlying biology, and not of bias introduced by other sources of variance (53, 71). Additionally, consistent normalization controls that are thawed and processed on each experiment run can help mitigate these issues by providing a known biological constant across experimental series, allowing normalization algorithms to calculate and apply corrections(60-62).

Unquantified variation in instrument performance has the potential to introduce significant artifacts into data analysis. Daily use of flow cytometers leads to wear and tear of lasers, detectors and fluidics, resulting in gradual drifts over time and leading to altered fluorescent signatures (47). In SFC, various approaches are implemented to mitigate these changes, including the use of quality control beads, which emit known fluorescent signatures. These beads are used to calibrate the instrumental gains daily, allowing the instrument to match lot-specific MFI target for each detector. Maintaining baseline standards long term is theoretically possible, allowing comparison between experiments

performed months apart and reducing the sources of batch effects that may confound an unsupervised analysis (55). In theory, reference control data can thus be saved to an instrumental reference library of Cytex Aurora instruments, which will account for changes in QC, allowing for their subsequent use in unmixing future experiments.

In practice, stored reference library controls often produce unmixing errors due to unknown reasons within a few weeks of acquisition. Once the unmixing errors are encountered, it is often difficult to determine which errant control is responsible, as there is no way to visualize the stored fluorescent signature. This limits our ability to explain why library controls start to fail when they initially performed well.

QC is normally performed only during the morning, and instruments are continuously operating throughout the day. It is unknown to what extent instrumental drift may occur during this time, and whether re-running QC during the evenings is necessary. If the evening fluorescent signatures vary substantially from the QC baseline, it could result in inaccurate unmixing. Additionally, as gains are adjusted to meet MFI thresholds based on QC bead samples, anything that damages the beads prior to running QC analytics could result in improperly set gains that could affect the final unmixed dataset.

Despite the realities of continuous cytometer usage, there are limited profiling tools to evaluate the QC measurements longitudinally for many spectral flow cytometry platforms, with vendors being slow to implement community-suggested feedback and best-practices(71, 75). Having the ability to profile both fluorescent bead signatures and other quality control metrics (Gain, MFI and robust coefficient variation variance (RCV) for each detector) would permit the early identification of a range of instrumental issues (such as laser failure, misalignment, fluidics problems, etc.) allowing pre-emptive action to be

taken before the bias is introduced into data interpretation. In order to reduce batch effect errors, I sought to develop software tools to visualize QC data, detect changes in QC bead signals, and estimate how these sources of variation might affect the unmixing within our datasets.

Results

I modified two existing Luciernaga functions: one that enabled the extraction of the stored gain/voltage settings from within individual fcs files, and another that enabled the automated gating of a target bead reference control population.

After combining the two Luciernaga functions, I incorporated the ability to both derive the median raw fluorescent signature across detectors for the gated beads (Figure 7A), as well as quantify their spread as %RCV (Figure 7B,C). I could then iterate over a collection of .fcs files within a folder, to create a data-frame object containing the recent QC data for a given instrument. I then modified existing plotting functionalities within Luciernaga to enable graphing these calculated values linearly, creating Levey-Jennings tracking plots. Levey-Jennings plots offer a visual means of identifying deviations, shifts and trends in analytical performance during data acquisition.

By leveraging the plotly R package, I converted the plot into interactive data visualizations. I then proceeded to build a website using the open-source, next generation version of R markdown, Quarto, to create a dashboard layout that could simultaneously visualize interactive plots for all instrument detectors organized by laser.

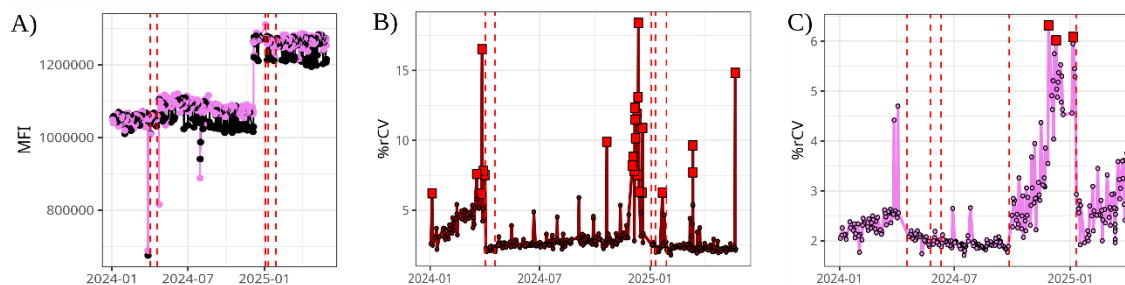


Figure 7: Levey-Jennings Plots. Graphical depictions of analytical performance are shown as follows: A) Example of maintenance of bead lot specific target MFI on a 3-laser Cyttek Aurora spectral flow cytometer, before morning QC measurements in black, after QC measurements in pink, with time (date of measurement) on the x-axis, and MFI on the y-axis; and B) An example of bleach induced spikes in %rCV for the last detector on the red laser; and C) Example of a failing laser on a 3-laser Cyttek Aurora spectral flow cytometer, with time (date of measurement) on the x-axis, and gain on the y-axis. The red-dashed vertical lines depict field service engineer visits for maintenance.

I next identified previously acquired QC .fcs files and processed all available files which for some instruments represented over 7 years of data. Due to the size of the datasets, I subdivided data analysis into individual years, enabling a rapid download of the underlying data into .csv files and visualizing plots as pdf files for reporting purposes.

After optimizing the technique, I created analytic scripts for all the cytometers within the core facility including Cytex, BD and Amnis instruments. While processing individual .fcs files was the preferred route, I also created programs to parse the required data from existent .pdf and .xml files. The extracted data was then visualized and incorporated as additional instrument-specific webpages for dashboard.

Once the previously acquired data was analyzed and organized, I implemented an automation pipeline to daily collect and process new QC data from each instrument. The Microsoft Windows Task Scheduler was used to schedule morning processing script runs, with newly acquired data being processed and submitted in a coordinated fashion to GitHub, enabling preservation of backup files to the cloud. I then wrote a GitHub actions script that when triggered, enabled GitHub to provision a virtual machine to carry out the rendering and publishing of the updated data to the website, avoiding the use of in-house computing (Figure 8A).

Leveraging retrospective data, I sought to estimate instrumental divergence from the previous day's QC measurements to the following day before QC measurement. Substantial changes from baseline were observed 66% of the time. Simulating the effect of these changes on reference signatures showed that instrumental fluctuations have the potential to substantially differ from their original signatures. When these distorted signatures were utilized to unmix samples acquired the day before, unmixing errors were

encountered, due to signature mismatch with full-stained samples. By contrast, drifts between morning and evening QC were minor, which when simulated against existing reference controls did not significantly alter signatures or result in unmixing errors.

While these signature simulation experiments were ongoing, the core cytometry facility switched QC bead lots, and major unmixing errors were encountered. From the dashboard, the appearance of spikes in the %rCV was observed for the last detector of each laser. By repeated comparison of signatures before and after QC, we observed that an APC-Fire 810-like molecule was being stripped from the beads. When QC was run using the damaged QC beads, the process resulted in the gains being dramatically increased for the detectors (UV16, V16, B14, YG10, R8) in an attempt to match bead target MFI. The overcorrection due to the damaged beads resulted in grotesquely distorted signatures. To validate whether the unmixing errors were due to the use of damaged beads, we simulated changes observed for the reference library signatures, and iteratively unmixed samples that had unmixed well the day before, with the effect phenocopying the original unmixing results.

Finally, it was of interest to determine whether this phenomenon was an anomaly or part of a recurring trend. We leveraged the retrospective datasets for the previous seven years and queried events where similar changes in QC signatures occurred. We found evidence of damaged beads used during QC for previous lots, but occurring on different detectors, although to a lesser extent and less frequency than what was observed with the 2006 bead lot.

As interest was voiced by the broader cytometry community in being able to build their own similar dashboards and given my acquisition of significant knowledge of navigating Quarto, GitHub and R, in addition to documentation for individuals with existing coding

skills, I set out to write an R package, CytometryQC, that would programmatically write the code needed to build individual flow cytometry core websites for those with no-to-little coding experience. Preliminary testing shows promise, and the package enabled significantly fewer edits to existing files. The program was also broadened to include various platforms from different manufacturers, including Agilent, BD and ThermoFisher.

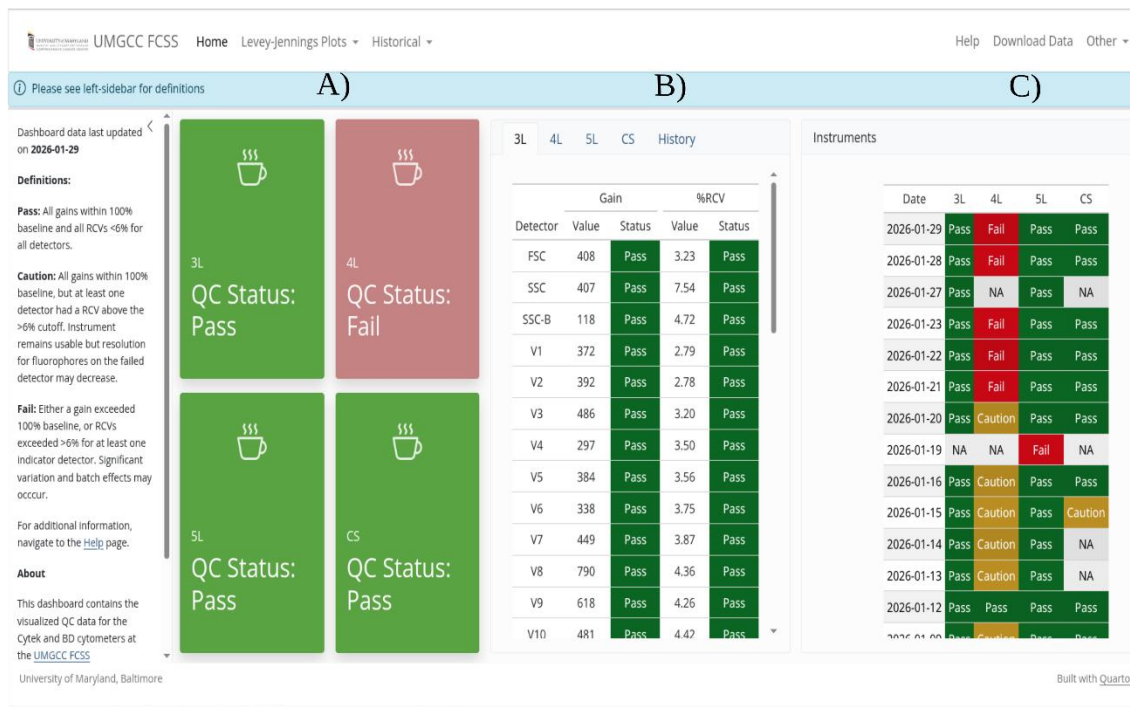


Figure 8: CytometryQC-generated dashboard landing page. Depicted is an example of the landing-page for a CytometryQC generated instrument QC dashboard. Legend and how-to-interpret description are located on the far-left sidebar. Individual instrument statuses for a particular day are to the right, being color-coded as green = pass, red = fail. To the right, the daily QC results for individual detectors are displayed, with each instrument occupying a tab slot. The far-right side bar displays a summary of instrument status for individual days on the y axis.

Discussion

Instrument quality control usually relies on manufacturer technical support, with limited contribution by individual shared resource laboratories (SRL). With the implementation of CytometryQC, we enable both users and SRL staff to explore quality control data and leverage it. As instrument users, this matters as it can help reduce variance that may otherwise affect our experiments, preventing batch effects rather than mitigating them after they occur.

The implementation of an automated-tracking dashboard has allowed the evaluation of Aurora QC functions, such as the assessment that designated target values are being maintained over longer periods of time. The longitudinal tracking of gain has enabled the early detection of failing lasers and detector arrays and enabled the scheduling of instrument repair visits by field service engineers well in advance of when manufacturer QC would have first detected an issue. Ability to intervene before issues arise that result in significant instrumental downtime is incredibly valuable, since staff time and ability to troubleshoot is limited, and longer an instrument is down for maintenance more potential revenue is lost.

The monitoring of % rCV permitted the detection of debris accumulation within the fluidic lines (allowing for enhanced cleaning protocols) and enabled the quick identification of the issue caused by switching to a new QC bead lot. Specifically, the signal spread (as identified by % rCV) for the last detector of each laser suddenly spiked. This issue was entirely missed by manufacturer QC, resulted in significant experimental unmixing, and would have led to a substantial number of user experiments being impacted, if not for the early intervention.

Our ability to visually track analytic performance permitted the retrospective analysis of QC bead signatures, and enabled detection of significant numbers of “damaged bead” days. Although these “bad bead” days were less frequent than what we encountered with the 2006 bead lot, our signature simulation and iterative unmixing approach showed that these likely impacted the unmixing performance for those experiments analyzed during this period, in a manner similar to what we saw for our own failed experiment. As these “bad bead” days also did not trigger the QC fail criteria of the instrument manufacturer, the missed-set QC targets would have resulted in altered fluorophore signatures and unexplained unmixing errors for experiments acquired on those days. In cases where impacted markers were critical, these experiments would need to be repeated, expanding time, cost, reagents, etc.

While similar approaches showed the necessity of running QC daily on spectral flow cytometers, only minor detector drifts occurred within a single day of continuous use. For the majority of fluorophores, there was no appreciable change in signature at the evening QC. In our testing, even noticeable changes did not impact a 29-color panel. Whether this would be the case for significantly larger panels remains an open question. We found reasons to justify daily QC on our spectral cytometers, and satisfactorily demonstrated that daily QC drift is limited, at least in context of our 29-color SFC panel

We demonstrated the benefit of providing ready access to instrumental QC data, and plan to distribute this “best practice” procedure among the broader cytometry community by facilitating similar implementation of open-source tools. If these datasets, currently hidden away in stored folders, are made available, the community will benefit as thousands of datasets become available, allowing for further specialization of our detection tools for a

wider array of problems than the ones we have encountered. While implementation of a multi-instrument dashboard would still require Shared Resource Laboratory agreement, advantages are sufficient to drive community uptake, as barrier entry is reduced with future improvements.

Chapter 6: Cord Blood Innate-Like T cells

Introduction

Due to previously discussed technical limitations, little is known about cord blood ILT. Previous studies have reported they are present at low frequencies (6, 7, 50). Data has demonstrated that upon stimulation, both NKT and V δ 2s produced Th1-like cytokines, and both were capable of producing IL-17 under certain conditions (e.g., in the presence of IL-23) (6, 7, 50). For the V δ 2s, this was a particular feature of neonatal cells, which in addition to IL-17 also produced perforin and granulysin, suggesting important mechanisms of immune defense during the neonatal period (97). After birth, in response to microbiota, both V δ 2 and MAITs undergo expansion, and acquire maturation markers (4, 8, 14, 98). By contrast, NKTs are already differentiated into memory at birth, and appear to remain at a stable frequency over the first few months of life. Given their presence and capacity to produce cytokine after activation, as well as the knowledge that ILT in adults secrete cytokine and cytotoxic granules in response to infection and inflammation, cord blood ILT may play similarly important roles during the neonatal period.

With technological improvements that have occurred in recent years, a detailed characterization these rare cell populations from limited clinical biospecimen is now attainable. The capacity to acquire and analyze greater numbers of cells (events) using SFC may permit more extensive examination of heterogeneity within these cell subsets, and their ability to respond to antigenic stimulation. It is unclear whether the cells that expand during early life in response to microbial exposure are already present during the neonatal period or emerge separately. Whether prenatal exposure to inflammation skews the ratio of non-responsive vs. responsive precursor cells is unknown but could explain differences

in susceptibility to infection due to disparate capacity to mount functional responses during early life. A deep characterization of ILT in the neonatal period will allow us to better understand the role that these cells play in both healthy individuals, as well as those with chronic infection.

Given the increased morbidity in both lower respiratory tract infections and diarrheal diseases seen in HEU infants during the first six months of life, many have hypothesized that a contribution of altered immune responses at birth compared to HU infants living in the same communities may be causative, even after accounting for socio-environmental factors. Since the same disease outcomes can result from either immune hypo-responsiveness (where pathogen mediates cellular damage) or hyper responsiveness (where immune cell activation results in collateral cellular damage), understanding the parameters of a healthy immune system can help inform the perturbations of a compromised immune response and the choice of clinical intervention.

With the placental environment of women with HIV presenting inflammatory features (including the production of inflammatory cytokines $\text{TNF}\alpha$, $\text{IL-1}\beta$, IL-6) (32, 33, 35), there might be cytokine differences in cord blood ILT in HEU infants, given that both viral proteins and ART can cross the placenta. Whether through direct effect or resulting from innate immune activation, this can result in early activation of the neonatal immune system. Whether these early activated cells persist as memory differentiated cells, or undergo activation-induced cell death is unclear, but would ultimately result in an atypical immune response compared to healthy infants without HIV exposure. Infants born to HIV-infected mothers on ART with suppressed viral loads throughout pregnancy, may have reduced immunologic perturbation (24).

In order to characterize subtle differences in ILT responses arising from a gradient of HIV exposure (as opposed to congenital infection) there is a need to broadly characterize the cord blood of neonates born both to healthy women, and those living with HIV. To our knowledge, no prior study has characterized cord blood ILT populations in an in-depth fashion, and only two prior studies examined ILT in HEU infants at six months and two years of age (21, 25).

Our access to a unique clinical East African mother-fetal cohort, allowed for leveraging our expertise in SFC, and we designed a 29-color SFC panel to characterize cord blood ILT and other T cell subsets. We analyzed over 60 cord blood specimens and assessed the immune heterogeneity present using both supervised and semi-supervised analytical approaches.

Results

ILT frequencies were determined by applying a manual gating strategy to a detailed 29-color SFC panel (Figure 9A, Supplementary Figure 1). In neonates born to healthy mothers, V δ 2s were the most abundant ILT subset in T cells (0.693% [0.421, 1.30]) (Figure 9A), detectable in all CBMC specimens analyzed. MAITs were present at lower frequencies than V δ 2 cells. Of note, the proportion of MAITs identified by tetramer staining (CD3+ CD161hi hMR1 5-OP-RU+ cells, 0.022% [0.017, 0.027]) was lower (p-value: 1×10^{-6}) than the proportion identified by the antibody specific for the semi-invariant V α 7.2 TCR chain (CD3+ CD161hi V α 7.2+ cells, 0.078% [0.056, 0.108]). NKTs were the least abundant of the three ILT populations (0.053% [0.044, 0.073]) (Figure 9A, B). The proportion of NKTs stained with tetramers (CD3+ hCD1d PBS-57+ cells) was not significantly different from

the proportion identified by an antibody specific for the semi-invariant V α 24J α 18 (CD3+ V α 24J α 18+ cells; 0.029% [0.023, 0.057]).

While the V δ 2 frequency appeared comparable between HU and HEU neonates by SFC, an elevated V δ 2 cell frequency was observed in HEU-hi infants by CFC analysis, which included a distinct subset of specimens than the SFC analysis (Figure 9B). The frequency of MAITs by tetramer staining appeared similar across exposure groups, but it tended to be lower in HEU infants compared to HU infants with anti-TCR V α 7.2, staining and reached a statistically significant difference when HEU-lo and HEU-hi specimens were pooled. A similar trend was also noted for NKTs, which were also significantly lower in HEU after pooling the HEU-lo and HEU-hi groups, but only when staining with CD1d tetramers.

To evaluate cell differentiation, we measured the expression of CD45RA and CD27 in ILT subsets (Figure 9A, Supplementary Figure 3A). CD62L was expressed at low frequencies in ILT, in contrast to adaptive T cell subsets, minimizing its utility for the assessment of differentiation. Most V δ 2 cells were naive (CD45RA+ CD27+, 54.0% [51.0, 70.2]), with a sizable subset of central memory cells (T_{cm}) (CD45RA- CD27+, 25.1% [16.3, 38.4]) in most individuals (Supplementary Figure 3A). Cord blood MAITs were almost all naive (CD27+ CD45RA+, 87.5% [81.8, 92.5]), with only a few neonates displaying any sizable, differentiated subset. Results were comparable for cells identified by either anti V α 7.2 or tetramer staining (Supplementary Figure 3B). Conversely, NKTs were predominantly T_{cm} (CD45RA- CD27+, 67.4% [60.3, 77.0]), followed by naive, Tem, and Temra (CD45RA+CD27-) cells. NKTs identified by either anti-V α 24J α 18+ or tetramers had comparable memory phenotypes (Supplementary Figure 9C). We did not observe any

statistically significant difference between the HIV/ART exposure groups' memory differentiation phenotype for any ILT population, by either antibody or tetramer staining.

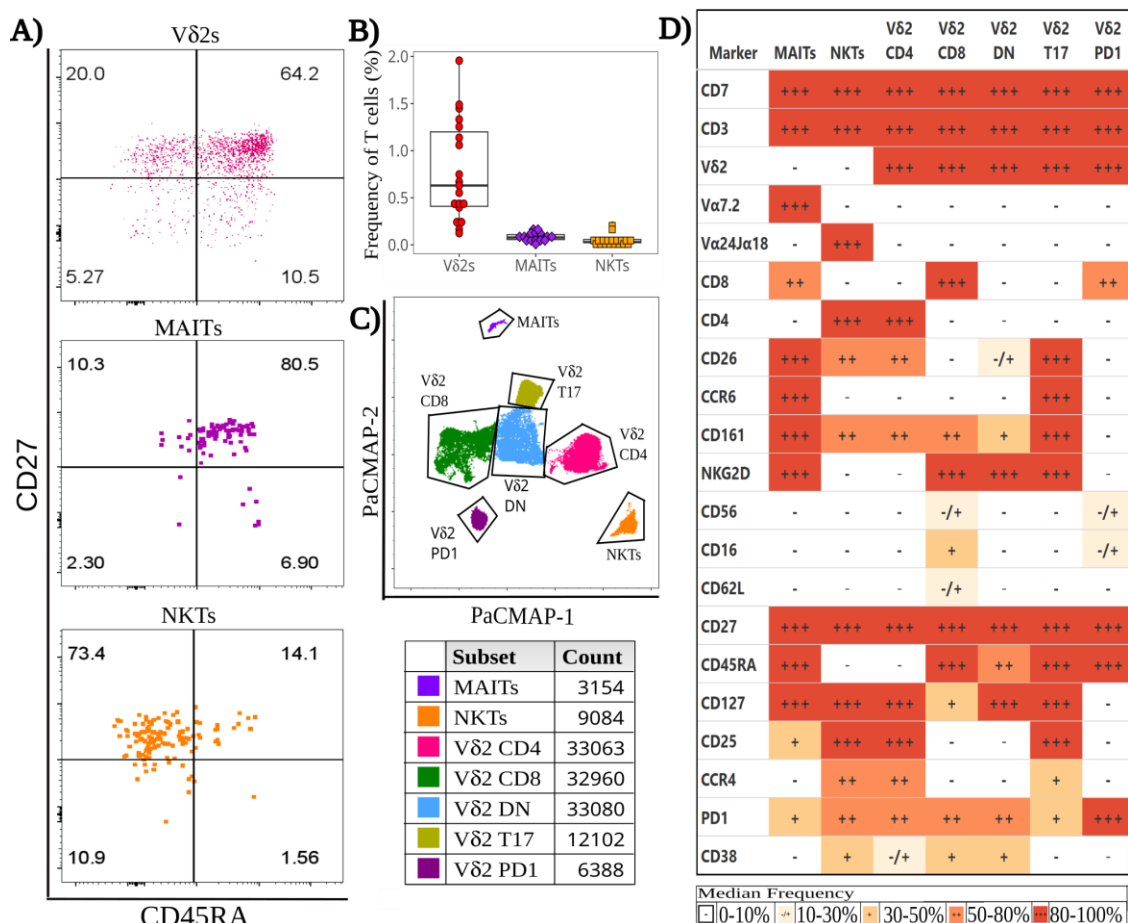


Figure 9: Baseline heterogeneity of cord blood Innate-like T (ILT) cells. CBMC were stained with a panel of 29 mAbs after a 6-hour incubation at 37°C in the presence of Golgi inhibitors. (A) Dotplots of a representative specimen obtained from an unexposed (HU) infant, show the expression of CD45RA and CD27 for Vδ2s (top), MAITs (middle), and NKTs (bottom). (B) The beeswarm plot displays the frequency of ILT subsets, with individual symbols representing unique HIV unexposed (HU) study participants, boxplots depicting median and IQR, whiskers showing the +/- 1.5 IQR range. (C) A PaCMAP dimensionality reduction plot is used to visualize ILT subsets in a concatenated file, including NKT and MAITs as identified by tetramer staining; the gates are based on density-distribution and marker expression, with color code and cell

counts included in the legend. (D) The heatmap summarizes marker expression for all ILT clusters delineated in (C) for all study participants, using a combination of colors and symbols to display, for each marker, the median of individual frequencies of positive cells.

In order to rigorously compare HEU and HU infants, we first employed a semi-supervised analysis to assess the heterogeneity of ILT subsets during homeostasis (Figure 9C). Comparing multiple dimensionality reduction and visualization methods (Supplemental Figure 4), individual algorithms differed in the extent of splitting or coalescing, but all computational models grouped MAITs and NKTs in a single cluster per subset (Supplementary Figure 4). Vδ2s clustered by CD4, CD8 and CCR6 expression, which appears to align with prior studies on Vδ2 cell function. Comparing the results of the PaCMAP and the heatmap (Figure 9C,D), it appears that markers driving cluster formation in the PaCMAP were those expressed at the highest (+++, red) or lowest (-, white) levels by cells within the cluster, aligning with marker co-expression patterns (Figure 9D).

Within the PaCMAP (Figure 9C), cord blood MAITs were primarily CD8+, and most expressed CD161, CD26, CCR6, NKG2D, and CD127 (Figure 9D), while the median proportion of MAITs expressing CD25 or PD1 fell in the 20%-50% range. Of note, <20% of MAITs expressed CD56, CD16, CD62L, CD4 and CCR4 (Figure 9D). NKTs were primarily CD4+ and uniform in their expression of CD127, with most cells expressing CD25, CCR4 and PD1 (Figure 9D). The median frequency of expression for CD26 and CD161 was 40%-60%, while CD38 was expressed by 30%-50% of the cells, with the remaining markers expressed by <20% of the cells.

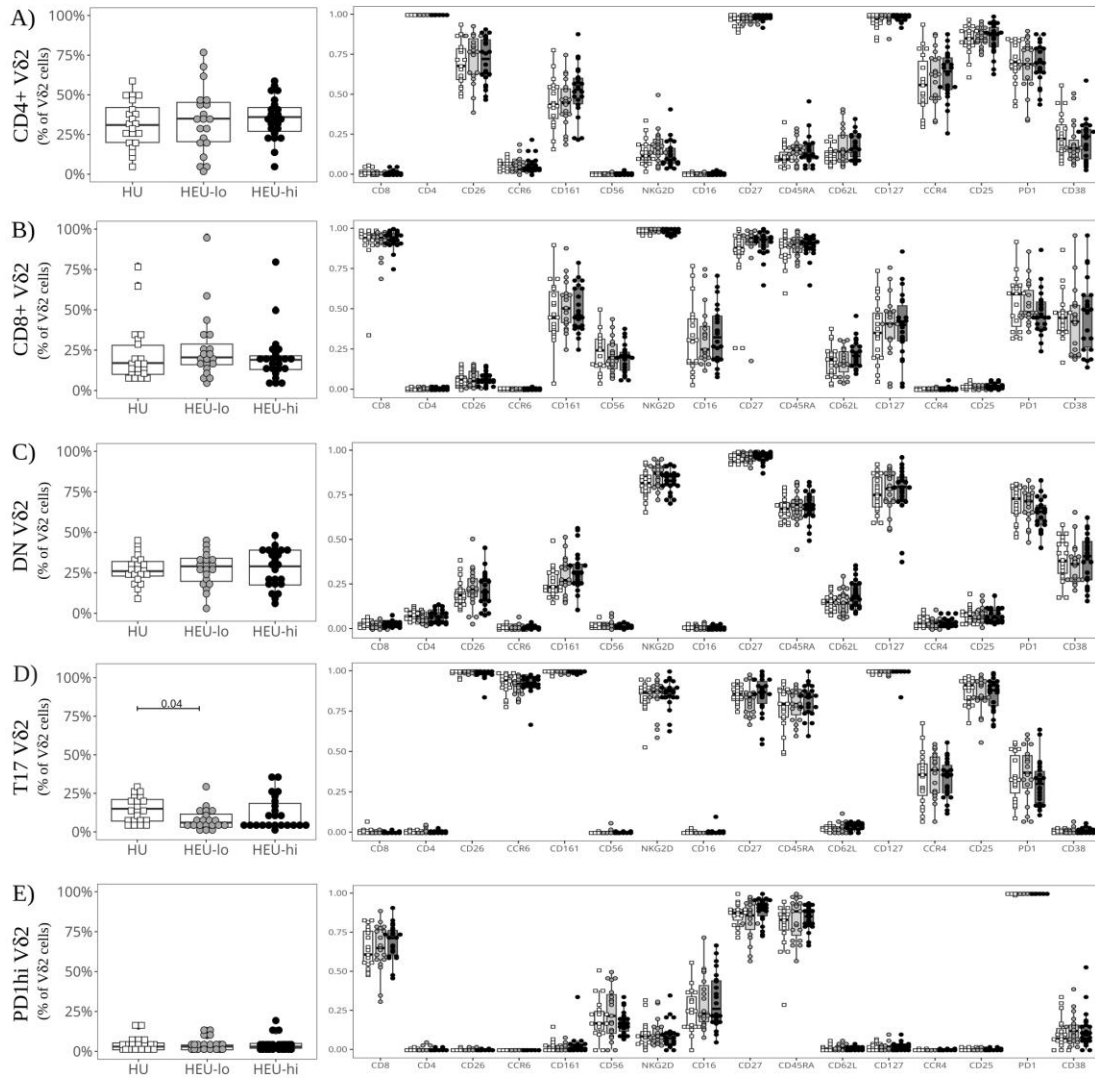


Figure 10: Semi-supervised analysis highlights cord blood Vδ2 cell phenotypic heterogeneity between clusters. The beeswarm plots on the left show the proportion of cells in each Vδ2 PaCMAP cluster (identified in Figure 9C) comparing the three exposure groups. The beeswarm plots on the right display the proportion of Vδ2 cells in the respective cluster expressing the individual markers listed on the x-axis. The symbols show individual values, boxplots show median and IQR, with whiskers indicating the

+/- 1.5 IQR range. Shape and color identify HIV-exposure status. (A) depicts results for the CD4+ PaCMAP cluster, (B) for CD8+, (C) for DN, (D) for T17, and (E) for PD1hi.

Semi-supervised analysis revealed that Vδ2 cells, the most abundant ILT population, were distributed into five clusters (Figure 9D, Figure 10A-E). The three most abundant subsets were primarily defined by the presence or absence of CD4 and CD8, the fourth by the expression of CCR6, and the smallest by high levels of PD1. The CD4⁺ Vδ2 cluster (35.0% [23.5, 45.0] of total Vδ2s) (Figures 9D, 10A), similar in marker expression to NKTs, was predominantly Tcm, with uniform expression of CD127 and most cells positive for CD25. Most cells also expressed CD26, CCR4 (a marker of Th2 polarization) and PD1 (likely akin to a differentiation marker in Vδ2 cells) (40, 69, 70). Approximately 50% of CD4⁺ Vδ2 cells displayed CD161. NKG2D and CD38 were limited to <20% of the cells, while markers of cytotoxic potential (CD56 and CD16) were minimally expressed. The CD8⁺ Vδ2 cluster (18.5% [12.7, 25.0]) was primarily naïve with uniform expression of NKG2D and intermediate expression of CD161, CD38 and PD1 (Figures 9D, 10B). Minimal to no expression of CD26, CCR6, or CCR4 was noted, suggestive of a Th1-like polarization. A cluster of double-negative (DN; CD4⁻CD8⁻, 28.0% [19.7, 35.0]) Vδ2 cells were mostly naïve, with Tcm cells being the second most represented differentiation subset (29.4%). The majority expressed NKG2D, CD127, and PD1, with intermediate expression of CD26, CD161, CD38 (Figures 9D, 10C). A Vδ2 cluster segregated from the rest of the DN cells, due to its high expression of CCR6, CD26 and CD161, was suggestive of Th17 polarization (70). Cells in this cluster, identified as T17 (9.0% [4.0, 16.5]), were CD4⁻, CD8⁻, CD127⁺ and, despite being naïve, displayed NKG2D and CD25 on most cells (Figures 9D, 10D). Less than 50% of the cells expressed PD1. The last Vδ2 cluster was a small subset (present in >85% of individuals, 3.0% [1.5, 5.0]) of mostly naïve CD8⁺ cells characterized by high PD1 and minimal NKG2D expression (similar to the CD4⁺ cluster) (Figures 9D, 10E). The

CD8⁺ V δ 2 and PD1^{hi} V δ 2 clusters were the only two that included a sizable proportion of CD56⁺ and CD16⁺ cells (20.1% and 30.2% for the CD8⁺ and 17.1% and 23.8% for the PD1^{hi} cluster), suggestive of cytotoxic potential, albeit modest. When we compared the exposure groups by using metadata to stratify events within clusters, we observed small differences (Figure 11). HEU-hi neonates, compared to the other two groups, showed significantly higher expression of CD62L in the DN and in the T17 V δ 2 cluster (Figure 11B), as well as higher CD161 and lower CD56 in the DN cluster (Figure 11A, B). We additionally observed a significantly lower frequency of CD161⁺ NKT cells in HEU-lo infants compared to the other two groups (Figure 11c).

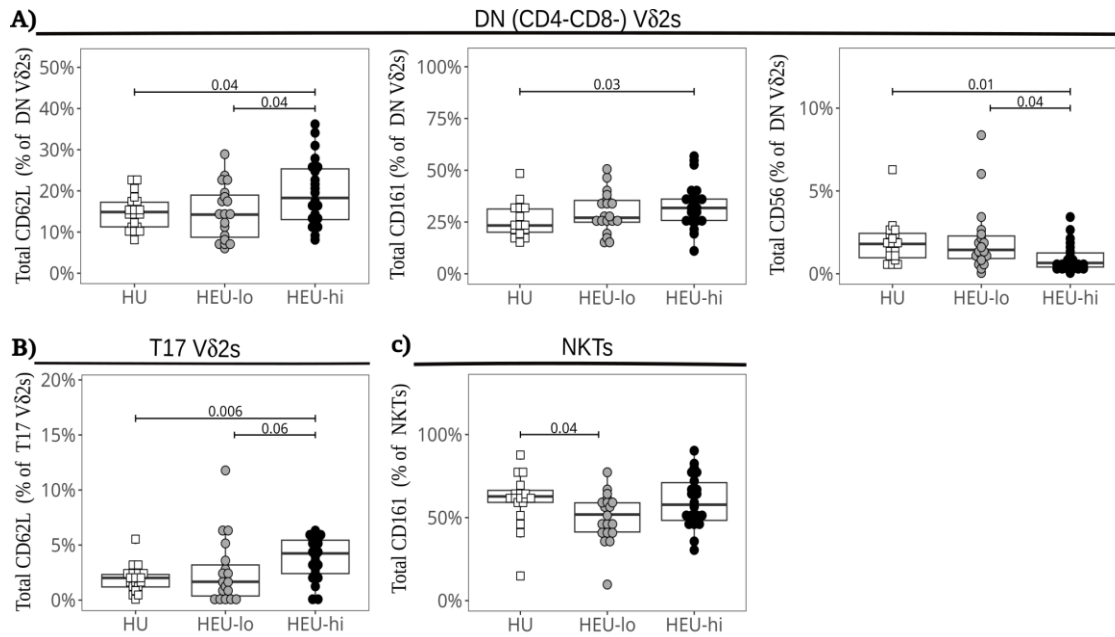


Figure 11: Altered marker expression of cord blood ILT subsets in HEU infants. The beeswarm plots compare the expression of specific markers between exposure groups. Symbols show individual values, boxplots show median and IQR, with whiskers indicating the ± 1.5 IQR range. Shape and color identify HIV-exposure status. (A) Proportions of CD62L⁺ (left), CD161⁺ (center), and CD56⁺ (right) cells in the DN Vδ2 cluster. (B) Proportions of CD62L⁺ cells in the Vδ2 T17 cluster. (C) Proportions of CD161⁺ cells in the NKT PaCMAP cluster

Discussion

This study demonstrates that use of SFC for specimens with low cell numbers, such as human pediatric specimens, allows in-depth examination of low frequency subsets, including ILT, that neither mass cytometry nor single cell RNA seq approaches could analyze in a cost-effective manner without prior enrichment. Consistent with our previous findings, we observed extensive heterogeneity in V δ 2 cells, in contrast to the more homogeneous features of NKT and MAIT cells.

We characterized the phenotype and functional profiles of cord blood ILT at baseline and following polyclonal stimulation, using both supervised and semi-supervised analytical approaches. We observed extensive heterogeneity in V δ 2 cells based on the expression of surface markers, consistent with prior single-cell RNA seq data published for a small number of samples (99). Notably, we observed memory differentiated CD4⁺ population, a T17-like CD4-CD8-CD26+CCR6⁺ population, as well as a rare PD1 high population. Some V δ 2 T cell subsets displayed marker expression profiles reminiscent of cord blood NKT and MAIT cells, which may be indicative of similar roles or cytokine responses.

While both NKT and MAIT cells appeared more homogeneous, the low abundance of these two subsets may hinder our ability to identify and measure minor NKT and MAIT clusters. CITE-seq analyses of pre-enriched subsets could validate and expand our current findings. In contrast with a previous study (100), we observed no significant differences in the marker expression of cord blood MAITs or NKTs defined by tetramer or by their corresponding anti-TCR antibody staining. The tetramers performed consistently based on internal normalization and unloaded controls. For MAITs, the apparent discrepancy

between our results and those of *Swarbrick et al.* is most likely the result of differences in experimental design and gating strategy (100). The results they reported for both US and South African neonates compared cytokine responses of tetramer defined MAITs (hMR1+) to phenotypically defined MAITs (TRAV1/2+CD161+CD26+). Since the way hMR1+ was gated independent of additional markers, it is hard to determine whether the cells ending up in this gate were staining in response to 5-OP-RU, to the tetramer itself, or were the result of non-specific binding. Given that we observed greater frequency of TNF α production after PMA-ionomycin than what they observed, and that their data is not available, we are unable to ascertain whether some of the differences are due to technical differences or poor tetramer performance.

Likewise, many global health groups with an immunology focus continue to identify as NKTs to any cell stained as positive for CD3+CD56+. Our data directly contradicts this, as CD56 is not expressed by either antibody or tetramer-gated NKTs. Furthermore, the marker expression profile of NKTs is similar to that of the CD4 V δ 2 T cells, which also do not express CD56. The only CD56 expression for ILT subsets was a low level of expression found for the CD8 and PD1hi V δ 2 T cells.

Despite a heightened pro-inflammatory milieu reported at the fetal-maternal interface among pregnant women living with HIV, we only observed modest differences at the global ILT level between HU and HEU infants. The differences were represented primarily by altered MAIT and NKT frequencies for pooled HEU infant groups. The few markers that were differentially expressed when viewed in isolation may not represent biologically meaningful data, as they were frequently only seen for individual ILT subsets. While this defies a concise explanation of changes that occur in HEU infants in response to viral/ART

exposure before birth, these individual phenotypic marker changes at the subset level may denote functional changes that may be present after activation.

In summary, we expanded our knowledge of rare cell subsets in humans, extensively and robustly profiling the marker expression pattern of these cell subsets in neonates born to healthy women and those living with HIV. We additionally demonstrated the suitability of our toolsets when it comes to comprehensively assessing rare cell populations.

Chapter 7: Activated Cord Blood Innate-Like T cells

Introduction

In adults, in response to infection and inflammation, ILT rapidly produce Th1-like cytokines and release cytotoxic granules, making them important first-responders in the early immune response (4, 12). Technological limitations have hampered our ability to profile these cells in cord blood, but prior studies have demonstrated that they are present, and are capable of producing Th1-like cytokines, following stimulation. This is in contrast to the muted CD4 and CD8 T cell responses following stimulation during infancy. ILT may play an important role in response to infections during early life before other immune system branches are fully mature.

In utero exposure to infections may contribute to altered immune responses during infancy. For example, HEU infants experience higher morbidity to lower respiratory tract infections during the first six months of life compared to HU infants (even controlling for socio-economic factors), suggesting that immune system dysfunction may play a role (26). These disease manifestations could conceivably arise by either immune system hypo-responsiveness (with ensuing pathogen mediated damage) or hyper-responsiveness (via immune system collateral damage) during early life. Before birth, anything that alters the potential to produce pro-inflammatory cytokines (and cytotoxic mediators under correct circumstances like presence of IL-23) can impact what cells are present, and their responsiveness during this critical window. Given our limited understanding of these cell subsets and functionality during early life, it may be important to study the role of ILT. To our knowledge, no studies have examined cytokine responses of cord blood ILT in HEU infants. However, studies of cord blood Alpha Beta ($\alpha\beta$) T cells have conflicting data with

some demonstrating increased frequency of cytokine-producing cells in HEU infants, and other studies finding little demonstrable response following activation. Given the technological limitations (as well as unmeasured confounders in the clinical cohort recruiting), it is possible and understandable that some discrepancies would arise.

Confounding variables that can affect the heterogeneity of HIV exposure during pregnancy might include the presence and timing of maternal ART initiation and the kinetics of viral load during pregnancy. Additionally, recent studies suggest that ART may have some immune-activation effects (36), primarily via mitochondrial metabolism pathways, which can result in inflammatory responses via the caspase cascade. Both viral proteins and ART are capable of crossing the placental barrier. They may elicit fetal immune system activation either directly or through activation of innate cell subsets, which may in turn activate ILTs in an antigen-independent fashion. Whether these prematurely activated fetal cells persist and mature, or undergo activation induced cell death is unknown, but the resulting functional responses in early life would be noticeably different. Consequently, controlling for confounding variables in order to ascertain differences in cytokine responses is necessary.

To do this, we designed a 29-color SFC panel that contained the markers necessary to identify the various cord blood ILT populations, as well as assess activation, degranulation and cytokine production upon activation. To do so, we controlled for various factors (including screening for *Plasmodium falciparum*, Syphilis (*Treponema pallidum*), SARS-COV-2, Dengue (DENV), and Chikungunya (CHKV), stratifying by maternal HIV and ART exposure, and selecting specimens of similar sex and gestational age. We processed

over 60 cord blood specimens, stimulating them for 6 hours with PMA-ionomycin, before analyzing using both supervised and semi-supervised toolsets.

Results

To evaluate ILT function, we assessed their production of cytokines and mobilization of cytotoxic granules in response to polyclonal stimulation. Comparing the ILT at the global level (as defined in materials and methods), TNF α ⁺ cellular expression was the most commonly detected cytokine and the frequency of TNF α ⁺ cells was highest in NKTs, with a median frequency of 79.4% [71.4, 82.2], while medians ranged between 40% and 55% in the V δ 2 (51.3% [43.4, 58.8]) and MAIT (41.2% [33.0, 51.5]) populations (Figure 12A). The frequency of IFN γ ⁺ cells was lower than TNF α ⁺ cells in all ILT subsets, ranging from a median of 4.3% [2.4, 6.2] in MAITs, to 20.8% [14.5, 24.8] in NKTs, and up to 40.3% [32.1, 50.6] in V δ 2 cells. Polyfunctionality (defined as % TNF α ⁺IFN γ ⁺ cells) was highest in V δ 2s (25.2% [22.0, 32.4]), followed by NKTs (17.9% [12.7, 24.8]) and absent in MAIT cells. Extending the analysis to NK (CD7+CD3-NKG2D+CD161+) and conventional T cells (CD7+CD3+ excluding ILT), the median proportion of TNF α ⁺ cells was 46.8% [38.7, 55.1] for NK cells, 16.7% [12.0, 23.2] for CD4 T cells and 6.4% [3.6, 10.1] for CD8 T cells (Figure 12A), while the median proportion of IFN γ ⁺ cells was 74.8% [63.0, 83.3] for NK cells, 9.8% [8.8, 17.6] for CD8 T cells, and 1.1% [0.9, 1.4] for CD4 T cells (Figure 12A). Among these three cell populations, only the NK cells had a sizable proportion of polyfunctional cells (37.9% [32.6, 50.5]). Thus, ILT subsets at birth display a cytokine profile more akin to NK cells than conventional T cell subsets.

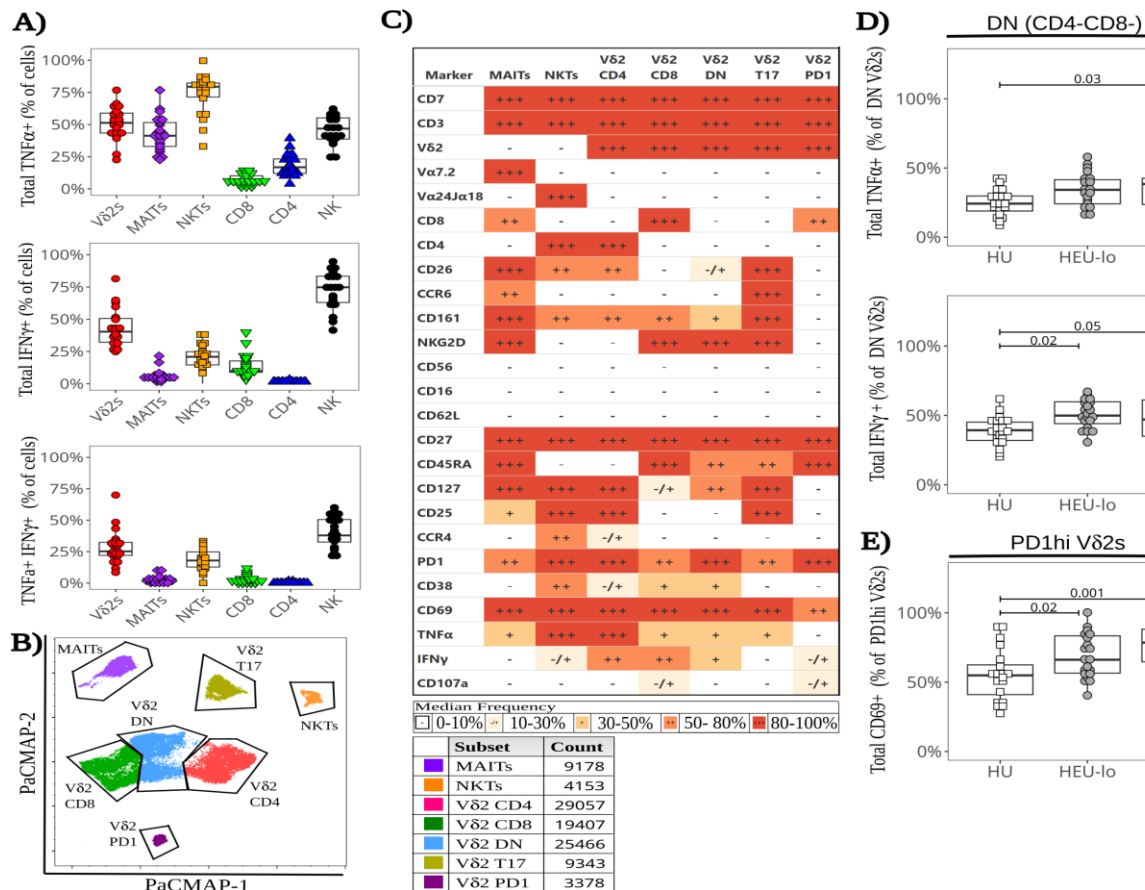


Figure 12: Functional heterogeneity of cord blood ILT in response to polyclonal stimulation ex-vivo with altered cytokine expression in HEU infants. CBMC were stained with a panel of 29 mAbs after a 6-hour stimulation with PMA plus ionomycin in the presence of Golgi inhibitors and CD107a. (A) The beeswarm boxplots compare the frequency of TNFα+ (top), IFNγ+ (middle), and TNFα+ IFNγ+ (bottom) cells for different manually gated cell subsets (listed on the x axis). Symbols depict individual values, boxplots show median and IQR, with whiskers indicating the +/- 1.5 IQR range. Shape and color identify HIV-exposure status. (B) A PaCMAP is used to visualize ILT subsets after PMA+ionomycin stimulation, including NKTs and MAITs (identified by anti-TCR staining). The gates are based on density-distribution and marker expression,

with color code and cell counts included in the legend. (C) The heatmap summarizes marker expression for all ILT clusters delineated in 4B for all study participants, using a combination of colors and symbols to display, for each marker, the median of individual frequencies of positive cells. (D, E) The beeswarm plots compare at the local level the expression of specific markers between exposure groups. The symbols show individual values, boxplots show median and IQR, with whiskers indicating the ± 1.5 IQR range. Shape and color identify HIV-exposure status. (D) Proportion of $\text{TNF}\alpha^+$ (top) and $\text{IFN}\gamma^+$ (bottom) cells in V δ 2 DN cluster. (E) Proportion of CD69^+ cells in the V δ 2 PD1hi cluster.

Dimensionality visualization plots of stimulated cells (Figure 12B) were generated including cytokines and most of the same markers used to generate the PaCMAP for unstimulated cells (Figure 9C). The algorithm formed the same clusters identified for unstimulated ILT, with a Vδ2 Th17 cluster segregated from the other Vδ2s (suggestive of distinct features compared to other clusters). MAIT and NKT cells each coalesced in a single cluster, and the results were consistent with the global ILT analysis performed at baseline (Figure 12A-C). The analysis of the Vδ2 clusters at the local level highlighted nuanced cytokine and degranulation profiles across subsets, consistent with our previously published data (48). In the CD4⁺ Vδ2 cluster, cells uniformly expressed TNFα (86.5% [78.4, 90.4]), with a lower proportion expressing IFNγ (46.2% [38.2, 55.7]) (Figures 12C, 13A). Less than 4% of the cells expressed CD107a, consistent with a low cytotoxic potential for CD4⁺ Vδ2 cells (48). Additionally, PD1 was upregulated compared to the unstimulated control (p-value 2.4×10^{-8}) (Figures 10A, 13A). Cells in the CD8⁺ Vδ2 cluster primarily expressed IFNγ (68.4% [59.6, 77.3]) (Figures 12C, 13B), with a lower proportion expressing TNFα (42.1% [32.6, 50.2]). A modest frequency of CD107a⁺ cells was present (18.8% [12.8, 25.1]), consistent with the frequency of CD56⁺ cells in the control treatment, and suggestive of cytotoxic potential (Figures 10B, 13B). The DN Vδ2 cluster had a cytokine profile comparable to the CD8⁺ Vδ2 population, but with less mobilization of cytotoxic granules (5.4% [3.4, 9.0]) (Figures 12C, 13C). Stimulation induced a decrease in CD127 expression (p-value 2.0×10^{-6}) (Figures 10C, 13C). Cells in the T17 Vδ2 cluster produced intermediate levels of TNFα (42.4% [34.2, 48.9]) (Figures 12C, 13D), but no IFNγ, and did not mobilize cytotoxic granules. The frequency of PD1⁺ cells was elevated after stimulation (p-value 6.0×10^{-7}) (Figures 10D, 13D). This functional

profile paralleled the activation pattern observed for MAITs. The PD1hi Vδ2 cluster had lower levels of CD69 expression (65.7% [54.6, 82.2]) (Figures 12C, 13E) compared to the other Vδ2 clusters (which were uniformly positive for this activation marker) and displayed a distinct cytokine profile: <1% of the cells produced TNFα, only 16.6% [11.6, 23.7] produced IFNγ and 22.2% [12.8, 32.4] were CD107a+, suggestive of modest cytotoxic potential and consistent with the levels of CD56 and CD16 expression observed for unstimulated PD1hi cells (Figures 10E, 13E).

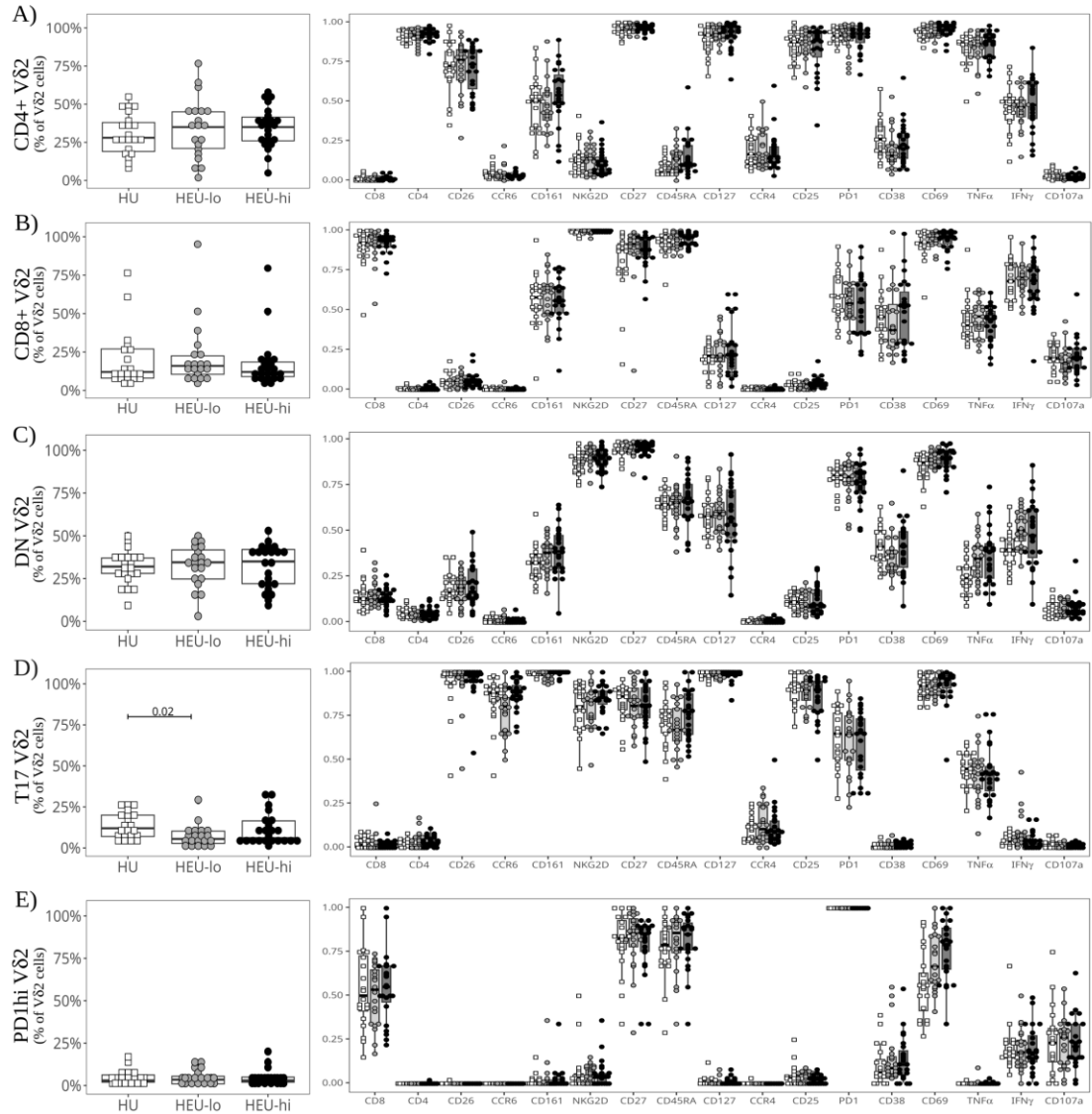


Figure 13: Semi-supervised analysis highlights cord blood V δ 2 T cell functional heterogeneity between clusters after polyclonal stimulation. Depicted are two columns of beeswarm plots with the left show the proportion of cells in each V δ 2 PaCMAP cluster (from Figure 12C), comparing the three exposure groups. The beeswarm plots on the right display the proportion of V δ 2 cells in the respective cluster expressing the individual markers listed on the x-axis. The symbols show individual values, boxplots show median and IQR, with whiskers indicating the ± 1.5 IQR range. Shape and color

identify HIV-exposure status. (A) depicts results for the CD4⁺ PaCMAP cluster, (B) for CD8⁺, (C) for DN, (D) for T17, and (E) for PD1hi.

When comparing exposure groups at the ILT global level (as described in materials and methods), we observed no significant difference in cytokine production for any ILT population. HEU-hi neonates, however, displayed substantial heterogeneity between specimens in the frequency of IFN γ ⁺ V δ 2 cells, with a significantly wider range of positivity (20%-80% of IFN γ ⁺ cells) compared to the other two groups (30%-60% among HEU-lo and HU, p-value 0.03). While CCR6 was downmodulated after stimulation in all exposure groups, the proportion of CCR6⁺ V δ 2 cells was significantly lower in HEU-lo infants compared to HU infants. We observed no significant differences across exposure groups in the frequency of individual clusters (Figure 13 A-E), with the exception of the T17 V δ 2 population, which was lower in HEU-lo infants after stimulation (p-value 0.02) (Figure 13D), consistent with a CCR6 down-modulation. However, a more granular comparison of PaCMAP V δ 2 clusters between exposure groups revealed differential upregulation of the activation marker CD69 in response to stimulation and differences in cytokine responses. Importantly, the DN V δ 2 cluster displayed a significantly higher proportion of TNF α ⁺ and IFN γ ⁺ in both HEU groups compared to the HU neonates, consistent with higher frequencies of CD69⁺ cells (p-value 0.05 for both HU vs. HEU-lo and HU vs. HEU-hi) (Figure 12D). For the PD1hi V δ 2 cluster, the proportion of CD69⁺ cells was higher in both HEU groups compared to the HU infants, possibly suggestive of a lower activation threshold in the HEU neonatal clusters (Figure 12E).

Discussion

Leveraging our semi-supervised analytical approach, we observed a level of functional heterogeneity in V δ 2 cells comparable to the phenotypic heterogeneity noted *ex vivo* in the previous chapter describing baseline expression. CD4⁺ V δ 2, which exhibited a greater breadth of memory differentiation than the other V δ 2 subsets, produced greater frequency of TNF α ⁺ and IFN γ ⁺. Both the CD8⁺ and CD4-CD8⁻ (DN) V δ 2 produced more IFN γ ⁺ than TNF α ⁺. By contrast, the T17-like CD26⁺ CCR6⁺ CD4-CD8⁻ population produced only TNF α ⁺ while the PD1 high V δ 2 subset showed minimal cytokine production, but a higher CD107a frequency, suggestive of degranulation.

We similarly saw a demarcation of cytokine responses for NKT and MAIT, with NKT paralleling the cytokine responses of CD4⁺ V δ 2, and the MAIT subset showing similarity to the T17 V δ 2 subset. Given that both baseline marker expression and cytokine responses exhibited the same demarcation pattern suggests that the phenotype and functional potential of these cell subsets are interrelated. Cord blood ILT subsets, despite featuring separate TCR arrangements, responsive to different antigens, may occupy similar functional niches before birth, while retaining the potential to mediate effector responses upon antigen exposure. When viewed in context of the broader $\alpha\beta$ T cell population, cells sharing the same collection of markers can be seen beyond the ILT subsets, for example, T17-like CD26⁺ CCR6⁺ CD161⁺ cells can be found across CD4⁺, CD8⁺ and CD4-CD8⁻ cord blood T cells, however their role is unknown, and it is not known whether stimulation would result in similar cytokine expression responses under appropriate conditions.

Despite a heightened pro-inflammatory milieu reported at the fetal-maternal interface among pregnant women living with HIV, we only observed modest differences at the

global ILT level between HU and HEU infants. At the local Vδ2 level, the large cluster of DN cells displayed significantly higher proportions of IFNγ+, TNFα+ and polyfunctional cells in HEU-hi compared to HU and HEU-lo neonates. This trend was seen to a lesser extent when it came to activation, with greater frequency of CD69+ cells being encountered for PD1 high Vδ2 in the HEU-hi. Given that we similarly saw an increase in IFNγ+ cells in CD4-CD8- T cells that were not ILT for the HEU-hi infants suggest that there might be an additional explanation impact of early life viral exposure seems to affect cells with these marker combinations to a greater extent.

HEU-hi infants also displayed a significantly wider frequency range of IFNγ+-expressing Vδ2 cells following ex vivo polyclonal stimulation compared to the other two groups, which was not accounted for by the Vδ2 cluster composition. The heterogeneity in maternal cumulative viremia and time of ART initiation during pregnancy likely resulted in disparate HIV/ART exposure among HEU-hi neonates, which may have contributed to our observation.

Our findings overlap, in part, with those of other HEU studies that observed greater Th1 cytokine production in HEU infants compared to their HU counterparts. At the same time, this difference was only noticeable for smaller subsets of Vδ2 T cells and may explain the disparate outcomes, as previous studies often only enrolled a small number of participants. Cytokine differences that were reported based on measurements at the whole subset level would have been driven by the share of the Vδ2 subsets (differential capacity to produce cytokines) present within the individual donor. For studies that were already small, a couple specimens skewed toward one particular subset would have been sufficient to drive the appearance of being statistically significant, when it was just subset composition all along.

Together, these findings highlight the functional complexity of innate-like T cells in cord blood, demonstrating that phenotypic heterogeneity within and across ILT subsets is mirrored by distinct cytokine response profiles. The impact of HIV exposure in utero on cytokine expression appears in specific V δ 2 subsets, underscoring the importance of investigating functional differences between HU and HEU neonates at a greater cellular depth, especially for future studies targeting other cell subsets beyond cord blood ILTs.

Conclusions and Future Directions

This study demonstrates that the use of SFC for samples of limited cell quantity, such as human pediatric specimens, allows for in-depth examination of low frequency subsets, including ILT, that neither mass cytometry nor single-cell RNA seq approaches could analyze in a cost-effective manner without prior enrichment.

We characterized the phenotype and functional profiles of cord blood ILT, in neonates born to mothers with or without HIV, at baseline and following polyclonal stimulation. The V δ 2 cell phenotypic heterogeneity appeared to correspond, at least in part, to functional characteristics, with different V δ 2 clusters mediating distinct cytokine responses. The three main clusters of V δ 2 cells identified by PaCMAP (CD4+, CD8+, DN) were skewed towards a Th1-biased cytokine response, producing both IFN γ and TNF α at varying levels. The T17 cluster, characterized by expression of CCR6 and CD26, produced TNF α but no IFN γ , while the PD1hi subset expressed very low levels of IFN γ , with no TNF α production. Of note, the CD4+ cluster, which includes the most differentiated V δ 2 cells based on surface markers, shared phenotypic (CCR4, CD25, and CD26) and functional features with NKT cells, with lymphocytes in both clusters secreting more TNF α than IFN γ . Conversely, T17 cells shared phenotypic properties with MAIT cells (CD161, CCR6, CD26) and neither subset appeared to produce IFN γ in any of the donors analyzed.

Our data suggest that ILT cells are mildly perturbed in neonates born to women with HIV who started ART before conception. However, modest differences in cytokine production (ex vivo or after culture) and in a few phenotypic markers, noted for HEU-hi neonates, suggest that immune perturbation may be more marked in this exposure group. The results are in agreement with prior studies that showed small or no differences between HU and

HEU infants in terms of immune responses and/or clinical outcomes when the mothers are on ART (19, 24). For example, a study conducted in Belgium showed no difference in the risk of hospitalization between HU and HEU infants born to women that started ART before pregnancy, but a significantly higher risk for HEU infants born to women that initiated ART during pregnancy, correlating with heightened cord blood monocyte activation (24). Therefore, ART initiation before pregnancy may be beneficial not only for the mother, but also for the infant. Initiation of ART preconception may mitigate fetal immune perturbation by limiting exposure to high maternal viremia and/or inflammation. However, preconception initiation could also prevent unintended effects on the developing fetal immune system that may occur when ART starts during pregnancy. The current study design does not allow us to discriminate between these possibilities.

Several study limitations should be considered when interpreting our results. While we have maternal clinical and HIV screening information at enrollment and delivery (as well as screening for other common pathogens), we cannot exclude that some women contracted undiagnosed or subclinical infections during pregnancy. Our sample size was relatively small, and, given the observed heterogeneity for multiple variables, we were likely underpowered to detect small but biologically relevant differences across exposure groups. All the specimens were cryopreserved and thawed, which may reduce cell viability and affect marker expression. Our observations for ILT, that respond to microbial metabolites, may be impacted by dietary/microbiota, environmental and geographic factors. Additionally, there may be temporal variability, and we assessed infant responses only at a single time point (at birth). These limitations, however, do not diminish the robustness of our analytical approach towards baseline neonatal ILT physiology, which allowed us to

rigorously assess these rare cell subsets and characterize them in neonates born to women living with HIV.

While the first full spectrum analyzer became commercially available in 2013, the implementation of this approach has become significantly more widespread in the last 5 years. The current trend is to repurpose packages and data analysis tools originally developed for conventional/mass cytometry or single cell RNA-seq for use in SFC. This practice, however, may not fully leverage SFC's core strengths, which include increased breadth compared to CFC and enhanced time/cost effectiveness compared to MC (often enabling improved depth). In addition, repurposing does not appropriately mitigate the variation in unmixing (due to autofluorescence, signature variance of unmixing controls, and tandem fluorophore degradation), which is unique to spectral cytometry. For example, data analysis tools initially intended for MC rarely handle the number of events acquired in SFC, making down sampling an obligatory step to use those tools. This approach does not fully leverage the depth of information obtained with SFC, potentially limiting our understanding of the immune system heterogeneity. Building tools capable of handling the unique SFC features and scalable to increasingly complex panels would enable a smoother integration into existing bioinformatic workflows for downstream analysis.

Our study leveraged new toolsets to characterize rare cell subsets from limited biospecimen. In the process of accomplishing this goal for our target cohort, large .fcs files containing millions of non-target cells were acquired. While not the focus of the thesis, same dataset and tools can be reanalyzed, allowing us to further understand how viral and ART exposure may impact other non-ILT, T cell subsets in cord blood, both at baseline and following activation. Alongside the Coereba R package method of defining cell

clusters, it can permit the incorporation of other publicly available datasets (both SFC and MC) to form a more comprehensive understanding of HEU infants.

Throughout the process of this PhD, we have tackled numerous questions pertaining to analytical challenges that would have been unanswerable at the time utilizing commercially available tools. In the process of addressing each challenge, we have created R packages (Luciernaga, Coereba, CytometryQC) that allow for troubleshooting of our existing workflows in a time efficient manner, pre-emptively identifying and correcting issues that could hamper data quality, and allowing for targeted interventions as needed. Besides being useful for their dedicated applications, they have potential for public use cases beyond the study of rare cell subsets, especially in the context of big data sets maintained in public repositories where provenance and metadata may not be well documented, limiting the ability to automate analysis.

One potential use case is the future application of Coereba to identify specimens where errors in processing may have contributed to cellular apoptosis, biasing cellular frequencies of cell populations, resulting in unrepresentative data points being included in the statistical analysis, as there are currently no tools that could detect these conditional aliquots that would have failed QC. Additional analysis of specimens beyond those collected in our study would be needed for validation but would enable researchers to avoid misinterpretation of poor experimental processing as biological heterogeneity.

In tackling these challenges, we have highlighted the power of open-source and open-science analytics, providing frameworks and tools that can be extended and shared with the wider community, where uptake will determine the usefulness of these tools over time.

In conclusion, we leveraged SFC to characterize rare subsets in neonates and optimized a panel (as well as analysis tools and code) that provides a backbone for ILT analysis and can be built upon as new fluorophore and additional conjugated antibodies become available. Our results, obtained comparing HIV-unexposed and two groups of HEU infants, suggest that ILT are mildly perturbed in HEU-lo, born to women who started ART before conception, but showed that the Vδ2 cell subset displays modest differences in HEU-hi neonates, born to women who started ART during pregnancy. This data, in alignment with prior literature, is consistent with the hypothesis that starting women on ART before conception is an effective approach to mitigating infant immune perturbation.

Appendix

All SFC .fcs files generated for the Cord-Blood Innate-Like T cell manuscript are available via the ImmPort repository (study accession SDY3080). The SFC data consists of the raw .fcs files for fully stained samples with the matched unstained and single-color unmixing controls. We have attempted to format the metadata in accordance with both ImmPort and MIFlow guidelines.

The visualized quality control data for the 5-laser Cytex Aurora on which the specimens were acquired in 2023 is available via the UMGCC FCSS dashboard (<https://umgccfcss.github.io/Aurora5L/>)

A full list of R packages and the R code used for the data visualizations and statistical analysis for this paper can be found at <https://github.com/DavidRach/CordBloodILT>, along with additional relevant and insightful information about laboratory protocols, instrumental quality, etc.

The Luciernaga (<https://github.com/DavidRach/Luciernaga>), Coereba (<https://github.com/DavidRach/Coereba>) and CytometryQC (<https://github.com/DavidRach/CytometryQC>) R packages are available on GitHub, under the free AGPL-3.0-or-later copyleft license.

Observation: Vδ2 After Expansion

After expansion with Bacillus Calmette-Guerin (BCG) or zoledronic acid monohydrate (ZOL) (Supplementary Figure 2, Supplementary Table 2), the frequency of naïve Vδ2 cells declined (median 54.0% ex vivo, 15.4% BCG, 10.9% ZOL; p-value 7×10^{-12} and 1.3×10^{-12})

for BCG and ZOL respectively) as they differentiated into Tcm (median 25.1% ex vivo, 63.4% BCG, 70.0% ZOL; p-value 7.1×10^{-7} and 7.1×10^{-7} for BCG and ZOL respectively) (Supplementary Figure 3). Expansion with either ZOL and BCG led to a lower proportion of NKG2D⁺ cells and a higher proportion of CD28⁺ cells compared to the IL2 control (Figure 14A). ZOL stimulation also resulted in a higher frequency of PD1⁺ cells compared to BCG and IL2, consistent with our previously published data (50)(Figure 14A). We did not observe any statistically significant differences in Vδ2 cell frequencies between exposure groups following expansion (data not shown).

In terms of marker expression, after ZOL expansion, HEU-hi infants displayed higher proportions of CD28⁺ Vδ2 cells compared to HEU-lo infants (Figure 14B), while HEU-lo infants had higher NKG2D levels compared to the HEU-hi infants (Figure 14B). This suggests that distinct underlying functional potential arising from differentiation may be present between the HEU groups (71). However, we did not observe statistically significant differences in perforin expression between exposure groups for any culture condition. We next re-stimulated expanded Vδ2 cells with plate-bound anti- $\gamma\delta$ TCR to assess IFN γ and TNF α production as well as degranulation (surface CD107a expression) (Figure 14C). We did not observe statistically significant differences between exposure groups for the ZOL or IL-2 expanded cells (Figure 14C). However, following culture with BCG + IL-2, HEU-hi infants displayed lower frequencies of TNF α ⁺, IFN γ ⁺, and polyfunctional TNF α ⁺IFN γ ⁺ cells compared to HU infants (Figure 14D).

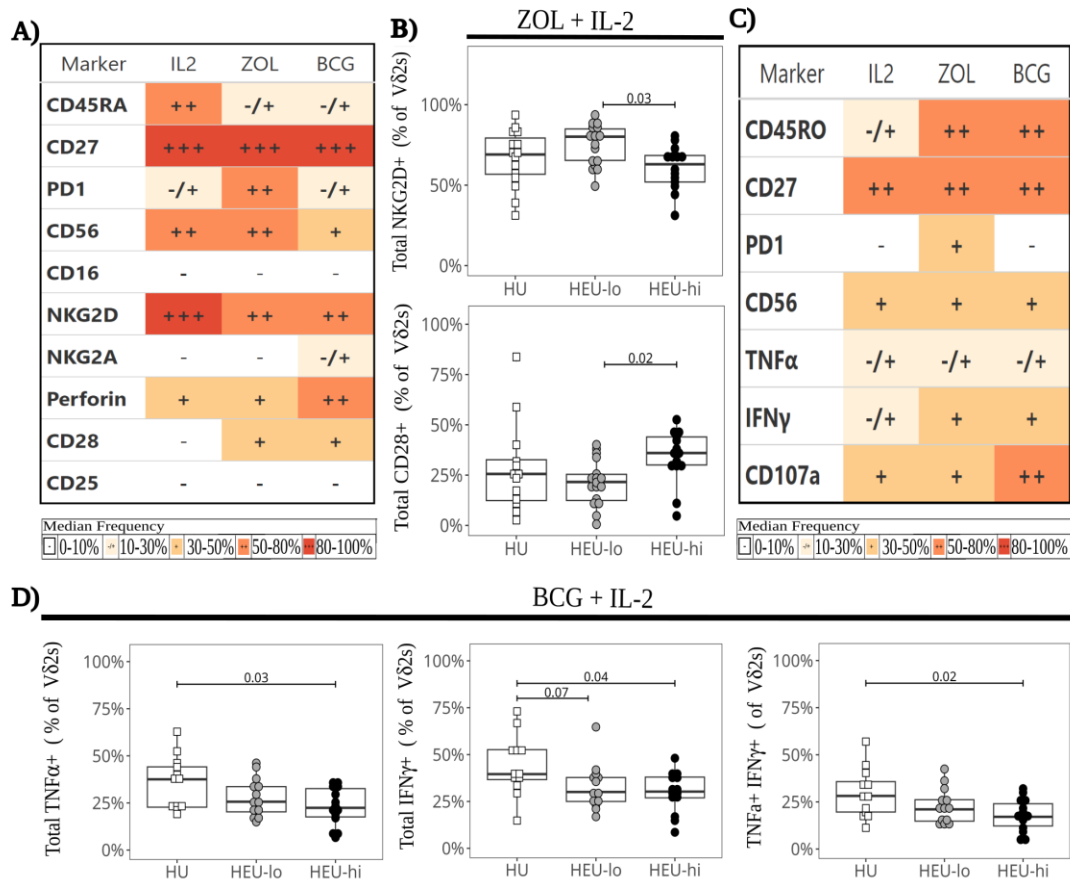


Figure 14: Lower cytokine production in BCG-expanded Vδ2 cells following restimulation in HEU infants. Vδ2 cells were stimulated in vitro with ZOL+IL-2, BCG+IL-2, or IL-2 alone and expanded in culture for 17 days. Expanded Vδ2 cells were restimulated with plate-bound anti-γδ TCR antibody (A) The heatmap summarizes marker expression of Vδ2 cells by culture condition on day 17 (before restimulation), using a combination of colors and symbols to display, for each marker, the median of individual frequencies of positive cells. (B) The beeswarm plots compare the expression of specific markers between exposure groups 17 days after ZOL stimulation (before restimulation), showing individual values, median and IQR, with whiskers showing the +/- 1.5 IQR range for NKG2D+ (left) and CD28+ (right) Vδ2 cells. (C) The heatmap summarizes by culture condition the Vδ2 cell cytokine production and degranulation in

response to restimulation, using a combination of colors and symbols to display, for each marker, the median of individual frequencies of positive cells. (D) The beeswarm plots compare exposure groups for cytokine production by BCG-expanded V δ 2 T cells. The response to restimulation is displayed as the frequency of TNF α ⁺ (left), IFN γ ⁺ (middle), and polyfunctional TNF α ⁺ IFN γ ⁺ (right) V δ 2 cells. The symbols show individual values, boxplots show median and IQR, with whiskers indicating the $\pm$ 1.5 IQR range. Shape and color identify HIV exposure status.

Discussion

We did not find substantial differences in markers, although we did observe small differences in CD28 expression upon ZOL stimulation. We did observe decreased cytokine responses in HEU-hi infants after expansion.

Unfortunately, we were unable to profile these cell cultures using the larger spectral flow cytometry panel, that would have provided substantially greater marker coverage than the four small panels in combination. Likewise, this might have permitted to ascertain which V δ 2 subsets made up the majority of the expanded V δ 2 in the culture, shedding insight into whether the decrease in cytokine frequency after restimulation in the HEU-hi was being driven by compositional differences, or the cells simply not responding as they had *ex vivo*. After culture with BCG, cells derived from the HEU-hi group displayed lower proportions of TNF α ⁺ and IFN γ ⁺ V δ 2 cells compared to HU neonates in response to TCR-mediated restimulation, suggesting a possible defect in recall responses. ZOL, an inhibitor of the enzyme farnesyl pyrophosphate synthase (FPPS), indirectly and potently stimulates V δ 2 cells by inducing upregulation of isopentenyl phosphate (IPP), the endogenous intermediate of cholesterol biosynthesis sensed by V δ 2 cells. BCG directly stimulates V δ 2 cells via the production of (E)-4-Hydroxy-3-methyl-but-2-enyl pyrophosphate (HMBPP), an intermediate of the non-mevalonate pathway of cholesterol biosynthesis, which is less potent than ZOL for cord blood V δ 2 cells stimulation. Since no differences in response to restimulation were evident for cells expanded with the potent zoledronate, the defect is likely subtle and only measurable after mild stimulation.

Observation: Sex-specific Differences

We compared ILT frequencies and marker expression at baseline by infant sex, pooling all infants regardless of HIV exposure groups (given the lack of major differences between exposure groups at the global level) (Figure 15). We observed the following differences in female compared to male neonates: significantly higher frequencies of CD8+, NKG2D+, and CD45RA+ cells and lower proportions of CD62L+ and CD25+ cells in the V δ 2 cell subset (Figure 15A); and significantly lower frequencies of PD-1+ MAIT cells (Figure 15B). These markers (except CD62L and PD1) were still differentially expressed after PMA+ionomycin stimulation (data not shown).

To confirm our observations, we employed a generalized-linear model comparison to account for the contribution of HIV/ART exposure. The results remained statistically significant for sex for all the above markers, with CD62L in V δ 2 cells also exhibiting an interaction with HIV/ART exposure for the HEU-hi group (Supplementary Table 3). Of note, there were no differences between males and females in the production of cytokines or cytotoxic granule release after stimulation both ex-vivo or post-expansion.

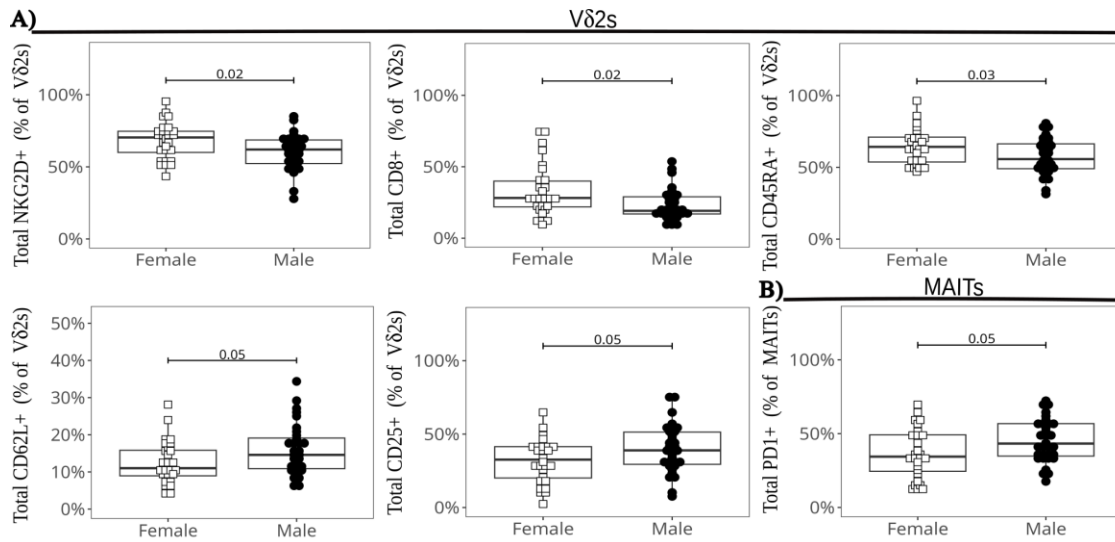
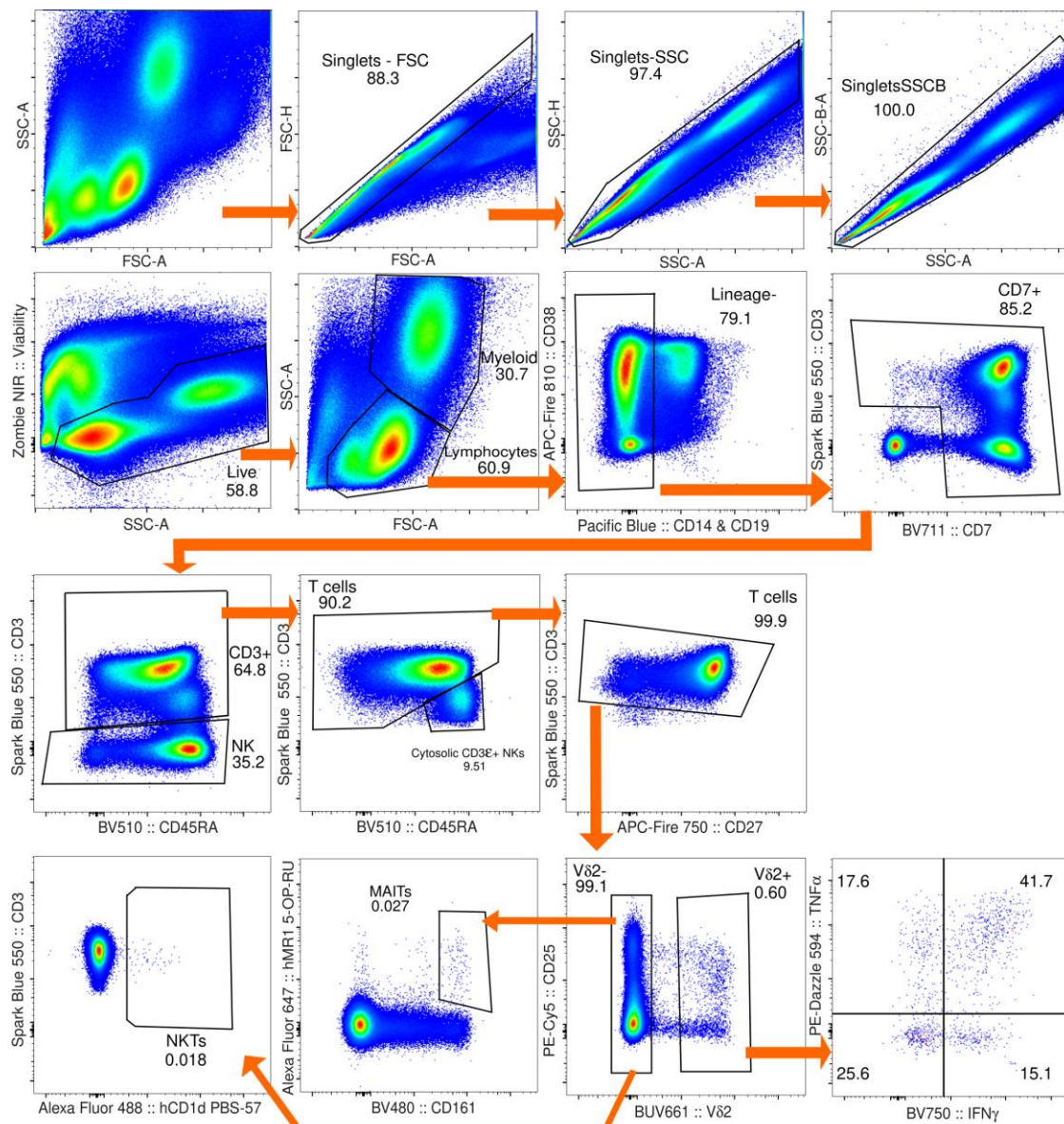


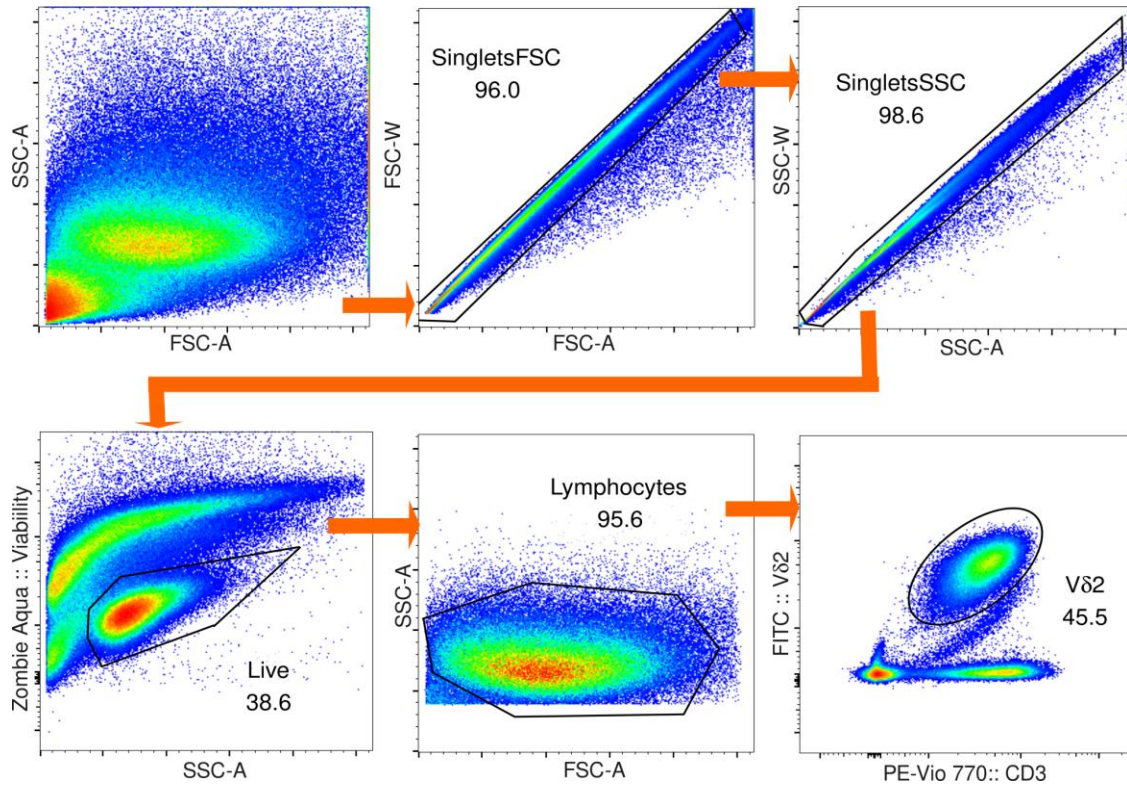
Figure 15: Differences in ILT global marker expression by infant sex. The beeswarm plots show the differentially expressed markers in Vδ2 cells (A), and in MAIT cells (B). The symbols show individual values and the boxplots show median and IQR, with whiskers indicating the ± 1.5 IQR range. Shape and color identify infant sex.

Discussion

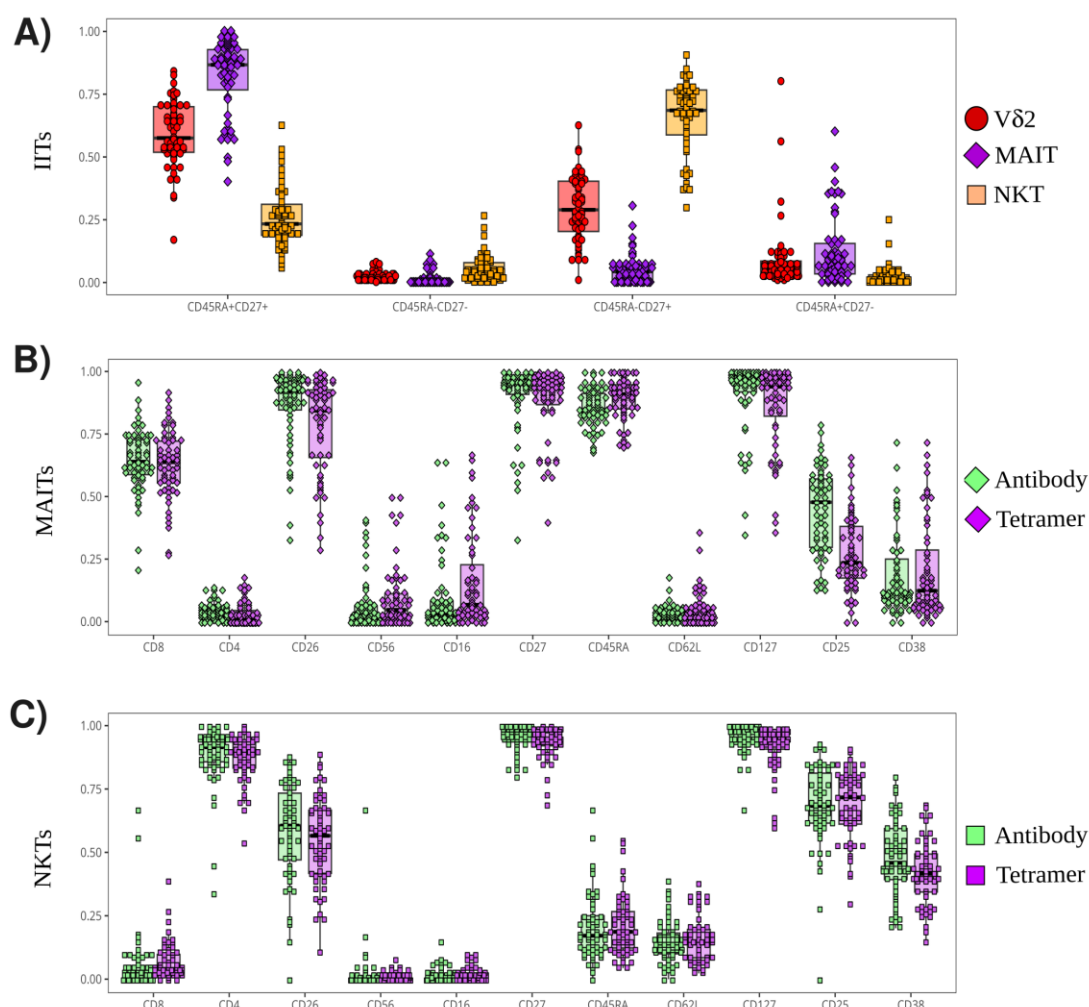
Throughout this study we noted small but internally consistent differences across conditional aliquots between male and female neonates in the expression of surface markers (CD62L, PD1, NKG2D, CD45RA and CD8) across multiple subsets, but predominantly V δ 2 cells. ILT subset frequencies did not account for the observed differences. These disparities remained significant after controlling for other variables, such as exposure and gestational age. Preliminary analyses of other cell subsets (NKs and conventional $\alpha\beta$ T cells) confirmed differential expression of the same markers between male and female neonates. Importantly, the sex-related differences did not appear to have functional implications, since cytokines, CD69, and CD107a expression were comparable between males and females. Previous studies have reported differences in inflammatory cytokines, androgen receptors, and FoxP3⁺ Tregs during pregnancy with male versus female neonates. Whether our observation is generalizable, or unique to our system, remains to be determined.



Supplementary Figure 1: Gating strategy for spectral flow cytometry (SFC) panel. The dot plots illustrate the gating strategy used to identify the main ILT nodes for a representative HIV-unexposed (HU) infant, with descending hierarchical gates denoted by arrows. Individual gates and frequency in the parent gate are shown.

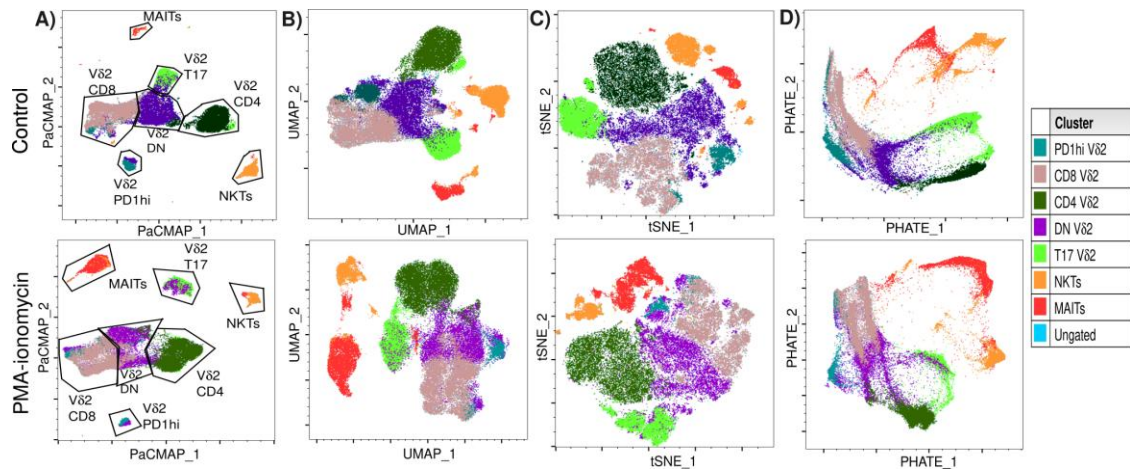


Supplementary Figure 2: Gating strategy for cultured day 17 conventional flow cytometry (CFC) panel. The dot plots illustrate the gating strategy used to identify expanded Vδ2 cells (D17) for a representative HU infant specimen, with descending hierarchical gates denoted by arrows. Individual gates and frequency in the parent gate are shown.



Supplementary Figure 3: Differentiation of Cord Blood ILTs. (A) The beeswarm plot displays the memory differentiation across ILT subsets, with each subset designated by a color and a symbol. Individual symbols represent the individual proportions of cells falling into each differentiation state listed on the x-axis. (B) The beeswarm plots compare the expression of specific markers in MAITs defined by either antibody (CD161hi Vα7.2+) or tetramer (CD161hi hMR1 5-OP-RU+), designated by shape and color. (C) The beeswarm plots compare the expression of specific markers in NKTs identified by either antibody (Vα24Jα18+) or tetramer (hCD1d PBS-57+), designated by

shape and color. Symbols represent individual values, boxplots depict medians and IQR, with whiskers showing the ± 1.5 IQR range.



Supplementary Figure 4: Different visualization algorithms identify comparable ILT clusters. The results of multiple visualization algorithms are shown side by side, with the same ILT cell gates (manually drawn in concatenated ILT files) overlaid on the different maps. (A) PaCMAP, (B) UMAP, (C) tSNE, (D) and PHATE results are shown for unstimulated (top) and PMA+ionomycin stimulated ILT.

Supplementary Table 1 SFC Antibodies

Marker	Alternate Name	Clone	Isotype	Fluorophore	Manufacturer	Catalog Number	Temperature (°C)	Location
CD62L		SK11	Mouse IgG2a, k	BUV395	BD Biosciences	565219	4°	Surface
CD8	CD8a	RPA-T8	Mouse IgG1, k	BUV496	BD Biosciences	612942	4°	Surface
CD69		FN50	Mouse IgG1, k	BUV563	BD Biosciences	748764	4°, RT	Surface, IC
CD194	CCR4	1G1	Mouse IgG1, k	BUV615	BD Biosciences	613000	37°	Surface
V82		B6	Mouse IgG2, k	BUV661	BD Biosciences	750056	4°, RT	Surface, IC
CD183	CXCR3	1C6/CXCR3	Mouse IgG1, k	BUV737	BD Biosciences	741866	4°	Surface
CD4		SK3	Mouse IgG1, k	BUV805	BD Biosciences	612887	37°, RT	Surface, IC
CD127	IL-7Ra	A019D5	Mouse IgG1, k	BV421	BioLegend	351310	37°	Surface
CD14		M5E2	Mouse IgG2a, k	Pacific Blue	BioLegend	301828	4°	Surface
CD19		SJ25C1	Mouse IgG1, k	Pacific Blue	BioLegend	363036	4°	Surface
CD161		HP-3G10	Mouse IgG1, k	BV480	BD Biosciences	748279	37°	Surface
CD45RA		HI100	Mouse IgG2b, k	BV510	BioLegend	304142	4°	Surface
CD56		5.1H11	Mouse IgG1, k	BV605	BioLegend	362538	37°, RT	Surface, IC
CD197	CCR7	G043H7	Mouse IgG2a, k	BV650	BioLegend	353234	37°, RT	Surface, IC
CD7		M-T701	Mouse IgG1, k	BV711	BD Biosciences	564018	37°	Surface
IFN γ		B27	Mouse IgG1, k	BV750	BD Biosciences	566357	RT	Intracellular
CD196	CCR6	11A9	Mouse IgG1, k	BV786	BD Biosciences	563704	37°	Surface
CD3		SK7	Mouse IgG1, k	Spark Blue 550	BioLegend	344852	4°, RT	Surface, IC

Supplementary Table 1 SFC Antibodies

Marker	Alternate Name	Clone	Isotype	Fluorophore	Manufacturer	Catalog Number	Temperature (°C)	Location
hCD1d PBS-57			Tetramer	Alexa Fluor 488	NIH Tetramer Core		RT	Surface
Vα24Jα18		6B11	Mouse IgG1, k	FITC	BioLegend	342906	RT	Surface
CD314	NKG2D	1D11	Mouse IgG1, k	PE	BioLegend	320806	37°, RT	Surface, IC
CD26		BA5b	Mouse IgG2a, k	PerCP-Cy5.5	BioLegend	302716	RT	Intracellular
CD25	IL-2Ra	M-A251	Mouse IgG1, k	PE-Cy5	BD Biosciences	555433	37°, RT	Surface
TNFα		MAb11	Mouse IgG1, k	PE-Dazzle 594	BioLegend	502946	RT	Intracellular
CD279	PD1	PD1.3.1.3	Mouse IgG2b	PE-Vio 770	Miltenyi Biotec	130-117-698	37°, RT	Surface, IC
CD16	FcyRIII	3G8	Mouse IgG1, k	APC	BioLegend	302012	4°	Surface
hMR1 5-OP-RU			Tetramer	Alexa Fluor 647	NIH Tetramer Core		RT	Surface
Vα7.2		3C10	Mouse IgG1, k	Alexa Fluor 647	BioLegend	351726	RT	Surface
CD107a		H4A3	Mouse IgG1, k	APC-R700	BD Biosciences	565184	37°	Surface
CD27		O323	Mouse IgG1, k	APC-Fire 750	BioLegend	302846	37°	Surface
CD38		HIT2	Mouse IgG1, k	APC-Fire 810	BioLegend	303550	37°	Surface
CD3*		UCHT1	Mouse IgG1, k	Alexa Fluor 488	BioLegend	300415	4°	Surface
CD3*		OKT3	Mouse IgG2a, k	Alexa Fluor 647	BioLegend	317312	4°	Surface

RT = Stained at room temperature; 4°, RT = Antibody incubated at 4°C for surface staining and at room temperature for intracellular staining; 37°, RT = Antibody incubated at 37°C for surface staining and at room temperature for intracellular staining; IC = Intracellular. *Only used as unmixing controls for tetramers on respective fluorophores.

Supplementary Table 2 CFC Antibodies

Marker	Alternate Name	Clone	Isotype	Fluorophore	Manufacturer	Catalog Number	Panel	Location
CD279	PD1	EH12.2H7	Mouse IgG1, k	BV421	BioLegend	329920	CK0, CK17, 1, 4b	Surface
V82		B6	Mouse IgG1, k	BV421	BioLegend	331428	2	Surface
Viability				Zombie Aqua Fixable Viability dye	BioLegend	423101	CK0, CK17, 1, 4b	Surface
CD3		UCHT1	Mouse IgG1, k	BV510	BioLegend	300448	2	Surface
CD56	NCAM	5.1H11	Mouse IgG1, k	BV650	BioLegend	362532	CK17, 4b	Surface
CD27		O323	Mouse IgG1, k	BV650	BioLegend	302828	CK0	Surface
CD16		3G8	Mouse IgG1, k	BV650	BioLegend	302042	1	Surface
CD314	NKG2D	1D11	Mouse IgG1, k	Biotin	BioLegend	320804	2	Surface
Streptavidin				BV650	BioLegend	405232	2	Surface
CD107a	LAMP-1	H4A3	Mouse IgG1, k	Alexa Fluor 488	BioLegend	328610	CK17	Surface
Perforin		dG9	Mouse IgG2b, k	Alexa Fluor 488	BioLegend	308108	2, 4b	Intracellular
V82		B6	Mouse IgG1, k	FITC	BioLegend	331406	CK0, 1	Surface
CD3		OKT3	Mouse IgG2a, k	PerCP-eFluor 710	Invitrogen	46-0037-42	CK0, 4b	Surface
CD25		CD25-4E3	Mouse IgG2b, k	PerCP-eFluor 710	Invitrogen	46-0257-42	1	Surface
TNF α		MAb11	Mouse IgG1, k	PerCP-Cy5.5	BioLegend	502926	CK17	Intracellular
CD56	NCAM	TULY56	Mouse IgG1, k	PerCP-eFluor 710	Invitrogen	46-0566-42	2	Surface
Granzyme B		GB11	Mouse IgG1, k	PE	BD Biosciences	561142	4b	Intracellular
CD56		5.1H11	Mouse IgG1, k	PE	BioLegend	362508	CK0	Surface

Supplementary Table 2 CFC Antibodies

Marker	Alternate Name	Clone	Isotype	Fluorophore	Manufacturer	Catalog Number	Panel	Location
CD28		"15E8"	Mouse IgG1, k	PE	Miltenyi Biotec	130-126-172	1	Surface
V82		B6	Mouse IgG1, k	PE	BioLegend	331408	CK17	Surface
CD279	PD1	EH12.2H7	Mouse IgG1, k	PE	BioLegend	329906	2	Surface
CD27		O323	Mouse IgG1, k	PE-Dazzle 594	BioLegend	302844	1	Surface
CD159a	NKG2A	REA110	Recombinant human IgG1	PE-Vio 770	Miltenyi Biotec	130-113-567	2, 4b	Surface
IFN γ		45-15	Mouse IgG1, k	PE-Vio 770	Miltenyi Biotec	130-113-494	CK0, CK17	Intracellular
CD3		BW264/58	Mouse IgG2a, k	PE-Vio 770	Miltenyi Biotec	130-113-130	1	Surface
V82		B6	Mouse IgG1, k	APC	BioLegend	331418	4b	Surface
CD27		O323	Mouse IgG1, k	APC	BioLegend	302810	CK17	Surface
TNF α		MAb11	Mouse IgG1, k	Alexa Fluor 647	BioLegend	502916	CK0	Intracellular
V81		REA173	Recombinant human IgG1	APC	Miltenyi Biotec	130-118-968	1, 2	Surface
CD16		3G8	Mouse IgG1, k	APC-Fire 750	BioLegend	302060	4b	Surface
CD45RO		UCHL1	Mouse IgG2a, k	APC-Fire 750	BioLegend	304250	CK0, CK17	Surface
CD45RA		HI100	Mouse IgG2b, k	APC-Fire 750	BioLegend	304152	1	Surface
Viability				BD Horizon 780	BD Biosciences	565388	2	Surface
CD3*		UCHT1	Mouse IgG1, k	BV421	BioLegend	300434		Surface
CD3*		UCHT1	Mouse IgG1, k	BV510	BioLegend	300448		Surface
CD3*		OKT3	Mouse IgG2a, k	BV650	BioLegend	317324		Surface
CD3*		UCHT1	Mouse IgG1, k	FITC	BioLegend	300406		Surface

Supplementary Table 2 CFC Antibodies

Marker	Alternate Name	Clone	Isotype	Fluorophore	Manufacturer	Catalog Number	Panel	Location
CD3*		OKT3	Mouse IgG2a, k	PerCP-eFluor 710	Invitrogen	46-0037-42		Surface
CD3*		UCHT1	Mouse IgG1, k	PE	BioLegend	300441		Surface
CD3*		UCHT1	Mouse IgG1, k	PE-Dazzle 594	BioLegend	300450		Surface
CD3*		BW264/58	Mouse IgG2a, k	PE-Vio 770	Miltenyi Biotec	130-113-130		Surface
CD3*		OKT3	Mouse IgG2a, k	APC	BioLegend	317318		Surface
CD3*		UCHT1	Mouse IgG1, k	APC-Fire 750	BioLegend	300470		Surface

1 = Phenotype, 2 = NK markers, 4b = Cytotoxicity, CK = Cytokine, 0 = *Ex Vivo*, 17 = Post expansion

* Only used for compensation controls

Supplementary Table 3. Generalized Linear Model Results

Marker ~ Infant Sex + HIV/ART Exposure
Separate models for each dependent variable

Tested	Predictor	OR	SE	z- value	p-value	OR [95% CI]	¹
V82							
NKG2D	Intercept	2.01	0.03	20.62	< 0.001	2.01 [1.88–2.15]	***
NKG2D	Sex: Male	0.92	0.03	-2.55	0.014	0.92 [0.87–0.98]	*
NKG2D	Exposure: HEU-lo	1.01	0.04	0.20	0.845	1.01 [0.93–1.09]	
NKG2D	Exposure: HEU-hi	0.95	0.04	-1.43	0.158	0.95 [0.88–1.02]	
CD8	Intercept	1.40	0.04	8.18	< 0.001	1.40 [1.29–1.52]	***
CD8	Sex: Male	0.90	0.04	-2.61	0.012	0.90 [0.84–0.97]	*
CD8	Exposure: HEU-lo	1.01	0.05	0.18	0.855	1.01 [0.92–1.11]	
CD8	Exposure: HEU-hi	0.97	0.05	-0.55	0.582	0.97 [0.89–1.07]	
CD45RA	Intercept	1.93	0.03	18.99	< 0.001	1.93 [1.80–2.07]	***
CD45RA	Sex: Male	0.93	0.03	-2.29	0.026	0.93 [0.87–0.99]	*
CD45RA	Exposure: HEU-lo	0.99	0.04	-0.13	0.896	0.99 [0.92–1.08]	
CD45RA	Exposure: HEU-hi	0.97	0.04	-0.66	0.509	0.97 [0.90–1.05]	
CD62L	Intercept	1.11	0.02	5.98	< 0.001	1.11 [1.07–1.15]	***
CD62L	Sex: Male	1.04	0.02	2.18	0.034	1.04 [1.00–1.07]	*
CD62L	Exposure: HEU-lo	1.02	0.02	0.95	0.346	1.02 [0.98–1.06]	
CD62L	Exposure: HEU-hi	1.04	0.02	2.10	0.04	1.04 [1.00–1.08]	*
CD25	Intercept	1.36	0.04	7.07	< 0.001	1.36 [1.25–1.48]	***
CD25	Sex: Male	1.09	0.04	2.04	0.046	1.09 [1.00–1.18]	*
CD25	Exposure: HEU-lo	0.99	0.05	-0.19	0.852	0.99 [0.90–1.10]	
CD25	Exposure: HEU-hi	1.04	0.05	0.74	0.464	1.04 [0.94–1.14]	
MAIT							
PD1	Intercept	1.46	0.04	8.69	< 0.001	1.46 [1.34–1.58]	***
PD1	Sex: Male	1.09	0.04	2.00	0.05	1.09 [1.00–1.18]	*
PD1	Exposure: HEU-lo	1.03	0.05	0.60	0.554	1.03 [0.93–1.14]	
PD1	Exposure: HEU-hi	0.96	0.05	-0.73	0.468	0.96 [0.88–1.06]	
¹ ***p < 0.001; **p < 0.01; *p < 0.05							

References

1. Awad W, Ler GJM, Xu W, Keller AN, Mak JYW, Lim XY, Liu L, Eckle SBG, Le Nours J, McCluskey J, Corbett AJ, Fairlie DP, Rossjohn J. The molecular basis underpinning the potency and specificity of MAIT cell antigens. *Nat Immunol.* 2020;21(4):400-11. Epub 2020/03/04. doi: 10.1038/s41590-020-0616-6. PubMed PMID: 32123373.
2. Eckle SB, Birkinshaw RW, Kostenko L, Corbett AJ, McWilliam HE, Reantragoon R, Chen Z, Gherardin NA, Beddoe T, Liu L, Patel O, Meehan B, Fairlie DP, Villadangos JA, Godfrey DI, Kjer-Nielsen L, McCluskey J, Rossjohn J. A molecular basis underpinning the T cell receptor heterogeneity of mucosal-associated invariant T cells. *J Exp Med.* 2014;211(8):1585-600. Epub 2014/07/23. doi: 10.1084/jem.20140484. PubMed PMID: 25049336; PMCID: PMC4113946.
3. Melandri D, Zlatareva I, Chaleil RAG, Dart RJ, Chancellor A, Nussbaumer O, Polyakova O, Roberts NA, Wesch D, Kabelitz D, Irving PM, John S, Mansour S, Bates PA, Vantourout P, Hayday AC. The gammadeltaTCR combines innate immunity with adaptive immunity by utilizing spatially distinct regions for agonist selection and antigen responsiveness. *Nat Immunol.* 2018;19(12):1352-65. Epub 2018/11/14. doi: 10.1038/s41590-018-0253-5. PubMed PMID: 30420626; PMCID: PMC6874498.
4. Godfrey DI, Uldrich AP, McCluskey J, Rossjohn J, Moody DB. The burgeoning family of unconventional T cells. *Nat Immunol.* 2015;16(11):1114-23. Epub 2015/10/21. doi: 10.1038/ni.3298. PubMed PMID: 26482978.
5. Mayassi T, Barreiro LB, Rossjohn J, Jabri B. A multilayered immune system through the lens of unconventional T cells. *Nature.* 2021;595(7868):501-10. Epub 2021/07/23. doi: 10.1038/s41586-021-03578-0. PubMed PMID: 34290426; PMCID: PMC8514118.
6. Ben Youssef G, Tourret M, Salou M, Ghazarian L, Houdouin V, Mondot S, Mburu Y, Lambert M, Azarnoush S, Diana JS, Virlovvet AL, Peuchmaur M, Schmitz T, Dalle JH, Lantz O, Biran V, Caillat-Zucman S. Ontogeny of human mucosal-associated invariant T cells and related T cell subsets. *J Exp Med.* 2018;215(2):459-79. Epub 2018/01/18. doi: 10.1084/jem.20171739. PubMed PMID: 29339446; PMCID: PMC5789419.
7. Moreira-Teixeira L, Resende M, Coffre M, Devergne O, Herbeuval JP, Hermine O, Schneider E, Rogge L, Ruemmele FM, Dy M, Cordeiro-da-Silva A, Leite-de-Moraes MC. Proinflammatory environment dictates the IL-17-producing capacity of human invariant NKT cells. *J Immunol.* 2011;186(10):5758-65. Epub 2011/04/12. doi: 10.4049/jimmunol.1003043. PubMed PMID: 21478400.
8. Papadopoulou M, Dimova T, Shey M, Briel L, Veldtsman H, Khomba N, Africa H, Steyn M, Hanekom WA, Scriba TJ, Nemes E, Vermijlen D. Fetal public Vgamma9Vdelta2 T cells expand and gain potent cytotoxic functions early after birth. *Proc Natl Acad Sci U S A.* 2020;117(31):18638-48. Epub 2020/07/16. doi: 10.1073/pnas.1922595117. PubMed PMID: 32665435; PMCID: PMC7414170.
9. Raffetseder J, Lindau R, van der Veen S, Berg G, Larsson M, Ernerudh J. MAIT Cells Balance the Requirements for Immune Tolerance and Anti-Microbial Defense During Pregnancy. *Front Immunol.* 2021;12:718168. Epub 2021/09/10. doi: 10.3389/fimmu.2021.718168. PubMed PMID: 34497611; PMCID: PMC8420809.
10. Solders M, Gorchs L, Erkers T, Lundell AC, Nava S, Gidlof S, Tiblad E, Magalhaes I, Kaipe H. MAIT cells accumulate in placental intervillous space and display a highly cytotoxic phenotype upon bacterial stimulation. *Sci Rep.* 2017;7(1):6123. Epub 2017/07/25. doi: 10.1038/s41598-017-06430-6. PubMed PMID: 28733576; PMCID: PMC5522401.
11. Solders M, Gorchs L, Tiblad E, Gidlof S, Leeansyah E, Dias J, Sandberg JK, Magalhaes I, Lundell AC, Kaipe H. Recruitment of MAIT Cells to the Intervillous Space of the Placenta by Placenta-Derived Chemokines. *Front Immunol.* 2019;10:1300. Epub 2019/06/28. doi: 10.3389/fimmu.2019.01300. PubMed PMID: 31244846; PMCID: PMC6563723.

12. Gumperz JE, Miyake S, Yamamura T, Brenner MB. Functionally distinct subsets of CD1d-restricted natural killer T cells revealed by CD1d tetramer staining. *J Exp Med*. 2002;195(5):625-36. Epub 2002/03/06. doi: 10.1084/jem.20011786. PubMed PMID: 11877485; PMCID: PMC2193772.
13. Anderson J, Thang CM, Thanh LQ, Dai VTT, Phan VT, Nhu BTH, Trang DNX, Trinh PTP, Nguyen TV, Toan NT, Harpur CM, Mulholland K, Pellicci DG, Do LAH, Licciardi PV. Immune Profiling of Cord Blood From Preterm and Term Infants Reveals Distinct Differences in Pro-Inflammatory Responses. *Front Immunol*. 2021;12:777927. Epub 2021/11/19. doi: 10.3389/fimmu.2021.777927. PubMed PMID: 34790206; PMCID: PMC8591285.
14. Odorizzi PM, Feeney ME. Impact of In Utero Exposure to Malaria on Fetal T Cell Immunity. *Trends Mol Med*. 2016;22(10):877-88. Epub 2016/09/12. doi: 10.1016/j.molmed.2016.08.005. PubMed PMID: 27614925; PMCID: PMC5048621.
15. Lim AI, McFadden T, Link VM, Han SJ, Karlsson RM, Stacy A, Farley TK, Lima-Junior DS, Harrison OJ, Desai JV, Lionakis MS, Shih HY, Cameron HA, Belkaid Y. Prenatal maternal infection promotes tissue-specific immunity and inflammation in offspring. *Science*. 2021;373(6558). Epub 2021/08/28. doi: 10.1126/science.abf3002. PubMed PMID: 34446580.
16. Babik JM, Cohan D, Monto A, Hartigan-O'Connor DJ, McCune JM. The human fetal immune response to hepatitis C virus exposure in utero. *J Infect Dis*. 2011;203(2):196-206. Epub 2011/02/04. doi: 10.1093/infdis/jiq044. PubMed PMID: 21288819; PMCID: PMC3071071.
17. Garcia-Flores V, Romero R, Xu Y, Theis KR, Arenas-Hernandez M, Miller D, Peyvandipour A, Bhatti G, Galaz J, Gershater M, Levenson D, Pusod E, Tao L, Kracht D, Florova V, Leng Y, Motomura K, Para R, Faucett M, Hsu CD, Zhang G, Tarca AL, Pique-Regi R, Gomez-Lopez N. Maternal-fetal immune responses in pregnant women infected with SARS-CoV-2. *Nat Commun*. 2022;13(1):320. Epub 2022/01/20. doi: 10.1038/s41467-021-27745-z. PubMed PMID: 35042863; PMCID: PMC8766450.
18. Hurtado CW, Golden-Mason L, Brocato M, Krull M, Narkewicz MR, Rosen HR. Innate immune function in placenta and cord blood of hepatitis C--seropositive mother-infant dyads. *PLoS One*. 2010;5(8):e12232. Epub 2010/09/04. doi: 10.1371/journal.pone.0012232. PubMed PMID: 20814429; PMCID: PMC2923602.
19. le Roux SM, Abrams EJ, Donald KA, Brittain K, Phillips TK, Zerbe A, le Roux DM, Kroon M, Myer L. Infectious morbidity of breastfed, HIV-exposed uninfected infants under conditions of universal antiretroviral therapy in South Africa: a prospective cohort study. *Lancet Child Adolesc Health*. 2020;4(3):220-31. Epub 2020/01/15. doi: 10.1016/S2352-4642(19)30375-X. PubMed PMID: 31932246; PMCID: PMC7235356.
20. Slogrove AL, Goetghebuer T, Cotton MF, Singer J, Bettinger JA. Pattern of Infectious Morbidity in HIV-Exposed Uninfected Infants and Children. *Front Immunol*. 2016;7:164. Epub 2016/05/21. doi: 10.3389/fimmu.2016.00164. PubMed PMID: 27199989; PMCID: PMC4858536.
21. Amenyogbe N, Dimitriu P, Cho P, Ruck C, Fortuno ES, 3rd, Cai B, Alimenti A, Cote HCF, Maan EJ, Slogrove AL, Esser M, Marchant A, Goetghebuer T, Shannon CP, Tebbutt SJ, Kollmann TR, Mohn WW, Smolen KK. Innate Immune Responses and Gut Microbiomes Distinguish HIV-Exposed from HIV-Unexposed Children in a Population-Specific Manner. *J Immunol*. 2020;205(10):2618-28. Epub 2020/10/18. doi: 10.4049/jimmunol.2000040. PubMed PMID: 33067377; PMCID: PMC7653510 other authors have no financial conflicts of interest.
22. Dirajlal-Fargo S, Mussi-Pinhata MM, Weinberg A, Yu Q, Cohen R, Harris DR, Bowman E, Gabriel J, Kulkarni M, Funderburg N, Chakhtoura N, McComsey GA, Protocol NL. HIV-exposed-uninfected infants have increased inflammation and monocyte activation. *AIDS*. 2019;33(5):845-53. Epub 2019/01/17. doi: 10.1097/QAD.0000000000002128. PubMed PMID: 30649056; PMCID: PMC6494115.
23. Garcia-Knight MA, Nduati E, Hassan AS, Gambo F, Odera D, Etyang TJ, Hajj NJ, Berkley JA, Urban BC, Rowland-Jones SL. Altered Memory T-Cell Responses to *Bacillus*

Calmette-Guerin and Tetanus Toxoid Vaccination and Altered Cytokine Responses to Polyclonal Stimulation in HIV-Exposed Uninfected Kenyan Infants. *PLoS One*. 2015;10(11):e0143043. Epub 2015/11/17. doi: 10.1371/journal.pone.0143043. PubMed PMID: 26569505; PMCID: PMC4646342.

24. Goetghebuer T, Smolen KK, Adler C, Das J, McBride T, Smits G, Lecomte S, Haelterman E, Barlow P, Piedra PA, van der Klis F, Kollmann TR, Lauffenburger DA, Alter G, Levy J, Marchant A. Initiation of Antiretroviral Therapy Before Pregnancy Reduces the Risk of Infection-related Hospitalization in Human Immunodeficiency Virus-exposed Uninfected Infants Born in a High-income Country. *Clin Infect Dis*. 2019;68(7):1193-203. Epub 2018/09/15. doi: 10.1093/cid/ciy673. PubMed PMID: 30215689.

25. Mazzola TN, da Silva MT, Abramczuk BM, Moreno YM, Lima SC, Zorzeto TQ, Passeto AS, Vilela MM. Impaired *Bacillus Calmette-Guerin* cellular immune response in HIV-exposed, uninfected infants. *AIDS*. 2011;25(17):2079-87. Epub 2011/08/26. doi: 10.1097/QAD.0b013e32834bba0a. PubMed PMID: 21866040.

26. Abu-Raya B, Kollmann TR, Marchant A, MacGillivray DM. The Immune System of HIV-Exposed Uninfected Infants. *Front Immunol*. 2016;7:383. Epub 2016/10/14. doi: 10.3389/fimmu.2016.00383. PubMed PMID: 27733852; PMCID: PMC5039172.

27. Dzanibe S, Lennard K, Kiravu A, Seabrook MSS, Alinde B, Holmes SP, Blish CA, Jaspan HB, Gray CM. Stereotypic Expansion of T Regulatory and Th17 Cells during Infancy Is Disrupted by HIV Exposure and Gut Epithelial Damage. *J Immunol*. 2022;208(1):27-37. Epub 2021/11/26. doi: 10.4049/jimmunol.2100503. PubMed PMID: 34819390; PMCID: PMC8702481.

28. Jones CI, Rose SL, Shutt A, Cairo C, Bourgeois NM, Charurat M, Sodora DL, Wood MP. Maternal HIV status skews transcriptomic response in infant cord blood monocytes exposed to *Bacillus Calmette-Guerin*. *AIDS*. 2021;35(1):23-32. Epub 2020/10/14. doi: 10.1097/QAD.0000000000002706. PubMed PMID: 33048873; PMCID: PMC7718394.

29. Miles DJ, Gadama L, Gumbi A, Nyalo F, Makanani B, Heyderman RS. Human immunodeficiency virus (HIV) infection during pregnancy induces CD4 T-cell differentiation and modulates responses to *Bacille Calmette-Guerin* (BCG) vaccine in HIV-uninfected infants. *Immunology*. 2010;129(3):446-54. Epub 2009/12/17. doi: 10.1111/j.1365-2567.2009.03186.x. PubMed PMID: 20002789; PMCID: PMC2826689.

30. Tchakoute CT, Sainani KL, Osawe S, Datong P, Kiravu A, Rosenthal KL, Gray CM, Cameron DW, Abimiku A, Jaspan HB, team Is. Breastfeeding mitigates the effects of maternal HIV on infant infectious morbidity in the Option B+ era. *AIDS*. 2018;32(16):2383-91. Epub 2018/08/23. doi: 10.1097/QAD.0000000000001974. PubMed PMID: 30134300.

31. Eckard AR, Kirk SE, Hagood NL. Contemporary Issues in Pregnancy (and Offspring) in the Current HIV Era. *Curr HIV/AIDS Rep*. 2019;16(6):492-500. Epub 2019/10/21. doi: 10.1007/s11904-019-00465-2. PubMed PMID: 31630334; PMCID: PMC6938215.

32. Ikumi NM, Matjila M, Gray CM, Anumba D, Pillay K. Placental pathology in women with HIV. *Placenta*. 2021;115:27-36. Epub 2021/09/20. doi: 10.1016/j.placenta.2021.09.006. PubMed PMID: 34537469.

33. Ikumi NM, Pillay K, Tilburgs T, Malaba TR, Dzanibe S, Enninga EAL, Chakraborty R, Lamorde M, Myer L, Khoo S, Jaspan HB, Gray CM. T-Cell Homeostatic Imbalance in Placentas From Women With Human Immunodeficiency Virus in the Absence of Vertical Transmission. *J Infect Dis*. 2021;224(12 Suppl 2):S670-S82. Epub 2021/04/22. doi: 10.1093/infdis/jiab192. PubMed PMID: 33880544; PMCID: PMC8883807.

34. Yampolsky M, Shlakhter O, Deng D, Kala S, Walmsley SL, Murphy KE, Yudin MH, MacGillivray J, Loutfy M, Dunk C, Serghides L. Exploring the impact of HIV infection and antiretroviral therapy on placenta morphology. *Placenta*. 2021;104:102-9. Epub 2020/12/15. doi: 10.1016/j.placenta.2020.12.004. PubMed PMID: 33310298.

35. Ikumi NM, Malaba TR, Pillay K, Cohen MC, Madlala HP, Matjila M, Anumba D, Myer L, Newell ML, Gray CM, Group PS. Differential impact of antiretroviral therapy initiated before or during pregnancy on placenta pathology in HIV-positive women. *AIDS*. 2021;35(5):717-26. Epub 2021/03/17. doi: 10.1097/QAD.0000000000002824. PubMed PMID: 33724257; PMCID: PMC8630811.
36. Richter E, Bornemann L, Korencak M, Alter G, Schuster M, Esser S, Boesecke C, Rockstroh J, Gunzer M, Streeck H. Reduction of CD8 T Cell Functionality but Not Inhibitory Capacity by Integrase Inhibitors. *J Virol*. 2022;96(5):e0173021. Epub 2022/01/13. doi: 10.1128/JVI.01730-21. PubMed PMID: 35019724; PMCID: PMC8906395.
37. Schoeman JC, Moutloatse GP, Harms AC, Vreeken RJ, Scherpbier HJ, Van Leeuwen L, Kuijpers TW, Reinecke CJ, Berger R, Hankemeier T, Bunders MJ. Fetal Metabolic Stress Disrupts Immune Homeostasis and Induces Proinflammatory Responses in Human Immunodeficiency Virus Type 1- and Combination Antiretroviral Therapy-Exposed Infants. *J Infect Dis*. 2017;216(4):436-46. Epub 2017/06/22. doi: 10.1093/infdis/jix291. PubMed PMID: 28633455; PMCID: PMC5853663.
38. Nolan JP. The evolution of spectral flow cytometry. *Cytometry A*. 2022;101(10):812-7. Epub 2022/05/15. doi: 10.1002/cyto.a.24566. PubMed PMID: 35567367.
39. Futamura K, Sekino M, Hata A, Ikebuchi R, Nakanishi Y, Egawa G, Kabashima K, Watanabe T, Furuki M, Tomura M. Novel full-spectral flow cytometry with multiple spectrally-adjacent fluorescent proteins and fluorochromes and visualization of in vivo cellular movement. *Cytometry A*. 2015;87(9):830-42. Epub 2015/07/29. doi: 10.1002/cyto.a.22725. PubMed PMID: 26217952; PMCID: PMC5132038.
40. Cossarizza A, Chang HD, Radbruch A, Abrignani S, Addo R, Akdis M, Andra I, Andreati F, Annunziato F, Arranz E, Bacher P, Bari S, Barnaba V, Barros-Martins J, Baumjohann D, Beccaria CG, Bernardo D, Boardman DA, Borger J, Bottcher C, Brockmann L, Burns M, Busch DH, Cameron G, Cammarata I, Cassotta A, Chang Y, Chirido FG, Christakou E, Cicin-Sain L, Cook L, Corbett AJ, Cornelis R, Cosmi L, Davey MS, De Biasi S, De Simone G, Del Zotto G, Delacher M, Di Rosa F, Di Santo J, Diefenbach A, Dong J, Dorner T, Dress RJ, Dutertre CA, Eckle SBG, Eede P, Evrard M, Falk CS, Feuerer M, Fillatreau S, Fiz-Lopez A, Follo M, Foulds GA, Frobel J, Gagliani N, Galletti G, Gangaev A, Garbi N, Garrote JA, Geginat J, Gherardin NA, Gibellini L, Ginhoux F, Godfrey DI, Gruarin P, Haftmann C, Hansmann L, Harpur CM, Hayday AC, Heine G, Hernandez DC, Herrmann M, Hoelsken O, Huang Q, Huber S, Huber JE, Huehn J, Hundemer M, Hwang WYK, Iannacone M, Ivison SM, Jack HM, Jani PK, Keller B, Kessler N, Ketelaars S, Knop L, Knopf J, Koay HF, Kobow K, Kriegsmann K, Kristyanto H, Krueger A, Kuehne JF, Kunze-Schumacher H, Kvistborg P, Kwok I, Latorre D, Lenz D, Levings MK, Lino AC, Liotta F, Long HM, Lugli E, MacDonald KN, Maggi L, Maini MK, Mair F, Manta C, Manz RA, Mashregi MF, Mazzoni A, McCluskey J, Mei HE, Melchers F, Melzer S, Mielenz D, Monin L, Moretta L, Multhoff G, Munoz LE, Munoz-Ruiz M, Muscate F, Natalini A, Neumann K, Ng LG, Niedobitek A, Niemz J, Almeida LN, Notarbartolo S, Ostendorf L, Pallett LJ, Patel AA, Percin GI, Peruzzi G, Pinti M, Pockley AG, Pracht K, Prinz I, Pujol-Autonell I, Pulvirenti N, Quatrini L, Quinn KM, Radbruch H, Rhys H, Rodrigo MB, Romagnani C, Saggau C, Sakaguchi S, Sallusto F, Sanderink L, Sandrock I, Schauer C, Scheffold A, Scherer HU, Schiemann M, Schildberg FA, Schober K, Schoen J, Schuh W, Schuler T, Schulz AR, Schulz S, Schulze J, Simonetti S, Singh J, Sitnik KM, Stark R, Starossom S, Stehle C, Szelinski F, Tan L, Tarnok A, Tornack J, Tree TIM, van Beek JJP, van de Veen W, van Gisbergen K, Vasco C, Verheyden NA, von Borstel A, Ward-Hartstonge KA, Warnatz K, Waskow C, Wiedemann A, Wilharm A, Wing J, Wirz O, Wittner J, Yang JHM, Yang J. Guidelines for the use of flow cytometry and cell sorting in immunological studies (third edition). *Eur J Immunol*. 2021;51(12):2708-3145. Epub 2021/12/16. doi: 10.1002/eji.202170126. PubMed PMID: 34910301; PMCID: PMC11115438.

41. Roederer M. Compensation in flow cytometry. *Curr Protoc Cytom.* 2002;Chapter 1:Unit 1 14. Epub 2008/09/05. doi: 10.1002/0471142956.cy0114s22. PubMed PMID: 18770762.
42. Kmet R, Novo D. Reducing Spreading: Removing the Impact of Irrelevant Dyes Improves Unmixed Flow Cytometry Data. *Cytometry A.* 2025;107(9):573-86. Epub 2025/08/22. doi: 10.1002/cyto.a.24957. PubMed PMID: 40843630.
43. Kmet R, Novo D. Matrix Recovery Algorithm for Reconstructing Mixing Matrices From Raw Observations and Ordinary Least Squares Unmixed Abundance Values. *Cytometry A.* 2026;109(1):80-4. Epub 2026/01/05. doi: 10.1002/cyto.a.70005. PubMed PMID: 41486099.
44. Novo D. A comparison of spectral unmixing to conventional compensation for the calculation of fluorochrome abundances from flow cytometric data. *Cytometry A.* 2022;101(11):885-91. Epub 2022/07/17. doi: 10.1002/cyto.a.24669. PubMed PMID: 35841160.
45. Novo D, Gregori G, Rajwa B. Generalized unmixing model for multispectral flow cytometry utilizing nonsquare compensation matrices. *Cytometry A.* 2013;83(5):508-20. Epub 2013/03/26. doi: 10.1002/cyto.a.22272. PubMed PMID: 23526804; PMCID: PMC4177998.
46. Roederer M. Spectral compensation for flow cytometry: visualization artifacts, limitations, and caveats. *Cytometry.* 2001;45(3):194-205. Epub 2001/12/18. doi: 10.1002/1097-0320(20011101)45:3<194::aid-cyto1163>3.0.co;2-c. PubMed PMID: 11746088.
47. Nolan JP, Condello D. Spectral flow cytometry. *Curr Protoc Cytom.* 2013;Chapter 1:1 27 1-1 13. Epub 2013/01/08. doi: 10.1002/0471142956.cy0127s63. PubMed PMID: 23292705; PMCID: PMC3556726.
48. Nampota-Nkomba N, Buchwald A, Nyirenda OM, Mkandawire FA, Masonga R, Meja S, Moyo D, Cairo C, Laufer MK, Team RS. Association between Maternal HIV and Adverse Birth Outcomes in the Era of Universal Antiretroviral Therapy in Malawi. *J Acquir Immune Defic Syndr.* 2025. Epub 2025/04/17. doi: 10.1097/QAI.0000000000003685. PubMed PMID: 40243285; PMCID: PMC12353428.
49. Hsu H, Boudova S, Mvula G, Divala TH, Mungwira RG, Harman C, Laufer MK, Pauza CD, Cairo C. Prolonged PD1 Expression on Neonatal Vdelta2 Lymphocytes Dampens Proinflammatory Responses: Role of Epigenetic Regulation. *J Immunol.* 2016;197(5):1884-92. Epub 2016/07/31. doi: 10.4049/jimmunol.1600284. PubMed PMID: 27474072; PMCID: PMC4992653.
50. Hsu H, Boudova S, Mvula G, Divala TH, Rach D, Mungwira RG, Boldrin F, Degiacomi G, Manganelli R, Laufer MK, Cairo C. Age-related changes in PD-1 expression coincide with increased cytotoxic potential in Vdelta2 T cells during infancy. *Cell Immunol.* 2021;359:104244. Epub 2020/11/29. doi: 10.1016/j.cellimm.2020.104244. PubMed PMID: 33248366; PMCID: PMC7811364.
51. Montoya CJ, Pollard D, Martinson J, Kumari K, Wasserfall C, Mulder CB, Rugeles MT, Atkinson MA, Landay AL, Wilson SB. Characterization of human invariant natural killer T subsets in health and disease using a novel invariant natural killer T cell-clonotypic monoclonal antibody, 6B11. *Immunology.* 2007;122(1):1-14. Epub 2007/07/31. doi: 10.1111/j.1365-2567.2007.02647.x. PubMed PMID: 17662044; PMCID: PMC2265989.
52. den Braanker H, Bongenaar M, Lubberts E. How to Prepare Spectral Flow Cytometry Datasets for High Dimensional Data Analysis: A Practical Workflow. *Front Immunol.* 2021;12:768113. Epub 2021/12/07. doi: 10.3389/fimmu.2021.768113. PubMed PMID: 34868024; PMCID: PMC8640183.
53. Ferrer-Font L, Small SJ, Lewer B, Pilkington KR, Johnston LK, Park LM, Lannigan J, Jaimes MC, Price KM. Panel Optimization for High-Dimensional Immunophenotyping Assays Using Full-Spectrum Flow Cytometry. *Curr Protoc.* 2021;1(9):e222. Epub 2021/09/08. doi: 10.1002/cpz1.222. PubMed PMID: 34492732.
54. Herzenberg LA, Tung J, Moore WA, Herzenberg LA, Parks DR. Interpreting flow cytometry data: a guide for the perplexed. *Nat Immunol.* 2006;7(7):681-5. Epub 2006/06/21. doi: 10.1038/ni0706-681. PubMed PMID: 16785881.

55. Liechti T, Weber LM, Ashhurst TM, Stanley N, Prlic M, Van Gassen S, Mair F. An updated guide for the perplexed: cytometry in the high-dimensional era. *Nat Immunol.* 2021;22(10):1190-7. Epub 2021/09/08. doi: 10.1038/s41590-021-01006-z. PubMed PMID: 34489590.
56. Hsu H, Zanettini C, Coker M, Boudova S, Rach D, Mvula G, Divala TH, Mungwira RG, Boldrin F, Degiacomi G, Mazzabo LC, Manganelli R, Laufer MK, Zhang Y, Marchionni L, Cairo C. Concomitant assessment of PD-1 and CD56 expression identifies subsets of resting cord blood Vdelta2 T cells with disparate cytotoxic potential. *Cell Immunol.* 2024;395-396:104797. Epub 2024/01/02. doi: 10.1016/j.cellimm.2023.104797. PubMed PMID: 38157646.
57. Davey MS, Willcox CR, Hunter S, Kasatskaya SA, Remmerswaal EBM, Salim M, Mohammed F, Bemelman FJ, Chudakov DM, Oo YH, Willcox BE. The human Vdelta2(+) T-cell compartment comprises distinct innate-like Vgamma9(+) and adaptive Vgamma9(-) subsets. *Nat Commun.* 2018;9(1):1760. Epub 2018/05/04. doi: 10.1038/s41467-018-04076-0. PubMed PMID: 29720665; PMCID: PMC5932074.
58. Finak G, Jiang W, Gottardo R. CytoML for cross-platform cytometry data sharing. *Cytometry A.* 2018;93(12):1189-96. Epub 2018/12/15. doi: 10.1002/cyto.a.23663. PubMed PMID: 30551257; PMCID: PMC6443375.
59. Emmaneel A, Quintelier K, Sichien D, Rybakowska P, Maranon C, Alarcon-Riquelme ME, Van Isterdael G, Van Gassen S, Saeys Y. PeacoQC: Peak-based selection of high quality cytometry data. *Cytometry A.* 2022;101(4):325-38. Epub 2021/09/23. doi: 10.1002/cyto.a.24501. PubMed PMID: 34549881; PMCID: PMC9293479.
60. Pedersen CB, Dam SH, Barnkob MB, Leipold MD, Purroy N, Rassenti LZ, Kipps TJ, Nguyen J, Lederer JA, Gohil SH, Wu CJ, Olsen LR. cyCombine allows for robust integration of single-cell cytometry datasets within and across technologies. *Nat Commun.* 2022;13(1):1698. Epub 2022/04/02. doi: 10.1038/s41467-022-29383-5. PubMed PMID: 35361793; PMCID: PMC8971492 Pharmacyclics. The remaining authors declare no competing interests.
61. Quintelier KLA, Willemsen M, Bosteels V, Aerts J, Saeys Y, Van Gassen S. CytoNorm 2.0: A flexible normalization framework for cytometry data without requiring dedicated controls. *Cytometry A.* 2025;107(2):69-87. Epub 2025/01/28. doi: 10.1002/cyto.a.24910. PubMed PMID: 39871681.
62. Van Gassen S, Gaudilliere B, Angst MS, Saeys Y, Aghaeepour N. CytoNorm: A Normalization Algorithm for Cytometry Data. *Cytometry A.* 2020;97(3):268-78. Epub 2019/10/22. doi: 10.1002/cyto.a.23904. PubMed PMID: 31633883; PMCID: PMC7078957.
63. Finak G, Frelinger J, Jiang W, Newell EW, Ramey J, Davis MM, Kalams SA, De Rosa SC, Gottardo R. OpenCyto: an open source infrastructure for scalable, robust, reproducible, and automated, end-to-end flow cytometry data analysis. *PLoS Comput Biol.* 2014;10(8):e1003806. Epub 2014/08/29. doi: 10.1371/journal.pcbi.1003806. PubMed PMID: 25167361; PMCID: PMC4148203.
64. Wang YF, Huang HY, Rudin C, Shaposhnik Y. Understanding How Dimension Reduction Tools Work: An Empirical Approach to Deciphering t-SNE, UMAP, TriMap, and PaCMAP for Data Visualization. *J Mach Learn Res.* 2021;22. PubMed PMID: WOS:000706883000001.
65. Belkina AC, Ciccolella CO, Anno R, Halpert R, Spidlen J, Snyder-Cappione JE. Automated optimized parameters for T-distributed stochastic neighbor embedding improve visualization and analysis of large datasets. *Nat Commun.* 2019;10(1):5415. Epub 2019/11/30. doi: 10.1038/s41467-019-13055-y. PubMed PMID: 31780669; PMCID: PMC6882880 Dickinson (BD); FlowJo is a subsidiary of BD. The remaining authors declare no competing interests.
66. Kobak D, Linderman GC. Initialization is critical for preserving global data structure in both t-SNE and UMAP. *Nat Biotechnol.* 2021;39(2):156-7. Epub 2021/02/03. doi: 10.1038/s41587-020-00809-z. PubMed PMID: 33526945.

67. Moon KR, van Dijk D, Wang Z, Gigante S, Burkhardt DB, Chen WS, Yim K, Elzen AVD, Hirn MJ, Coifman RR, Ivanova NB, Wolf G, Krishnaswamy S. Visualizing structure and transitions in high-dimensional biological data. *Nat Biotechnol.* 2019;37(12):1482-92. Epub 2019/12/05. doi: 10.1038/s41587-019-0336-3. PubMed PMID: 31796933; PMCID: PMC7073148.
68. Wickham H. Tidy Data. *Journal of Statistical Software.* 2014;59(10):1 - 23. doi: 10.18637/jss.v059.i10.
69. Seiler C, Ferreira AM, Kronstad LM, Simpson LJ, Le Gars M, Vendrame E, Blish CA, Holmes S. CytoGLMM: conditional differential analysis for flow and mass cytometry experiments. *BMC Bioinformatics.* 2021;22(1):137. Epub 2021/03/24. doi: 10.1186/s12859-021-04067-x. PubMed PMID: 33752595; PMCID: PMC7983283.
70. Eckle SB, Corbett AJ, Keller AN, Chen Z, Godfrey DI, Liu L, Mak JY, Fairlie DP, Rossjohn J, McCluskey J. Recognition of Vitamin B Precursors and Byproducts by Mucosal Associated Invariant T Cells. *J Biol Chem.* 2015;290(51):30204-11. Epub 2015/10/16. doi: 10.1074/jbc.R115.685990. PubMed PMID: 26468291; PMCID: PMC4683245.
71. Ferrer-Font L, Small SJ, Hyde E, Pilkington KR, Price KM. Panel Design and Optimization for Full Spectrum Flow Cytometry. *Methods Mol Biol.* 2024;2779:99-124. Epub 2024/03/25. doi: 10.1007/978-1-0716-3738-8_6. PubMed PMID: 38526784.
72. Jensen IJ, McGonagill PW, Lefebvre MN, Griffith TS, Harty JT, Badovinac VP. Worry and FRET: ROS Production Leads to Fluorochrome Tandem Degradation and impairs Interpretation of Flow Cytometric Results. *Immunity.* 2020;52(3):419-21. Epub 2020/03/19. doi: 10.1016/j.immuni.2020.02.003. PubMed PMID: 32187510; PMCID: PMC7251922.
73. Kalina T, Lundsten K, Engel P. Relevance of Antibody Validation for Flow Cytometry. *Cytometry A.* 2020;97(2):126-36. Epub 2019/10/03. doi: 10.1002/cyto.a.23895. PubMed PMID: 31577065.
74. Roederer M. Distributions of autofluorescence after compensation: Be panglossian, fret not. *Cytometry A.* 2016;89(4):398-402. Epub 2016/02/10. doi: 10.1002/cyto.a.22820. PubMed PMID: 26859757.
75. Jameson VJ, Luke T, Yan Y, Hind A, Evrard M, Man K, Mackay LK, Kallies A, Villadangos JA, McWilliam HEG, Perez-Gonzalez A. Unlocking autofluorescence in the era of full spectrum analysis: Implications for immunophenotype discovery projects. *Cytometry A.* 2022;101(11):922-41. Epub 2022/03/30. doi: 10.1002/cyto.a.24555. PubMed PMID: 35349225; PMCID: PMC9519814.
76. Roet JEG, Mikula AM, de Kok M, Chadick CH, Garcia Vallejo JJ, Roest HP, van der Laan LJW, de Winde CM, Mebius RE. Unbiased method for spectral analysis of cells with great diversity of autofluorescence spectra. *Cytometry A.* 2024;105(8):595-606. Epub 2024/06/12. doi: 10.1002/cyto.a.24856. PubMed PMID: 38863410.
77. Mage PL, Konecny AJ, Mair F. Measurement and prediction of unmixing-dependent spreading in spectral flow cytometry panels. 2025.
78. Fletez-Brant K, Spidlen J, Brinkman RR, Roederer M, Chattopadhyay PK. flowClean: Automated identification and removal of fluorescence anomalies in flow cytometry data. *Cytometry A.* 2016;89(5):461-71. Epub 2016/03/19. doi: 10.1002/cyto.a.22837. PubMed PMID: 26990501; PMCID: PMC5522377.
79. Meskas J, Yokosawa D, Wang S, Segat GC, Brinkman RR. flowCut: An R package for automated removal of outlier events and flagging of files based on time versus fluorescence analysis. *Cytometry A.* 2023;103(1):71-81. Epub 2022/07/08. doi: 10.1002/cyto.a.24670. PubMed PMID: 35796000; PMCID: PMC9823164.
80. Monaco G, Chen H, Poidinger M, Chen J, de Magalhaes JP, Larbi A. flowAI: automatic and interactive anomaly discerning tools for flow cytometry data. *Bioinformatics.* 2016;32(16):2473-80. Epub 2016/05/07. doi: 10.1093/bioinformatics/btw191. PubMed PMID: 27153628.

81. Seamer L. Data file standard for flow cytometry, FCS 3.0. *Curr Protoc Cytom.* 2001;Chapter 10:Unit 10.2. Epub 2008/09/05. doi: 10.1007/978-3-642-57049-0_6. PubMed PMID: 18770676.
82. Spidlen J, Moore W, Parks D, Goldberg M, Blenman K, Cavanaugh JS, Force IDST, Brinkman R. Data File Standard for Flow Cytometry, Version FCS 3.2. *Cytometry A.* 2021;99(1):100-2. Epub 2020/09/04. doi: 10.1002/cyto.a.24225. PubMed PMID: 32881398; PMCID: PMC8241566.
83. Spidlen J, Moore W, Parks D, Goldberg M, Bray C, Bierre P, Gorombey P, Hyun B, Hubbard M, Lange S, Lefebvre R, Leif R, Novo D, Ostruszka L, Treister A, Wood J, Murphy RF, Roederer M, Sudar D, Zigon R, Brinkman RR. Data File Standard for Flow Cytometry, version FCS 3.1. *Cytometry A.* 2010;77(1):97-100. Epub 2009/11/26. doi: 10.1002/cyto.a.20825. PubMed PMID: 19937951; PMCID: PMC2892967.
84. Orlova DY, Herzenberg LA, Walther G. Science not art: statistically sound methods for identifying subsets in multi-dimensional flow and mass cytometry data sets. *Nat Rev Immunol.* 2017;18(1):77. Epub 2017/12/23. doi: 10.1038/nri.2017.150. PubMed PMID: 29269766.
85. Saeys Y, Van Gassen S, Lambrecht B. Response to Orlova et al. "Science not art: statistically sound methods for identifying subsets in multi-dimensional flow and mass cytometry data sets". *Nat Rev Immunol.* 2017;18(1):78. Epub 2017/12/23. doi: 10.1038/nri.2017.151. PubMed PMID: 29269765.
86. Saeys Y, Van Gassen S, Lambrecht BN. Computational flow cytometry: helping to make sense of high-dimensional immunology data. *Nature Reviews Immunology.* 2016;16(7):449-62. doi: 10.1038/nri.2016.56.
87. Mair F, Liechti T. Comprehensive Phenotyping of Human Dendritic Cells and Monocytes. *Cytometry A.* 2021;99(3):231-42. Epub 2020/11/18. doi: 10.1002/cyto.a.24269. PubMed PMID: 33200508.
88. Quintelier K, Couckuyt A, Emmaneel A, Aerts J, Saeys Y, Van Gassen S. Analyzing high-dimensional cytometry data using FlowSOM. *Nat Protoc.* 2021;16(8):3775-801. Epub 2021/06/27. doi: 10.1038/s41596-021-00550-0. PubMed PMID: 34172973.
89. Van Gassen S, Callebaut B, Van Helden MJ, Lambrecht BN, Demeester P, Dhaene T, Saeys Y. FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data. *Cytometry A.* 2015;87(7):636-45. Epub 2015/01/13. doi: 10.1002/cyto.a.22625. PubMed PMID: 25573116.
90. Levine Jacob H, Simonds Erin F, Bendall Sean C, Davis Kara L, Amir E-ad D, Tadmor Michelle D, Litvin O, Fienberg Harris G, Jager A, Zunder Eli R, Finck R, Gedman Amanda L, Radtke I, Downing James R, Pe'er D, Nolan Garry P. Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. *Cell.* 2015;162(1):184-97. doi: 10.1016/j.cell.2015.05.047.
91. Samusik N, Good Z, Spitzer MH, Davis KL, Nolan GP. Automated mapping of phenotype space with single-cell data. *Nature Methods.* 2016;13(6):493-6. doi: 10.1038/nmeth.3863.
92. Carey VJ. Bioconductor: Planning a third decade of comprehensive support for genomic data science. *Patterns (N Y).* 2025;6(7):101319. Epub 2025/09/10. doi: 10.1016/j.patter.2025.101319. PubMed PMID: 40926969; PMCID: PMC12416078.
93. Weber LM, Robinson MD. Comparison of clustering methods for high-dimensional single-cell flow and mass cytometry data. *Cytometry A.* 2016;89(12):1084-96. Epub 2016/12/20. doi: 10.1002/cyto.a.23030. PubMed PMID: 27992111.
94. Chari T, Pachter L. The specious art of single-cell genomics. *PLoS Comput Biol.* 2023;19(8):e1011288. Epub 2023/08/17. doi: 10.1371/journal.pcbi.1011288. PubMed PMID: 37590228; PMCID: PMC10434946.
95. Montante S, Yokosawa D, Li L, Butyaev A, Malek M, Movassaghi R, Michalchuk Q, Hsu C-TJ, Shmil D, Rahim A, Cossarizza A, Esteban JB, Erhart K, Finnbogason B, Kelion G,

- Leifsson H, Rivers J, Ecker D, Szantner A, Waldspühl J, Brinkman RR. Citizen science gamers enable automated flow cytometry gating through machine learning. *bioRxiv*. 2025:2025.10.07.679685. doi: 10.1101/2025.10.07.679685.
96. Moore WA, Meehan SW, Meehan C, Parks DR, Walther G, Herzenberg LA. Automatic phenotyping using exhaustive projection pursuit. *Commun Biol*. 2025;8(1):1207. Epub 2025/08/13. doi: 10.1038/s42003-025-08581-z. PubMed PMID: 40797028; PMCID: PMC12343891.
97. Moens E, Brouwer M, Dimova T, Goldman M, Willems F, Vermijlen D. IL-23R and TCR signaling drives the generation of neonatal Vgamma9Vdelta2 T cells expressing high levels of cytotoxic mediators and producing IFN-gamma and IL-17. *J Leukoc Biol*. 2011;89(5):743-52. Epub 2011/02/19. doi: 10.1189/jlb.0910501. PubMed PMID: 21330350.
98. Patel O, Kjer-Nielsen L, Le Nours J, Eckle SB, Birkinshaw R, Beddoe T, Corbett AJ, Liu L, Miles JJ, Meehan B, Reantragoon R, Sandoval-Romero ML, Sullivan LC, Brooks AG, Chen Z, Fairlie DP, McCluskey J, Rossjohn J. Recognition of vitamin B metabolites by mucosal-associated invariant T cells. *Nat Commun*. 2013;4:2142. Epub 2013/07/13. doi: 10.1038/ncomms3142. PubMed PMID: 23846752.
99. Tan L, Fichtner AS, Bruni E, Odak I, Sandrock I, Bubke A, Borchers A, Schultze-Florey C, Koenecke C, Forster R, Jarek M, von Kaisenberg C, Schulz A, Chu X, Zhang B, Li Y, Panzer U, Krebs CF, Ravens S, Prinz I. A fetal wave of human type 3 effector gammadelta cells with restricted TCR diversity persists into adulthood. *Sci Immunol*. 2021;6(58). Epub 2021/04/25. doi: 10.1126/sciimmunol.abf0125. PubMed PMID: 33893173.
100. Swarbrick GM, Gela A, Cansler ME, Null MD, Duncan RB, Nemes E, Shey M, Nsereko M, Mayanja-Kizza H, Kiguli S, Koh J, Hanekom WA, Hatherill M, Lancioni C, Lewinsohn DM, Scriba TJ, Lewinsohn DA. Postnatal Expansion, Maturation, and Functionality of MR1T Cells in Humans. *Front Immunol*. 2020;11:556695. Epub 2020/10/13. doi: 10.3389/fimmu.2020.556695. PubMed PMID: 33042140; PMCID: PMC7524872.