Spatial Analysis of High Parameter Imaging Data to Understand Early HIV-Cell Interactions in Colorectal Tissue

Nicolas Paolo Canete

A thesis submitted to fulfil requirements for the degree of Master of Philosophy

Supervisors:
Prof. Andrew Harman & Dr. Ellis Patrick

The University of Sydney
Faculty of Medicine and Health
Westmead Clinical School

Centre for Virus Research
The Westmead Institute for Medical Research

2023

Declaration

This is to certify that to the best of my knowledge, the content of this thesis is my own work. This thesis has not been submitted for any degree or other purposes.

I certify that the intellectual content of this thesis is the product of my own work and that all the assistance received in preparing this thesis and sources have been acknowledged.

The following publications/manuscripts are contained within this thesis:

- Baharlou, H.*, **Canete, N.P.***, Cunningham, A.L., Harman, A.N, Patrick, E. (2019) Mass Cytometry Imaging for the Study of Human Diseases - Applications and Data Analysis Strategies.
- Baharlou, H.*, **Canete, N.P.***, Bertram K.M., Sandgren, K.J., Cunningham, A.L., Harman, A.N, Patrick, E. (2020) AFid: a tool for automated identification and exclusion of autofluorescent objects from microscopy images *Bioinformatics, Volume 37, Issue 4, 15 February 2021, Pages 559–567*
- **Canete, N.P.**, Iyengar, S.S., Ormerod, J.T., Baharlou, H., Harman, A.N., Patrick, E. (2022) spicyR: spatial analysis of in situ cytometry data in R. *Bioinformatics, Volume 38, Issue 11, 1 June 2022, Pages 3099–3105*
- Baharlou, H, **Canete, N.P** Vine, E.E.. et al. (2022) An *in situ* analysis pipeline for initial host-pathogen interactions reveals signatures of human colorectal HIV transmission. *Cell Rep. 2022 Sep 20;40(12):111385.*

*Co-first authors

Nicolas Paolo Canete

May 29, 2023

Authorship Attribution Statement

Chapter 1 of this thesis is adapted from a review published in Frontiers in Immunology (Baharlou, Canete et al. 2019; full citation above), in which I am a co-first author. While contributing to the whole review, my primary contribution was reviewing the analysis techniques that are available in high parameter imaging.

Chapter 2 of this thesis consists of a publication in Bioinformatics (Baharlou, Canete et al. 2020), in which I am co-first author. H. Baharlou conceived the project, collected data and wrote the first implementation of the algorithm. I made several intellectual contributions to the project, wrote the codebase in ImageJ, MATLAB and R, and performed the data analysis.

Chapter 3 of this thesis consists of a publication in Bioinformatics (Canete et al. 2022), in which I am the first author. I conceived the project, wrote the package codebase and performed the analysis. S. Iyengar created the shiny package.

Chapter 4 of this thesis consists of a publication in Cell Reports (Baharlou et al. 2022), in which I am second author. H. Baharlou conceived and guided the project. Experimental investigation was largely performed by H. Baharlou and E.E. Vine. Computational investigation was largely performed by myself and H. Baharlou.

In addition to the statements above, in cases where I am not the corresponding author of a published item, permission to include the published material has been granted by the corresponding author.

Nicolas Paolo Canete

May 29, 2023

As supervisor for the candidature upon which this thesis is based, I can confirm that the authorship attribution statements above are correct.

Andrew Harman

May 29, 2023

Abstract

Although Human Immunodeficiency Virus (HIV) was first discovered over 30 years ago, there remains no cure or vaccine for HIV. In Australia, HIV transmission almost exclusively occurs via sexual intercourse, with over 75% occurring as a result of anal intercourse. Due to technological limitations, there are still major gaps in understanding how HIV is transmitted across the human colorectum, with the role of key HIV target cells being unclear. Additionally, traditional microscopy methods have not been able to visualize the full repertoire of HIV target cells within tissue.

To investigate cellular interactions during early HIV transmission, we topically infect fresh human colorectal tissue with HIV as a model of early HIV transmission *in situ*. We then make use of Cyclic Immunofluoresecence, a high parameter microscopy imaging approach, to visualise key HIV target cells, namely dendritic cells, macrophages and CD4+ T lymphocytes could be observed within the same tissue section. Furthermore, through the *in situ* hybridization technique RNAscope, we are able to visualize HIV with single-virion sensitivity.

Cyclic Immunofluorescence and RNAscope provide a unique opportunity for visualizing the early events of HIV transmission, allowing the interactions of HIV with its target cells as well as the colorectal tissue structure to be captured. However, given the recent development of these techniques, the imaging data being produced provide many unique challenges in terms of computational analysis. In particular, the presence of autofluorescence makes it difficult to reliably assign marker expression to cells while spatial analytical techniques are limited when multiple cell types are measured simultaneously.

This thesis presents three publications which address these analytical gaps, while also detailing the computational analysis techniques utilized in the analysis of the HIV imaging dataset. The first publication presents an algorithm which identifies and removes autofluorescence from fluorescent microscopy images, necessary for robustly identifying the HIV and their target cells within each image. The second publication presents a novel approach for performing inferences on changes in spatial co-localization across different experimental conditions. Finally, the third publication outlines the experimental innovations used to visualize early HIV transmission in human colorectal tissue. Additionally, this publication presents a complete image analysis pipeline used to characterize the cellular composition and interactions within human colorectal tissue, integrating the tools developed in the first two publications. This paper presents the main biological findings of this thesis, revealing for the first time the early events that underly HIV transmission.

Foreword

The purpose of this thesis is two-fold. Firstly, the overarching aim of this thesis will be to utilize the high parameter imaging technique cyclic immunofluorescence to understand the early transmission of HIV in colorectal tissue. The complexity of the images acquired has necessitated the need for image analysis strategies to quantify the cell-cell and cell-HIV relationships captured with microscopy. Hence, the second overarching aim of this thesis will be to develop computational strategies for analyzing this dataset, as well as high parameter images in general.

Chapter 1 of this thesis will provide a brief introduction to what is currently understood with regards to early HIV transmission in human colorectal tissue. The remainder of this chapter will be a review of different computational approaches and analysis strategies that have been developed for analyzing high parameter microscopy images. This section is largely based on a review of Mass Cytometry Imaging published in Frontiers in Immunology (Baharlou, Canete et al. 2019).

Chapters 2-4 consists of 3 publications. Chapter 2 presents a publication in Bioinformatics (Baharlou, Canete et al. 2020) which describes a novel post-acquisition algorithm for identifying and removing autofluorescence from autofluorescent microscopy images. Chapter 3 presents a publication in Bioinformatics (Canete et al. 2022) which describes an R package for performing inference on changes in spatial co-localization between cell types. Finally, Chapter 4 presents a publication in Cell Reports (Baharlou et al. 2022) which describes the full pipeline used for the analysis of the HIV CyCIF image dataset acquired by the host lab. My contributions to each project is outlined by a brief preface at the beginning of each chapter.

While Chapters 2-4 contain a standalone discussion section, Chapter 5 of this thesis will contextualize the three publications present, discuss the overall significance of this thesis, and discuss future work.

Publications and Presentations

Publications

- Baharlou, H.*, **Canete, N.P.***, Cunningham, A.L., Harman, A.N, Patrick, E. (2019) Mass Cytometry Imaging for the Study of Human Diseases - Applications and Data Analysis Strategies.
- Baharlou, H.*, **Canete, N.P.***, Bertram K.M., Sandgren, K.J., Cunningham, A.L., Harman, A.N, Patrick, E. (2020) AFid: a tool for automated identification and exclusion of autofluorescent objects from microscopy images *Bioinformatics, Volume 37, Issue 4, 15 February 2021, Pages 559–567*
- **Canete, N.P.**, Iyengar, S.S., Ormerod, J.T., Baharlou, H., Harman, A.N., Patrick, E. (2022) spicyR: spatial analysis of in situ cytometry data in R. *Bioinformatics, Volume 38, Issue 11, 1 June 2022, Pages 3099–3105*
- Baharlou, H, **Canete, N.P**, Vine, E.E.. et al. (2022) An *in situ* analysis pipeline for initial host-pathogen interactions reveals signatures of human colorectal HIV transmission. *Cell Rep. 2022 Sep 20;40(12):111385.*

*Co-first authors

Presentations

2019:

- Annual Australian Centre for HIV and Hepatitis Virology Research Conference 2019 (Canungra, QLD) - Oral Presentation
- Westmead Research Showcase 2019 (Sydney, NSW) - Lightning Talk
- Bioimage Informatics 2019 (Seattle, Washington, USA) - Poster Presentation
- Australian Bioinformatics and Computational Biology Society Conference 2019 (Sydney, NSW) - Poster, Lightning Talk

2020:

- Australian Bioinformatics and Computational Biology Society Conference 2020 (Virtual) - Online Poster

2021:

- Australian Bioinformatics and Computational Biology Society Conference 2021 (Virtual) - Oral Presentation

Acknowledgements

I'd like to thank my two amazing supervisors Andrew and Ellis. Ever since starting Honours back in 2018, you have both been incredibly supportive of me in this not-so-conventional project. Thank you both for always having my back and for always pushing me with this project. I have learned a lot from the both of you, both in terms of working in research and in terms of life, and I would not be the person I am now without both of your guidance and support.

Heeva, you have been my partner in crime throughout this project. I will always appreciate and treasure all of the random talks and adventures we've had.

Also a big thank you for everyone who has passed through the Patrick lab. In particular, a big thanks to Adam, Harry, Elijah, Alex, Farhan and Sourish for making 2022 a very exciting year.

I would like to shout out everyone on Level 6 at WIMR - I never would have expected to have worked with such brilliant and cool people! Tom - I love all our talks about movies, books, and anything science-y. Jacinta - thanks for indulging me whenever I wanted to talk about Star Wars and other super dorky things over tea whenever I was bored. Hafsa, Sana - thanks for sharing my love for R&B and hip-hop, and for putting up with my opinionated (but superior) music taste and uninvited rants about the music world. Kevin, Sam, Ashley - thanks for bringing K-Pop back into my life (for better or for worse). Katie - I'll always appreciate all the car rides we've had listening to Triple J. Vincent - thanks for the nerd talk and MTG playing over the last few years and for many years to come. Emily - thanks for sharing my dream of one day ruling Westmead together. Liz - thanks for being an absolutely amazing friend throughout the last year; all our random talks about anything and everything has really been a highlight of 2022 for me. And of course, a big shout out to Jason, Konrad, Shaan, Laurentius, Kevo (not at WIMR but I'm throwing you in here too!) - you guys are honestly the bestest of friends; thank you for indulging my weirdness and nerdiness, and for letting me be me.

Thanks also to all my mates, who probably have no idea what I've been doing these last few years. Thanks for being so flexible and understanding with me, and for being a constant reminder that there's more to life outside of science and medicine.

Finally, a big thank you to my family - Mum, Dad, Chris, Janella. Thank you for putting up with my busy schedule, and for helping pick up the slack at home!

Contents

List of Figures

List of Supplementary Figures

List of Tables

List of Supplementary Tables

Chapter 1

Introduction

Publications incorporated into this chapter:

Baharlou, H.*, **Canete, N.P.***, Cunningham, A.L., Harman, A.N, Patrick, E. (2019) Mass Cytometry Imaging for the Study of Human Diseases - Applications and Data Analysis Strategies. *Front. Immunol. 10:2657*

*Co-first authors

Preface

This introductory chapter is divided into three main parts. Firstly, we provide a brief summary of what is currently understood about early HIV transmission. This will contextualize Chapter 4, in which we apply various image analysis techniques developed throughout the thesis to understanding early HIV transmission in human colorectal tissue. This will serve as the primary biological question addressed within this thesis. Secondly, we provide a summary of high parameter microscopy imaging methods, including mass cytometry and inmunofluorescence techniques, as well as the computational analysis methods that have been implemented so far for the analysis of the image data obtained. This is largely adapted from a review published in Frontiers in Immunology (Baharlou, Canete et al. 2019), in which I was co-first author. While the review is mostly focused on mass cytometry-based imaging methods, the analysis techniques described here can be utilized in any high parameter imaging modalities. Here, we summarize the key points provided in the review. Finally, we provide a summary of the aims of this thesis.

1.1 Early HIV Transmission

1.1.1 Introduction

Human Immunodeficiency Virus (HIV) was first discovered 40 years ago and is the causative agent of acquired immunodeficiency syndrome (AIDS). There remains no cure or vaccine for HIV, with an estimated 27.5 million people living with HIV globally and 1.5 million new HIV infections in 2020 [1]. Therefore, a vaccine is still critically needed and, in the meantime, better prophylactic strategies. HIV transmission almost exclusively occurs via sexual intercourse, and in Australia over 75% of HIV infections occur as a result of anal intercourse [2, 3]. However, our understanding of the immunological mechanisms which facilitate the sexual transmission of HIV is still limited, especially the initial interactions between the virus and its target innate immune cells and how these cells go on to initiate and adaptive immune response. A better understanding of these events will be critical for the development of better prophylactic strategies and vaccines.

The key biological enquiry motivating this thesis is to define the initial events governing HIV transmission in human colorectal tissue (Chapter 4) and in developing better image analysis tools to visualize these processes.

1.1.2 HIV and HIV Transmission in Human Colorectal Tissue

HIV is a lentivirus that consists of two single-strands of RNA which encode the virus genome, enclosed by a conical capsid composed of the viral protein p24. This structure is surrounded by a lipid bilayer viral envelope. The viral genome consists of 9 genes encoding structural proteins, accessory proteins and enzymes necessary for viral replication.

HIV can infect cells that express its entry receptor CD4 and chemokine co-receptor, usually CCR5 or CXCR4 [4–6]. It is mainly R5 strains which utilize the CCR5 co-receptor that are involved in transmission. During persistent infection, the virus mutates to X4 stains which utilize the CXCR4 co-receptor to infect cells [7]. HIV entry into its target cells is mediated by the viral envelope (Env) protein, comprised of the glycoprotein subunit gp120 and transmembrane subunit gp41. (i) by binding to CD4 and CCR5, resulting in fusion of the virus envelope with the plasma membrane and delivery of the viral capsid into the cytosol, or (ii) by binding to surface lectin receptors such as DC-SIGN/CD209 [8–10], mannose receptor/CD206 [11], Langerin/CD207 [12], Siglec-1/CD169 [13], as well as surface integrins such as $\alpha 4\beta 7$ [14] and resulting in endocytic uptake that does not lead to infection. Following infection, reverse transcription occurs followed by integration the HIV genome into the host cell genome. This subsequently leads to HIV replication a new virion formation which go onto infect new cells either via cell-free spreading cell-cell viral transfer.

Within colorectal tissue, the primary HIV target cells are CD4+ T lymphocytes (CD4 T cells). The virus is also taken up by mononuclear phagocytes (MNPs) which deliver the virus to CD4 T cells in association with their antigen presenting cell (APC) function. MNPs can be divided into multiple subsets with distinct phenotypes and function and distinct roles in HIV transmission. Once CD4 T cells are infected, HIV undergoes rapid lytic replication and the onset of AIDS occurs once these cells have been depleted below critical levels. This section will briefly outline the specific roles of HIV target cells and the human colorectum in the sexual transmission of HIV.

HIV Transmission in Human Colorectum

The human colorectum comprises of the distal portion of the sigmoid colon and the rectum [15]. The inner mucosal surface of the colorectum is thin and fragile and hence a less effective barrier against HIV infection compared to other surfaces such as the anus, vagina and foreskin, which contain a robust stratified squamous epithelial layer [16].

The colorectal mucosa consists of a single columnar epithelium overlaying the lamina propria, a layer of connective tissue richly populated by HIV target cells. HIV has been observed to cross the epithelium into the lamina propria via several mechanisms: (i) via breaches in the epithelium due to mechanical abrasions or trauma, or chemical irritation [17, 18]; (ii) via transient openings of the tight-junctions between epithelial cells [19, 20]; (iii) via productive infection of epithelial cells [21, 22]; (iv) via viral uptake by specialised epithelial cells such as microfold cells which transcytose the virus into vesicles and transferring it across the epithelium [23]; or (v) via direct sampling by DCs in the lamina propria [19, 24]. These mechanisms provide an opportunity for HIV to encounter key HIV target cells within the lamina propria.

T Lymphocytes

CD4 T cells are responsible for cell-mediated immunity which regulates the immune response. These cells are positive for the T cell co-receptor CD3 as well as CD4. Naive T cells express antigen-specific receptors and reside in lymphoid tissue and in blood. These cells are activated via APCs, especially dendritic cells (DCs), resulting in clonal expansion and differentiation into effector and memory helper subsets [25]. Effector T cells subsequently produce cytokines that activate B cells, macrophages, and other immune cell types and mediate the immune response. Memory T cells are inactive, long-lived immune cells that can be activated to rapidly initiate a secondary immune response [26]. In activated CD4 T cells, HIV undergoes rapid lytic replication resulting in cell death but the virus can remain in a latent state in memory CD4 T cells.

While CD4 T cells are the key target cells of HIV, it is unclear whether they represent the initial target cells within colorectal mucosa. The high frequency of CD4+CCR5+ T cells in the gut mucosa may facilitate direct HIV infection of T cells [27]. T cells that highly express the gut homing receptor $\alpha 4\beta 7$ have been shown to be more susceptible to productive infection [14, 28]. In human colonic mucosa explants, Kolodkin et al. [29] observed that cell-associated virus is transmitted more efficiently than cell-free virus, with CD4 T cells on the mucosal surface able to uptake and produce virus for initiating infection. 48 hours post-exposure to HIV, the authors found that 90% of infected cells were CD4 T cells. Furthermore, in a macaque study of the analogous simian immunodeficiency virus (SIV), dos Santos et al. [30] observed the early accumulation of SIV in rectal lymphoid aggregates, densely packed structures of CD4 T cells surrounding an inner B cell zone. Specifically, SIV associated with CD4 T cells and a follicular DC subset. In another macaque model, Stieh et al. [31] showed the preferential infection of Th17 cells by SIV in vaginal tissue. Furthermore, McKinnon et al. [32] observed the depletion of Th17 cells in the cervix of HIV+ female sex workers within 48 hours, while Gosselin et al. [33] observed that memory Th17 contributed to the pool of cells harbouring HIV DNA in colon, despite being lower in number compared to other T cell subsets. Thus, Th17 cells are implicated in early HIV uptake.

Mononuclear Phagocytes

There exist four distinct cell types of mononuclear phagocytes (MNPs): dendritic cells (DCs), Langerhans cells (LCs), monocytes, and macrophages. These cells detect and uptake pathogens through an array of pattern recognition receptors, including C-type lectin receptors (CLRs), sialic acid-binding immunoglobulin-type of lectins (Siglec) and Toll-like receptors (TLRs). Langerhans cells are not found in colorectal tissue and will not be discussed further. DCs and macrophages are located in close proximity to the epithelial surfaces of colorectal tissue and are among the first cell types that encounter HIV.

Monocytes circulate in the bloodstream and migrate into tissue to replenish tissue resident macrophages in response to inflammatory signals. In an inflammatory response, these cells differentiate into monocyte-derived macrophages (MDMs) or DCs (MDDCs) when recruited into tissue.

Macrophages are found in tissue throughout the body, derived from circulating monocytes or established prenatally. These cells have an enhanced ability to phagocytose apoptotic and necrotic cells as well as pathogens, but do not tend to migrate out of tissue and are weak APCs. In tissue they can be defined by their expression of surface CD68 or the transcription factor FXIIIA.

Although macrophages can be infected by HIV, there has been little investigation of the role in early HIV transmission due to their weak APC capacity. However, they may serve as a HIV reservoir, phagocytosing HIV-infected CD4 T cells [34], but this function remains understudied. Recently, 4 distinct macrophage subsets were identified (Mf1, Mf2, Mf3, Mf4) in human small intestine mucosa, with distinct function and phenotype [35]. While not discussed further, their roles in early HIV transmission may be of interest.

DCs are located in blood and in tissue, and are found in high concentrations in areas susceptible to antigen exposure, including the gastrointestinal tract [16]. In response to antigens, DCs produce inflammatory cytokines and migrate to lymphoid tissue, acting as strong APCs. Such cells can be identified in tissue through their expression of the marker CD11c. Here, they stimulate CD4 T cells to elicit an immune activation or immune tolerance. There exists 2 ontogenetically distinct subsets of DCs: plasmacytoid DCs (pDC) and conventional DCs, existing as 2 subsets cDC1 and cDC2 [36]. In particular, the cDC2 subtype has been implicated in inducing Th17 formation and may have a role in transferring HIV to T cells residing in lymph nodes [36]. Of interest are three recently identified DC subsets in human intestine, each with distinct lineages and functions: SIRPα-CD103+ DC (cDC1 lineage), SIRPα+CD103+ DC (cDC2 lineage) and SIRPα+CD103+ DC (monocyte lineage) [37]. While these subsets will not be discussed further, their role in HIV transmission in the colorectum is of interest.

DCs are postulated to be the major APC which interact with CD4 T cells [16] and have been shown to transfer HIV to CD4 T cells via two mechanisms [12, 38]. First-phase transfer occurs within 24 hours, involving the transient endocytic uptake of virus through pattern recognition receptors. Following uptake, the virus remains protected for up to 24 hours in neutral pH 'endocytic caves' from where it can be transmitted to CD4 T cells [39]. After 24 hours the virus becomes degraded via acid proteolysis. Second-phase transfer occurs after 72 hours as newly formed virions bud from the plasma membrane as a result of cells becoming infected via the CD4/CCR5 mediated entry pathway [40]. Furthermore, Shen et al. [41] identified a subpopulation of CD11c+HLA-DR+ cells expressing the HIV receptor CD4 and co-receptors CCR5 and CXCR4 in jejunal lamina propria. These cells were observed to rapidly uptake cell-free HIV. A follow up study [42] showed similar

results in vaginal explants, where identified CD13+CD11c+ DCs were found to rapidly bind to HIV and transport the virus into the lamina propria. In an intestinal explant model, Cavarelli et al. [19] showed that within 30 minutes of infection, CD14+ DCs migrate between epithelial cells to sample luminal HIV.

1.1.3 Concluding Remarks and Future Opportunities

Overall, many mechanisms have been implicated in the early transmission of HIV, particularly in human colorectal tissue. However, the specific mechanisms that facilitate transmission is still unknown, with the relative roles of each cell types being unclear. This is in part due to the difficulty in access to clinically relevant human tissue and viral strains, preventing real-world HIV transmission from being modelled.. Furthermore, the lack of sensitivity and specificity that target HIV means that the virus could only be visualized once within 72 hours once enough virus was present. Finally, traditional microscopy methods have not been able visualize the full repertoire of HIV target cells within tissue. As such, there is a unique investigative opportunity to investigate early HIV transmission within a clinically relevant tissue type.

1.2 High Parameter Imaging and Analysis

1.2.1 Introduction

Multiplexed imaging methods are becoming an increasingly important tool for both basic science and clinical research [43–52]. Recently, serial staining immunofluorescence approaches [45, 46, 48–51] and mass cytometry imaging (MCI) approaches [53, 54], have been developed. Such methods provide a unique opportunity for imaging human disease at subcellular resolution, extending the number of parameters visualized within an image beyond the 2-3 markers used in conventional fluorescence microscopy. Simple high parameter fluorescence imaging methods can extend the number of parameters visualized to 6-8 markers, while more complex fluorescence techniques and mass cytometry techniques can enable the visualization of 40-60 markers. These techniques are being rapidly adopted for various applications, including studies in cancer, autoimmune disease and the definition of complex immune subsets during development and homeostasis [55–63].

1.2.2 Serial Staining Immunofluorescence

Serial staining immunofluorescence approaches involve iterative rounds of staining, imaging, and removal of fluorescent signals. Established serial staining techniques include Cyclic Immunofluorescence (CyCIF) [45, 46], co-detection by imaging (CODEX) [49], iterative indirect immunofluorescence imaging (4i) [48], and GEMultiOmyx[50]. In these serial staining approaches, typically 2–3 parameters are acquired per round, thus requiring 13–20 rounds to acquire 40 parameters which is the current limit for MCI. Advantages of this approach relate to broad compatibility with many fluorescence-based imaging systems and the capacity to acquire large areas across multiple tissue sections in a short period of time, which allows parallel processing of many slides. However, there are several disadvantages including lengthy acquisition times which can span weeks, extensive tissue manipulation and perturbance of antigens between staining cycles, autofluorescence, and the lower dynamic range of fluorescence compared to MCI [45, 50, 64, 65]. Further,

considerable expertise and computing power is required to process the resultant large images, which if acquired at a high resolution in multiple Z planes, can form gigabytes and even terabytes of raw data, which must be deconvolved, projected and registered prior to analysis.

1.2.3 Mass Cytometry Imaging Technologies

There are two approaches for MCI—Imaging Mass Cytometry (IMC) [53] and Multiplexed Ion Beam Imaging (MIBI) [54]. In both methods, the first step is the labeling of tissue sections with up to 40 different antibodies conjugated to stable isotopes, mostly from the lanthanide series). In IMC, the tissue is then ablated using a laser with a 1 µm spot size, which rasterizes over a selected region of interest. Plumes of tissue matter are then aerosolized, atomized, and ionized, and then fed into a time-of-flight mass spectrometer for analysis of isotope abundance. In MIBI, an oxygen duoplasmatron primary ion beam rasterizes over the tissue, ablating a thin layer of the tissue surface, which then liberates antibody-bound metal isotopes as ions. Similar to IMC, these secondary ions are then fed into a time-of-flight mass spectrometer for the estimation of isotope abundance [55]. In both methods, the isotope abundance of each "spot" can then be mapped back to the original co-ordinates, producing a high dimensional image qualitatively similar to a fluorescence microscopy image.

The differences between IMC and MIBI have previously been reviewed [44]. However, MIBI has undergone extensive improvements since its initial description, overcoming many of the limitations relating to speed of acquisition and multiplexing capacity [55]. Two important differences we will mention relate to sample ablation and image resolution. IMC uses a laser for sample acquisition and is designed to ablate the entire sample with a fixed lateral resolution of 1,000 nm. However, MIBI utilizes a tuneable ion beam which can be adjusted for varying depth of sample acquisition and also ion spot size (image resolution). This means that the same area can be scanned at a lower resolution to gain an overview and then potential areas of interest rescanned at a higher resolution, reportedly as low as 260 nm, though with a trade-off of longer acquisition times. A comparative summary of features between IMC, MIBI, and serial staining immunofluorescence technologies is presented in Table 1.1.

1.2.4 Significance of High Parameter Imaging

Serial staining immunofluorescence and MCI are landmark developments because it allows for upward of 40-60 markers to be simultaneously stained, acquired and visualized, enabling a variety of distinct cell types to be analyzed concurrently in their native microenvironment. The microenvironment consists of a complex matrix of fluids, proteins and cells which provide signals that shape a given cells phenotype and function within an organ in both health and disease [66–70]. Indeed, there is increasing evidence that cellular functions are programmed not just by cell ontogeny but also by signals from the surrounding microenvironment. Examples include Monocytes and Dendritic Cells and T cells which exist in several functionally diverse subsets, which vary across tissues so as to meet the requirements of their local environment [71–75]. Specific subsets of Dendritic Cells, Innate Lymphoid Cells and T cells can carry out distinct functions at a given point in time, inducing either tolerance or inflammation depending on a host of signals derived from both cytokines and direct cell contact [76–79]. In the context of disease pathogenesis,

Table 1.1: **Highly multiplexed imaging technologies.**
* A smaller spot size (resolution) results in longer acquisition times. A lower limit of 260 nm is referenced in a recent publication, but the actual data acquired in the study was at a resolution of 500 nm [55].
** There is no hard upper limit for serial staining protocols, but published data has shown approximately 60 markers per section [46, 49, 50]. A limit of 40 markers for IMC and MIBI is derived from interpretation, based on both the indicated references and current reagent availability.
*** The rate-limiting step for serial staining protocols is the antibody incubation period which can take hours and is often performed overnight.

	Serial Staining immunofluorescence	**Metal tagged antibodies**	
Examples	CyCIF, GEMultiOmyx, 4i, CODEX	IMC	MIBI
Resolution	∼200 nm	∼1,000 nm	∼260 nm*
Simultaneous detection limit	1-5	40	40
Max number of epitopes imaged per section**	∼60	40	40
Throughput***	Hours or 1 day per cycle per tissue section	$1mm^2$/2h	$1mm^2$/5h (500 nm resolution)
References	[45, 46, 48, 50]	[53]	[54, 55]

the tumor microenvironment is now appreciated as a complex signaling network between transformed and non-transformed cells, with the latter being corrupted to promote tumor function [80, 81]. The importance of the microenvironment for cell function is clear. The major contribution of high parameter imaging is that it provides spatial data for a large number of parameters at subcellular resolution. As such, we are now positioned to discover interdependencies between complex cell subsets in health and disease. These interactions can be further investigated *ex vivo* to determine their functional outcome and contribution to disease progression.

These techniques are also an important development for practical reasons as it enables complete studies to be performed on archival samples. This is particularly useful as research questions evolve with time and it is invaluable to be able to repeatedly interrogate the same sample for different parameters. This feature will be particularly helpful for investigations of inflammatory disorders where significant heterogeneity can exist, making it difficult to accurately characterize the cell types involved and thus the immune motifs underlying the disease; such is the case for dendritic cell subsets which are partly defined by surface markers that are labile during inflammation [82]. Furthermore, many studies can only be performed using small biopsies or precious post-mortem samples, as in brain and pancreatic tissues, with samples typically curated through biobank networks [83, 84]). As such large gaps remain in our understanding of disease pathogenesis in these tissues; a gap which high parameter imaging is poised to fill.

1.2.5 Image Processing and Analysis

Due to its ease of use, high parameter imaging is poised to be a useful tool in clinical research. However, a key bottleneck is related to both image processing and the inherent difficulty of analyzing up to 40-60 parameters with added spatial dimensions. In this section, we first cover all techniques that have been used for processing high parameter image data data. Broadly, this includes image denoising, single-cell segmentation and

finally tissue and cell-type annotation. Next, we discuss approaches for the analysis of images that have been implemented for studying disease models. The analysis section is formatted as a series of general biological questions which can be answered using image analysis. For each question we discuss both its clinical significance and the specific techniques used in imaging studies to answer each question.

Processing

In this section, we outline the steps taken to process images, allowing downstream analysis with a single-cell approach.

Autofluorescence Autofluorescence is a phenomenon specific to immunofluorescence imaging which arises from tissue structures which have excitation and emission profiles that overlap with the excitation/emission filter setups in fluorescence microscopes. This arises from structural and connective tissue, cellular contents as well as fixatives used to preserve tissue [64, 85–87]. These structures can lead to false positive signals in data analysis.

While several experimental methods exist to remove autofluorescence [85, 88–90], these methods can quench real signal or shift autofluorescent signals. Hence, computational methods for autofluorescence have been proposed. These methods include spectral unmixing approaches [91, 92], as well as algorithmic subtraction of a background reference image acquired prior to staining [50, 93]. While robust, spectral unmixing methods currently require specialized instrumentation and proprietary software, limiting its use [91]. Furthermore, background subtraction methods requires the acquisition of additional images and will require precise image intensity scaling and image registration [50].

With serial staining immunofluorescence methods becoming more widely available, the removal of autofluorescence signals has become increasingly important to facilitate accurate downstream analysis.

Denoising in MCI An important issue common to all image analysis is the presence of noise and artifacts which must be removed prior to analysis (Figure 1.1A, left). Robust and stable methods for denoising will become increasingly important if high parameter is to be applied within the clinical setting, allowing for accurate patient sample characterization. While autofluorescence is avoided in MCI, there are noise profiles that are specific to MCI, in contrast to other imaging technologies, and may be specific to tissue types [55, 59, 60, 94]. To comprehensively analyse the images obtained, various computational methods for denoising to preserve real signal and remove technical artifacts have been proposed [55, 59, 60, 94]. At present there is no consensus on the most appropriate way to denoise images with research to date employing homebrew approaches based on the level and composition of noise observed by the investigators. Such approaches include correcting for channel cross-talk [94, 95], removing objects that differ from real signals in terms of size and pixel distributions [55, 60], and by using image filters to identify artifacts [55, 60]. Here, we describe methods proposed to eliminate noise and artifacts in high parameter images.

Crosstalk is the phenomenon in which signals from one channel are introduced into adjacent channels. This has been observed when comparing channels within $\pm$ 3 atomic masses from each other, occurring due to the presence of contaminating isotopes of similar masses [96–98]. Crosstalk can even occur within 16 atomic masses due to oxidation [98].

This confounding phenomenon needs to be corrected as it can lead to the misidentification of real signal within a single channel, particularly if adjacent channels correspond to markers that may be co-expressed. To correct for crosstalk, two methods have been proposed. Wang et al. observed a linearly correlated increase in pixel intensities at high pixel values for adjacent channels when plotting the intensities for two channels [60]. They classified these pixels as crosstalk, and compensation was performed by resampling their intensity values in the given channel, providing a post-acquisition method for correcting for crosstalk. Chevrier et al. presents a bead-based compensation workflow to account for crosstalk, made available as the CATALYST R/Bioconductor package [94]. Damond et al. implements this solution, measuring channel crosstalk using a slide with the metal isotopes used [59]. There are trade-offs between using a fully post-acquisition approach, as opposed to a bead-guided approach. The post-acquisition method by Wang et al. is advantageous as it minimizes IMC acquisition time and resources required. However, it is unclear if their approach is valid for other images, and it is difficult to assess if real signals are removed. Additionally, correction may not be necessary if the marker panel employed is well-designed and titrated. If certain markers are expressed at vastly different levels across samples, panel design alone may not eliminate crosstalk. Hence, users should make a judgement as to whether crosstalk correction is necessary for their study.

Background noise and the corrections required can be specific to certain tissue types and experimental setups. As such, several "homebrew" computational methods have been developed to identify and remove noise in images. Wang et al. observed horizontal streak artifacts within their image [60]. The authors accounted for this by using a 5×5 μm^2 median filter which excludes the middle row. For each pixel, the median pixel value within this filter is measured, and the central pixel is removed if it is brighter than this median and is in the top 2% of pixel intensity values, characteristic of these streaks. Keren et al. observed a background artifact in areas of the slide outside of tissue in all channels [55]. To correct for this, a background channel, not containing antibody derived signals, was obtained. The area corresponding to background was identified with a threshold, and the pixel intensities were reduced in all other channels within this area. The authors also observed that low density pixel signals (those with few neighboring pixel signals) were associated with noise, while real signals tended to aggregate together, corresponding to cellular staining. To remove the interfering low density pixels, each pixel in the image was assigned a score by calculating the average distance to the 25 nearest positive counts. A bimodal distribution was obtained, and pixels above the crossing point of the distributions were removed, corresponding to the low-density noise. This background removal method complements wet-lab based optimizations of blocking methods and antibody concentrations. Indeed, it can be very challenging to address all signal-to-noise issues for large antibody panels. As such, computational methods, as described here, are an important preprocessing step to ensure reliable downstream analysis.

While various custom algorithms have been successful for denoising in previous studies, the application of MCI within a clinical setting will require an improved understanding of the sources of noise for specific samples. The standardization of "best practice" procedures for sample processing, staining, and acquisition in addition to image post-processing methods will be necessary going forward.

Cell Segmentation Fundamental to the study of tissues in health and disease is the identification and characterization of individual cells. In microscopy this is achieved through single-cell segmentation, which involves identifying the boundary of individual

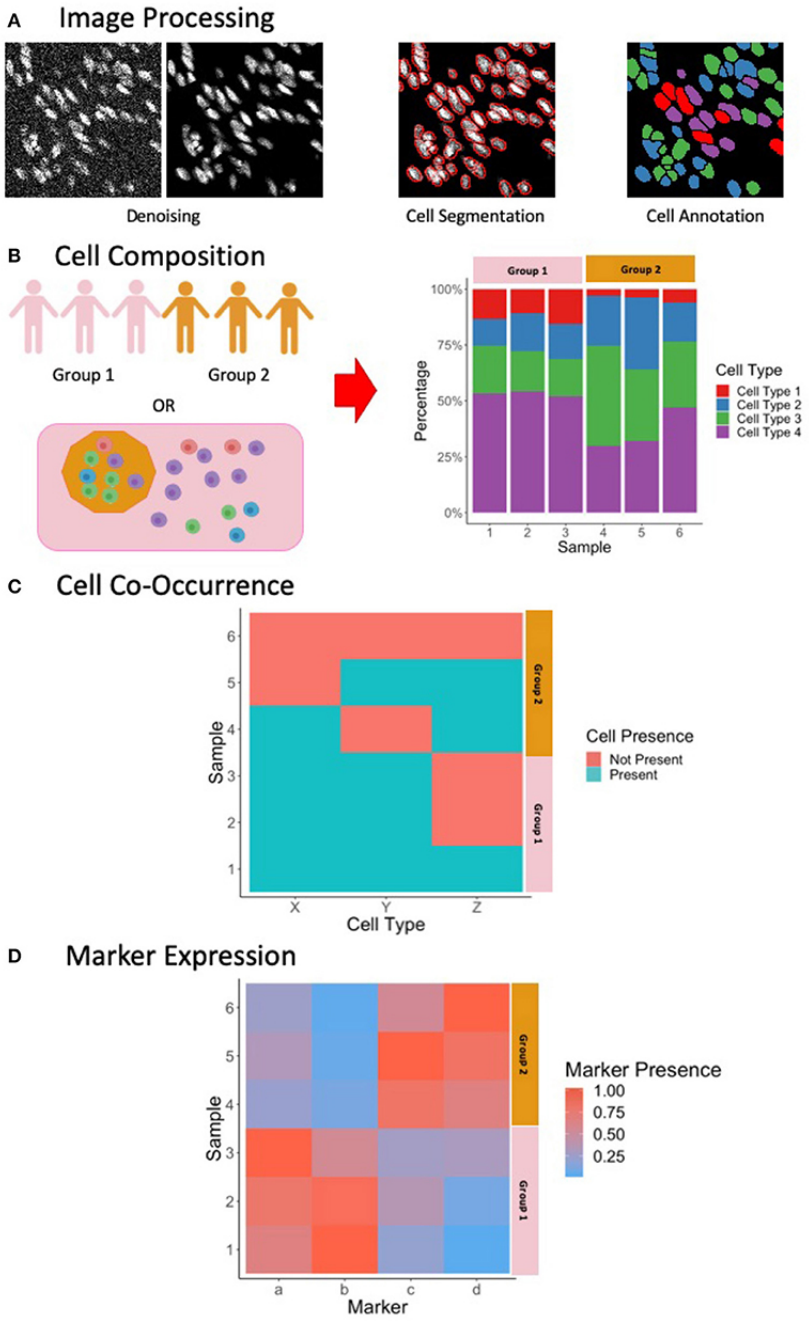


Figure 1.1: **Summary of Image Processing and Analysis Techniques.** **(A)** Following image acquisition, image processing is performed to denoise the images, perform single-cell segmentation to identify cell outlines, and to classify these cells based on marker expression. **(B)** One way of exploring cell composition between groups is to compare the change in the cell fractions. **(C)** Another way to explore cell composition is to classify patients as being positive and negative for a particular cell population. The co-occurrence of cells can be presented similar to what is presented here, and significance of co-occurrence can be identified using a chi-square test. **(D)** Differences in marker expression between patients can be visualized using a heatmap.

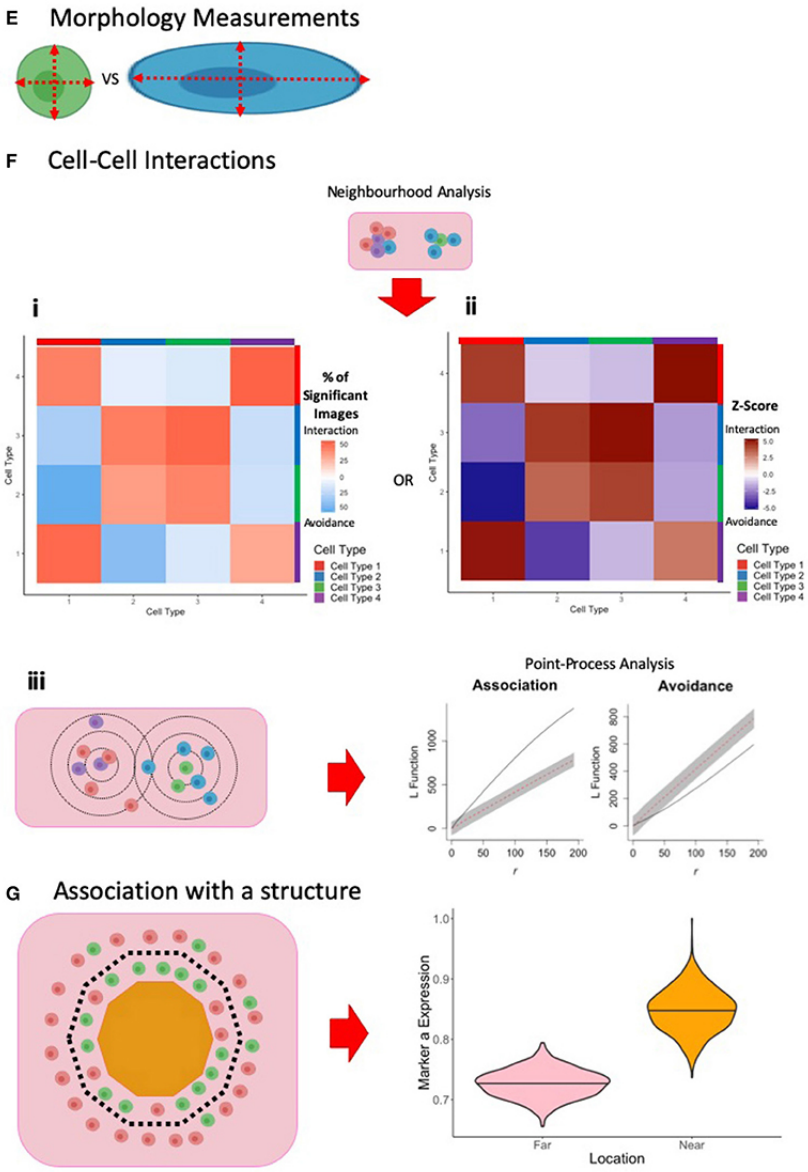


Figure 1.1: **Summary of Image Processing and Analysis Techniques.** **(E)** Cell morphology measurements can be used to explore cell phenotypes. **(F)** Cell-cell interactions can be measured using neighborhood analysis or point-process analysis. With a neighborhood analysis, percentage of significant images **(i)** or Z-scores **(ii)** of the cell-cell interactions can be represented as a heatmap, with significant associations associated with a more positive Z-score and significant avoidance is associated with a more negative Z-score. With a point-process analysis, an L function can be used to assess the significance of cell-cell interactions. The L function being above or below the gray envelope generated by bootstrapping corresponds to association and avoidance, respectively (iii). **(G)** One way of measuring cell or marker association with a marker is to classify cells as being near or far away from the border. A cell composition analysis can be used to explore differences, or differences in marker expression can be explored, as shown here. Parts of this figure were made Biorender.

cells (Figure 1.1A, middle). This section discusses several approaches which have been applied for segmenting high parameter image data (summarized in Table 1.2).

Many pipelines for single-cell segmentation have been established and adapted for high parameter imaging. These pipelines typically apply a threshold to a nuclear image and implement watershed segmentation to identify nuclear boundaries. Dilation of the cells, or the use of a cell-membrane marker, identifies the remaining cell body. The popular CellProfiler software [99] is often used for single-cell segmentation, with the user being able to provide inputs on the size filters, smoothing, and thresholding applied among other parameters to achieve segmentation. This is implemented by Wang et al., taking advantage of the many parameters used in IMC by using a range of non-immune and immune cell membrane markers for cell segmentation [60]. This approach has the advantage of not requiring user training, requiring few user inputs for implementation. However, CellProfiler may not be able to segment cells that are packed tightly, as in tumors and lymphoid tissues, especially when the resolution is low as in the case of MCI. Schüffler et al. proposes a method in which multiple membrane proteins are weighted together to define the cell membrane [100]. The proposed method performs an exhaustive search for an appropriate weighting and smoothing of all cell membrane channels, and provides a score based on how successful segmentation is performed. This self-reflective scoring may be useful for assessing the success of segmentation, but it is unknown whether it is successful for difficult, high-density images. Finally, Durand et al. employs an in-house-developed segmentation pipeline to achieve single-cell segmentation of tonsil tissue [101]. First, a Laplacian-of-Gaussian filter is applied, which resolves nuclei as spots with a local minimum. A h-minima transform is then applied to identify these local minima [102]. Finally, a single-cell segmentation mask is obtained by applying a watershed transformation to the linear combination distance map obtained from the h-minima transform and the average image of all membrane-bound marker proteins. The cellular regions are restricted by a defined radius of 8 pixels around each local minimum to avoid oversized cells. Ultimately, these pipelines allow cell boundaries to be identified without user training.

For more precise single-cell segmentation, supervised classifiers have been successful, particularly for MCI. These approaches require humans outlining single-cells to produce a set of well-annotated cells that can be used to train machine learning algorithms, with the advantage that humans may be better at identifying the subtle details that separate cells. Schulz et al. [58] and Damond et al. [59] implement the popular Ilastik toolkit [103, 104], employing a random forest classifier for cell segmentation, while Keren et al. [55] implements DeepCell [105], which employs deep-learning for cell segmentation. With both tools, training sets are developed using nuclear, cytoplasmic, and membrane markers, and a probability map is produced describing whether a pixel is nuclear, cytoplasmic, or background. CellProfiler, or conventional thresholding and watershed segmentation is then used to identify cells and their bodies based on the probability maps. This workflow of performing segmentation on probability maps was first demonstrated by Schapiro et al. [99]. These supervised methods have been successful at separating cells that are clustered together, and can be advantageous to using CellProfiler in a standalone manner. However, these techniques require users to generate substantial training data with a new classifier needing to be generated for each experimental panel and tissue type which can be time consuming.

In general, if the outline of cells is obvious, using CellProfiler may be sufficient for performing single-cell segmentation. However, if cell shapes are more complex, as in the case of neural tissue, or if dense cell structures are present within tissue structures,

Table 1.2: **Software for cell segmentation and cell classification.**

	Technique	Description	References
Cell Segmentation	CellProfiler	Identify primary object with nuclear marker, secondary object with membrane marker	[107]
	Weighted sum of membrane markers	Segments using a weighted sum of membrane markers	[100]
	Ilastik	Uses a random forest classifier, defining pixels as nuclear, cytoplasm, and background based on user training data. Probability maps can be used as an input for segmentation in CellProfiler	[103, 104]
	DeepCell	Identifies cell nuclei based on training data, using deep-learning	[105]
Cell Classification	Manual gating	Users manually identify their cells based on marker expression	
	Hierarchical clustering	Identifies clusters in a hierarchical cluster by grouping together cells or clusters that are most similar to each other	
	Phenograph	Models cells as a nearest-neighbour graph in high-dimensional space	[102]
	FlowSOM	Self-organizing maps used to identify cell populations. Meta-clustering is then performed to find a given number of populations	[108]
	Ilastik	Uses a trained random forest classifier to classify identified single cells	[103, 104]

then the use of classifiers will be more suitable. In fluorescence images, these classifiers have been shown to outperform classical methods for segmentation [106], but an extensive comparison using MCI has not yet been performed. As the use of high parameter imaging becomes more universal and applied within a clinical setting, there will be an increased need for more precise segmentation. It is likely that the most appropriate method will be to use a well-trained classifier. For generally applicable classifiers, users may have to contribute to an existing online classifier, creating a diverse training set to perform cell segmentation. Much investigation will hence be necessary in the future for improved and more generalized segmentation.

Tissue and Cell Annotation Immune cells exist in great diversity within both healthy and diseased contexts. Along with canonical cell types such as Dendritic Cells, Macrophages, T cells, and B cells, each cell type is comprised of diverse subsets which differ throughout the body. Importantly, specific subsets can play a crucial role in disease manifestation, even when their prevalence is extremely low. As such, accurate and high throughput methods for the annotation of cell types (Figure 1.1A, right) and the tissue compartments in which they reside, are essential. Here we discuss several approaches that have been employed for the annotation of high parameter image data (summarized in Table 1.2).

The simplest approach for identifying cells is by selecting manual gates based on scatter plots of marker expression, similar to other single-cell technologies such as flow cytometry. Marker expression is typically quantified by summing the ion counts within a single cell as outlined by segmentation and dividing by the area of the cell. The histoCAT package [99] provides a tool which allows users to gate on cells and visualize the presence of these cells within their image. Furthermore, the single cell data can be exported from histoCAT for downstream analysis using commercial platforms such as FlowJo or Cytobank, and also open source platforms such as Flowing Software. However, a key advantage of histoCAT is that cell selections can be visualized on the image in real-time, which facilitates greater accuracy when selecting gates. However, while manual gating provides a user with full control over the cells being classified, this can be time consuming, especially when

many markers are considered. Nevertheless, manual gating may be useful for exploratory analysis of image data.

One approach for semi-automated gating is by using a mixture model, such as the implementation by the mclust R package [109]. This package is used to classify cells as being positive or negative for a marker, based on the mean pixel intensity in that specific marker channel. Another approach is by Boolean rules based on whether cells are positive or negative for these markers to classify cell types. Wang et al. implement this method, but set additional manual cutoffs as informed by the mixture models to identify positive and negative populations [60]. This approach will only be applicable for markers with which cells can be discretely positive and negative for, but not when cell-type definition relies on a continuum of marker expression (e.g., low, mid, and high).

Automated gating strategies employing clustering techniques to group cells by similarities in marker intensity have become popular in all high-parameter imaging assays. This provides a quick and unbiased approach for classifying cells in tissue. Schulz et al. [58] employs PhenoGraph [102] to cluster cells, employing a nearest-neighbor graph to identify phenotypically coherent subpopulations. Here, they use both marker expression as well as RNA expression to cluster cells. Durand et al. use a hierarchical clustering approach on all markers, obtaining 60 clusters which was arbitrarily chosen to overclassify cells [101]. This allowed the authors to identify smaller yet distinct clusters with some similar clusters manually merged when the clusters were annotated based on known cellular phenotypes. Keren et al. [55] clusters cells into immune and non-immune cells using FlowSOM [108], which employs a self-organizing map to identify cell populations. Lineage marker expression was used to cluster cells. This was applied iteratively, first to distinguish between immune and non-immune cells, then to classify non-immune cells into epithelial, mesenchymal, endothelial, and unidentified cells, and finally to classify immune cells into specific subsets. The approach taken by Keren et al. employs only canonical cell markers, leading to the identification of canonical cell subsets. Expression of functionally significant markers was then assessed on the defined cell subsets in different tissue compartments. In contrast, Schulz et al. clusters using all markers, leading to canonical cells being divided by marker expression. For example, two CD3 high T cell clusters were obtained, one of which expressed CD3 only, and the other being a potential memory T cell subset. Importantly, this clustering revealed the identification of rare cells that express CXCL10 RNA. Durand et al., however, merges clusters with a similar phenotype. Hence, a choice needs to be made as to whether to include all markers or only lineage markers when investigating cell phenotypes. With most automated gating strategies being implemented in the R statistical environment, the R package cytomapper [110] allows single cell and image data to be stored in conventional data structures for clustering while also providing a framework for visualizing the distribution of the cell types identified by clustering.

Finally, users can employ supervised classifiers, providing training data to predict cell types based on both marker expression and the visual texture of the signals. For example, membrane markers will be localized only to the membrane of the cell. This can be achieved using an interactive classifier such as Ilastik, where users can annotate cells as the cell subsets they are interested in Ilastik uses both marker expression level and morphology to classify cells based on the provided training data. Damond et al. implements this classification iteratively, first to classify cells as islet, immune, exocrine, and "other" cells [59]. A second round of training and classification was then performed to classify the different immune, islet, and exocrine cells, and "other" cells were classified as endothelial,

stromal or unknown cells. The classifier is advantageous as classification is informed by both marker expression and texture as defined based on more reliable human judgement. However, the training of a classifier can be time consuming, and this approach will only be able to identify user-defined cells. Hence, supervised classifiers will not identify other cell marker phenotypes that automated gating may identify.

Following cell classification, tissue compartment identification can be performed. This is useful for exploring the role of tissue structures in the context of disease. Keren et al. [55] and Wang et al. [60] use classified tumor cells and islet cells to identify the tumor and islet areas, respectively. Damond et al. uses Ilastik to identify islets and blood vessels by constructing training data using a range of structural markers, while Durand et al. uses E-Cadherin, CD19, and CD3 to identify the crypt, B cell zone, and T cell zone of tonsil tissue. The identification of these tissue structures is important because of their role in disease pathology. For example, the tumor-immune boundary has been used as a prognostic indicator for tumor progression, and islet cell composition and morphology have been observed to change with disease progression [111–113]. Identifying these key compartments and their borders hence allow their role to be observed.

Similar to the segmentation of individual cells, cell-type annotation in the clinical setting would require automated and standardized methods for cell-type classification. At present, classifiers used for annotation are trained on a study-by-study basis. Although accurate, it has not been established that these approaches are generalizable or time efficient for use in the clinical setting. Ultimately, classifiers will need to be constructed and trained to account for patient and experimental variation.

Analysis

In this section, the key biological questions that are answered through image analysis are discussed (summarized in Table 1.3).

How to Stratify Data for Analysis? To understand the biological processes underlying disease, the appropriate stratification of patient data for analysis is important. The simplest method is to group data based on clinically defined categories such as "time since diagnosis" or "patient survival." This approach is implemented by both Wang et al. [60] and Damond et al. [59] in their study of T1DM. Here, they stratified their patient groups based on time since diagnosis, with an additional control group. Although this method is often appropriate, stratification based on a biologically meaningful model of disease can offer a powerful and complementary approach for revealing disease specific relationships that simple clinical groupings could miss. For example, as diabetes is a progressive disease, Damond et al. performed pseudotime analysis (discussed below) to group islets into three "pseudostages" of disease. This followed from their observation that islet profiles followed a spectrum during the early-stages of disease, resembling both healthy islets and late-stage islets as well-intermediate stages in between. Additionally, as tumor-immune organization is known to predict survival for certain cancers, Keren et al. performed a spatial enrichment analysis (discussed below), generating a metric for tumor-immune cell mixing and allowing the investigators to stratify patients based on tumor organization [55]. The decision on how to best group data for analysis is crucial for the discovery of disease specific immunological motifs. In reality, this part of the analysis stretches back to experimental design. To effectively use high parameter imaging as a primary research tool, it is important to carefully consider beforehand, choice of patient samples, availabil-

ity of clinical data and also panel design. These three aspects will inform the types of data stratification that are possible and therefore the scope of questions that can be asked and answered using high parameter imaging.

How Does Cell Composition Change With Disease Context? The prevalence of specific cell subsets is associated with disease outcomes, both in the clinical setting and in models of disease. As such, the basic analysis of cell composition is an important first step which can also inform downstream analyses. In present studies, this has taken two approaches. The first is to quantify cell compositions and then compare these between different patient groups. This is done as either absolute counts of a specific cell subset, a measure of its proportion among a larger group of cells, or as a cell density per mm^2 of tissue. The second is to examine the co-occurrence or anti-occurrence of cell types, providing an insight on any causal pathways that may underlie disease. In this section we summarize how imaging studies have explored cell composition within tissue.

The cell subset composition can be presented as the proportion of the total cells (or all immune, tumor, islet cells, etc.) (Figure 1.1B), the total number of cells, or the cell density. There are many advantages and disadvantages to these different approaches for quantification. Total counts can allow for patient-patient comparisons, allowing interpatient variations to be observed. When comparing between groups of patients, the cell proportion may be more appropriate for comparison, normalizing the data to account for interpatient variation. Cell density per mm^2 of tissue may be appropriate when comparing cells within compartments, with the data being normalized by the area of the compartment. The density measurement is also useful for comparing small changes that are overwhelmed by the abundance of another cell type. Ultimately, the choice of measurements used is dependent on the question being asked.

This cell composition analysis is implemented by both Damond et al. [59] and Wang et al. [60] in studying how the islet cell composition changes with T1DM progression. Both studies observed a decrease in beta cell fraction, and an increase in gamma cell fraction with disease progression, relative to all other islet cells. Damond et al. further observed a small decrease between pseudostage 1 and 2 islets, followed by a significant decrease between pseudostage 2 and 3 islets. Additionally, Damond et al. and Keren et al. present the proportion of immune cell subsets within their images, assessing the composition of immune infiltration within tissue. Data obtained from cell composition analysis can also reveal meaningful biological relationships. For example, Keren et al. ordered patients by number of infiltrating immune cells and found that patients with more immune cells were more likely to have a "compartmentalized" phenotype. Additionally, Damond et al. found that when ordering patients by the number of islet cells, stratified by patient diabetes status, mid-sized islets had a higher proportion of beta cells. Presentation of data in this manner can aid in the interpretation of single-cell MCI data.

To assess cell subset co-occurrence or anti-occurrences, two approaches have been used in present MCI studies. The first approach is to observe whether the count or proportion of one cell subset is correlated with that of another cell subset, assessed using Pearson's Correlation. This measurement is useful when investigators want to show that an increased presence of one cell type is accompanied by an increase or decrease of another cell type, and is appropriate when both cell types are often or always present within that tissue type. The second approach is to convert cell counts into categorical data by classifying images as being positive or negative for a given cell subset if the count exceeds a user-defined cutoff (Figure 1.1C). A chi-square test is then used to quantify the signif-

icance of co-occurrence. This measurement is not very useful when both cell types are often or always present within that tissue type. Hence, this measurement is suitable only when the cell types being investigated are not consistently present within that tissue type.

The co-occurrence approaches mentioned above have been applied by Keren et al. and Damond et al. In studying immune infiltration into tumors, Keren et al. observed that there was a correlated increase in CD4+ T cell proportion and a correlated decrease in macrophage proportion. Similarly, when studying immune cell infiltration into the islets, Damond et al. observed a correlated increase in CD4+ helper, and CD8+ cytotoxic T cells in pseudostage 2 islets. This revealed that both CD4+ and CD8+ T cells are recruited simultaneously into the islets during the onset of diabetes, potentially co-operating to mediate beta cell destruction. Furthermore, to assess co-occurrence of cells in tumor infiltration, Keren et al. classified each patient as being positive for a given immune cell if the cell count is 10, and negative otherwise. A chi-square test subsequently revealed relationships such as patients with B cell infiltration into their tumors also had CD4+ and CD8+ T cell infiltration. The relationships observed by these analyses reveal a potential coordination in the immune response in both tumors and islets, with the recruitment of several cell types occurring.

Does the Expression or Co-expression of Cell Markers Change With Disease Context? In addition to changes in cell composition, understanding variations in functionally relevant markers is essential for understanding disease pathology. Indeed, many interventions targeting cancer, infectious diseases and autoimmune diseases use antibodies and small molecule inhibitors targeting cytokines or cell-associated ligands/receptors [114–116]. Through the many markers afforded by high parameter imaging, these diverse markers can be studied within the disease pathology setting. This section will explore how marker expression is examined in images.

In studies so far, the exploration of cell marker expression has taken many pathways. One approach is to compare marker expression among canonical cell subsets, with fold-changes being expressed as a heatmap (Figure 1.1D). Marker expression can also be measured at the tissue compartment or patient level, with expression level visualized as a heatmap for each sample (Figure 1.1D). By stratifying samples into groups, direct comparisons can be made. To assist with the analysis of the many markers used by MCI imaging, dimensionality reduction techniques have been used. These include principal components analysis, t-Distributed Stochastic Neighbor Embedding (t-SNE), and pseudotime analysis. Finally, the investigation of preferential co-expression of markers can be assessed by classifying images as being positive or negative for a given marker and using a chi-square test to quantify the significance of co-occurrence. Each of these approaches can be used to investigate differences in marker expression within different samples, with each analysis telling different aspects of the overall pathophysiological story. Investigators should use the appropriate investigation required depending on the question being asked, and the cellular pathway being explored.

In their investigation of T1DM progression, Damond et al. studied the change of islet marker expression within islets [59]. While the investigators observed a decrease in beta cell fraction as described previously [113], they wanted to further investigate whether this was a result of beta cell loss, a downregulation of beta cell marker expression, or both. To investigate this, the authors performed a pseudotime analysis using the trajectory inference algorithm SCORPIUS [117, 118]. This was performed by measuring the islet marker expression profiles of each individual islet. The algorithm finally assigns

a value between 0 and 1 to each islet, relating the marker expression profile of islets to the T1DM development timeline, and allowing the investigators to stratify the islets into three pseudostages. Specifically, they observed a strong downregulation of beta cell markers between pseudostages 1 and 2, and stability between pseudostages 2 and 3. The authors concluded that progression from pseudostages 1 and 2 may be driven by the down regulation of beta cell markers, while the transition between pseudostages 2 and 3 is reflective of cell death. The assessment of changes in marker expression, combined with cell composition analysis, can reveal the mechanisms behind a disease timeline.

Keren et al. investigated the expression of the immunoregulatory proteins PD-1, PD-L1, IDO, and LAG3 in their study of breast cancer [55]. Through a chi-square test, they found that patients expressing one of these proteins expressed another, implying that multiple immunosuppressive pathways are present within the tumor environment. Additionally, it was found that the presence of regulatory T cells accompanied the presence of at least one of these markers, reflecting the potential for these proteins to induce the differentiation of naïve T cells toward a regulatory T cell phenotype. Such results provide insight as to the signaling pathways that are present within the disease setting, and relate molecular expression profiles to the histological structure of the tissue.

Ultimately, it is important to understand the distribution and expression level of functional markers relevant to disease. These may be chosen based on the literature as in the MCI studies discussed here, or alternatively using other omics technologies, such as genomics and proteomics platforms [119, 120], to pre-screen samples for suitable candidates. Importantly, the inclusion of such markers allows one to infer biological processes from static 2D images.

Does Cell or Structural Morphology Vary With Disease Context? Another important aspect of cellular phenotype is its morphology (Figure 1.1E). Just as with marker expression, cell morphology can also be associated with disease context or with drug treatment. Morphology measurements such as area, perimeter, solidity, eccentricity, and circularity can be made with analysis software such as histoCAT [99], as well as most image analysis packages [107, 121]. These measurements allow structural changes to cell or tissue to be identified with changing disease context, or with drug treatment. However, the reliability of the measurements is dependent on how accurately segmentation of objects are obtained. This can be difficult with the lower resolution of MCI images, but may be reliable when classifiers are used, as mentioned previously.

Morphological measurements can be used to assess the integrity of histological structures. Damond et al. applies these measurements to their islets to assess changes with diabetes progression [59]. The authors measured the islet extent (islet area divided by islet bounding box) and solidity (portion of pixels in the islet convex hull that are also in the islet), indicative of shape regularity. These two measurements were found to decrease between pseudostages 2 and 3, indicative of a more irregular islet shape, associated with beta cell loss and diabetes progression. Thus, morphology measurements can provide an unbiased quantification of tissue structure, identifying degradation as described here, but may also be used to highlight swelling or growth.

Cell morphology can be affected by drugs and has utility in drug-discovery [122]. Bouzekri et al. uses morphology measurements to assess drug effects on breast cancer cell lines as visualized by IMC. The authors found that certain drugs led to an increase in size, with morphological measurements such as area, perimeter, and major- and minor-axes increasing following drug application. In combination with protein measurements,

these observations may allow researchers to propose transduction pathways affected in response to drug treatment [123, 124].

Are There Any Interactions Between Specific Cell Types Within Tissue, and Does This Change With Disease Context? Within previous studies, two methods have been used to investigate cell-cell interactions. The first is through the neighborhood analysis algorithm described by Schapiro et al.[99]. This method identifies whether a cell of type X is within a user-defined neighborhood of cell type Y, and vice-versa. This is performed by dilating each cell in a single-cell mask by a user-defined number of pixels (usually 4–6) and counting the cell types that it overlaps with. To assess significance, a bootstrapping approach was implemented, in which the annotated cell labels are randomly reassigned. The mean number of cells of type X within the neighborhood of cells of type Y are calculated for each simulation and for the real distribution. The statistic obtained for the real distribution is then ranked against the simulated statistics with two one-tailed permutation tests to obtain a p-value. The upper-tailed test corresponds to interaction, while the lower-tailed test corresponds to avoidance. When applied to a large number of donors, this can be represented on a heatmap as the percentage of significant avoidance or interactions for each cell pair (Figure 1.1Fi). The second method was to count the number of cells of type X within a user-set distance away from cells of type Y. A similar bootstrapping approach was implemented, and the number of cells was remeasured to generate a distribution from which Z-scores are obtained. This relabeling can be performed with all cells, providing context of global organizational patterns of the cells, or by constraining within a specific group of cells (e.g., immune cells, tumor cells, T cells, etc.), providing a more context dependent answer. A negative Z-score corresponds to avoidance, while a positive Z-score corresponds to association, and these values can be visualized on a heatmap (Figure 1.1Fii). These two approaches are effective for identifying cell-cell interactions. However, they do not provide any context of the cell-cell interactions over a wider distance, and does not reveal whether cells traffick toward a particular target. Additionally, an arbitrary distance needs to be chosen, and the sign of the Z-score and hence the interpretation of cell-cell interactions, can vary with scale.

This neighborhood analysis technique has been applied by Damond et al., who observed reduced beta cell associations in the third pseudostage, representative of beta cell destruction, while immune cell associations with other immune cells was increased in the second and third pseudostages, indicative of an immune response [59]. They also found that the number of interactions of beta cells with CD4+ helper and CD8+ cytotoxic T cells was much higher during pseudostage 2, in line with their previous results.

Additionally, Keren et al. counted the number of cells positive for marker X located within 39 µm from marker Y. This resulted in the identification of three distinct levels of tumor and immune cell mixing: "cold," with low immune infiltration, "mixed," with high immune infiltration, and "compartmentalized," with tumor and immune cells forming distinct clusters separated from each other. The authors developed a mixing score to quantify this, defined as the number of immune-tumor interactions divided by the number of immune-immune interactions. Furthermore, when plotting Kaplan-Meier curves, which showed survival as a function of time for patients, they observed higher survivability in patients with "compartmentalized" tumors compared to patients with "mixed" tumors. Here, the spatial organization of tumor was related to patient survivability.

To explore avoidance or association at a range of distances, cells can be modeled as a marked point process model [125], in which cells are represented as labeled points on a

plane. One approach is to use Ripley's K and L functions to model cell-cell interactions, with the variance stabilized L function being a useful transformation to the K function (plotted in Figure 1.1Fiii). Simply, Ripley's K function is a function which models the number of cells of type X a certain distance away from cells of type Y, as a function of distance. Bootstrapping is once again used to generate significance. This was used by Setiadi et al. in fluorescence imaging to show that B cells cluster in tumor-draining lymph nodes compared to healthy lymph nodes [126]. While applied to the same cell type in this example, this can be applied to pairs of cells of different types, or with cells of a specific type to a pathogen. This can provide context of the significance of these interactions along a wider range, and to observe how interactions can change with scale, and may give insight to any cellular trafficking from a steady-state image. These functions, along with other functions and methods for comparison between samples, are readily available in the R package spatstat [127]. However, a disadvantage of these models is that no single Z-score is given, making visualization and interpretation difficult. Baddeley et al. [128] proposes envelope-based tests to measure the statistical power of the interaction or avoidance, while another strategy may be to determine the percentage of images with which interaction or avoidance was significant, similar to Schapiro et al. [99]. An investigation on the appropriate spatial statistic will be necessary to make robust conclusions about any cell-cell interactions, especially in the context of high parameter images where many cell subsets are being investigated simultaneously.

Spatial analysis can be applied to a wider range of cell subsets compared to conventional microscopy over a range of distances. This allows a diverse range of cell-cell interactions to be performed, with the possibility of cell-pathogen interactions to be investigated in the future. Although these images only provide a snapshot of the tissue environment, the identification of significant interactions may bypass the need for more complicated techniques using live imaging. As well, interactions observed in the native microenvironment provide a sound rationale for *ex vivo* co-culture experiments, to investigate the functional outcome of certain cell-cell interactions. Given that specific cell-cell interactions have already been associated with patient outcomes [55, 59], such interactions metrics could prove a useful prognostic indicator in a variety of disease settings.

Do Cells Localize to Histological Structures and Does This Vary With Disease Context? In addition to cell-cell or cell-pathogen interactions, it is useful to understand whether cells or pathogens localize to a specific histological structure, such as epithelium, tumors, and islets, which have been implicated to have an involvement in disease pathology.

In previous studies so far, the number of cells or the amount of cell expression was measured within user-selected binned distances from the structure to investigate cell localization. This can be visualized as a heatmap, if 2 bins are used to represent "near" or "far" from a border. Dividing distance from a structure into user-selected bins essentially turns the problem into a comparison of cell composition or marker expression between bins. While simple, this approach discretizes continuous data, and results can vary depending on the bins used. In particular, it is hard to ascertain whether there truly is a continuous trend in the change in cell composition or marker expression toward a border. The use of a point process model as described in the previous section may prove to be suitable for analyzing the spatial dependence of cells or markers from a structure, but further investigation is required to assess the robustness of such measurements.

Keren et al. hypothesized that there are differences in the cell phenotype near or

far from the tumor-immune border in "compartmentalized" patients [55]. To investigate this, the authors applied a cut-off of 39 µm to stratify cells as being close to or far from the border. In addition to counting cells, they observed whether marker expression was higher or lower away from the border (Figure 1.1G). The authors observed that the ratio of H3K27me3 (methylated DNA) to H3K9ac (acetylated DNA) increased for tumor cells that are far from the border in two patients, indicating that cells closer to the border may be more transcriptionally active. However, this strategy fails to provide any insight as to how this marker ratio varied continuously as a function of distance from the border. It would be interesting to see whether or not the marker ratio increased with distance, coinciding with the binned approach, or whether it alternates between increasing and decreasing. Furthermore, to simplify the spatial relationships observed, a principal component analysis was performed, revealing a subset of patients that had increased immunoregulatory protein expression in CD11c+CD11b+ immune cells. This is suggestive of myeloid derived suppressor cells, which may inhibit the immune response [129]. Hence, the examination of a spatial binning to analyse the spatial dependence of marker expression from a structure was able to reveal subgroups of patients with unique phenotypes. An interesting progression may be to compare how the survival varies between these subgroups.

What Is the Role of the Cell Microenvironment in a Diseased Setting? Multiparameter imaging provides the opportunity for cellular microenvironments to be examined within a diseased setting.

Spatial variance component analysis (SVCA) [130] is a technique that has been applied to MCI data which allows the sources of variation of gene or protein markers in an image to be identified, without the need for cell classification. The sources of variation of cell markers are decomposed into intrinsic effects, environmental effects, and cell-cell interactions. SVCA was applied by Park et al. to investigate how multiple sclerosis (MS) brain lesion environments influence variations in cell marker expression [131]. They found that toward the center of a lesion, the relative influence of intrinsic and environmental effects increased, while the relative influence of cell-cell interactions had decreased. The authors suggest that cells in the lesion rim are more responsive to cues from the microenvironment, such as cytokines or receptor-ligand interactions, while cells respond to cell-intrinsic programs in the lesion center. There are also additional methods proposed for measuring associations between cell microenvironment and marker expression [49, 132].

Another approach for identifying patterns of multicellular architecture is presented by Jackson et al. [133], where the authors define multicellular units as communities. These communities consist of multiple interactions between one or more different cell phenotypes. The authors identified these communities in breast cancer tissue using a graph-based approach implementing the Louvain algorithm. From this analysis, tumour communities and microenvironment communities were identified. These communities were then grouped phenotypically using PhenoGraph. Importantly, this analysis had led to the identification of different subgroups with distinct clinical outcomes.

Another approach is presented with the Local Indicators of Spatial Association Clustering (lisaClust) R package [134]. Here, the authors defined tissue microenvironments as regions of tissue enriched for a combination of certain cell types. The approach of lisaClust involves modelling cells as a multi-type Poisson point process model. This model is then used to estimate 'local indicators of spatial association', defined as the contribution of each cell to the model. Clustering is then used to group cells by their 'local indicators

of spatial association', leading to the identification of unique microenvironments. This was applied to the Damond et al. IMC diabetes dataset [59], as well as a CODEX spleen dataset [135], revealing unique tissue microenvironments.

Ultimately such analysis approaches can provide insight on the role of the microenvironment within a diseased state.

Opportunities in Image Analysis

Through image processing and analysis, researchers are in a position to interrogate high parameter image data in a single-cell manner. This approach allows key clinical and biological questions to be explored and answered, providing insight on the cellular dynamics that are present in the diseased context. In addition, these results can inform further experimentation within or outside the cytometry setting.

There is potential for the development of statistical tests to identify associations between disease outcomes and the spatial relationships between cells, implementing spatial information with multiple markers. Current methods are able to classify cells, but still perform simple spatial analysis that is implemented in other imaging cytometry assays [125]. Complex machine learning algorithms will eventually benefit from including both spatial and marker information provided by high parameter imaging, constructing predictive models in a higher dimensional space.

Deep learning has become a well-established tool for image analysis. Its consistent use in a variety of applications has been driven by its ability to deconstruct and model highly complex images [102–104, 108]. However, deep learning methods require many observations to train effective models. Deep learning is ideal for cell type prediction, where thousands of cells can be trained from a single image. Though, it is unclear whether it will be effective for classifying heterogeneous global spatial interactions in datasets with relatively small sample sizes, as observed in many exploratory clinical studies. Such approaches may become useful in large cohorts generated after high parameter imaging has been implemented in routine clinical use, allowing for improved accuracy.

There is still an exciting opportunity to develop analytic algorithms for summarizing spatial cell-cell interaction relationships into simple, easy to interpret summary statistics. Such algorithms are characterized by the discussed methods for tumor-immune mixing quantification and pseudotime analysis, which stratify patients into risk groups or assigns groups to a disease progression gradient. It is important to simplify such complex relationships as it will allow scores or statistics to be developed for interpretable decision making. This may also facilitate the ability for image data to be included in disease risk scores, incorporating the data with other clinical and pathological and genetic information [55, 106, 111, 112].

1.2.6 Concluding Remarks

Despite its recency, high parameter imaging has already been adopted in diverse contexts ranging from oncology to autoimmunity where it has shown promise for predicting clinical outcome and understanding the role of the immune system in disease progression. Underlying these studies are common questions relating to the composition, phenotype and location of cell subsets and how they interact. Given the fundamental similarities, these studies also share similar computational strategies which we have linked to the general biological questions they answer. Nevertheless, such computational strategies continue to remain limited, with researchers often requiring custom solutions to analysis problems.

This thesis will aim to introduce new techniques which can be generalized to various high parameter imaging modalities, applying this to a HIV CyCIF dataset. Furthermore, novel analysis techniques are also presented in the analysis of the dataset.

1.3 Summary and Thesis Aims

To date, the events of early HIV transmission in colorectal tissue has been understudied due to lack of clinically relevant tissue and viral strian access. Furthermore, traditional microscopy methods have not been able to visualize the diversity of HIV target cells. To address this, the host lab has developed an innovative experimental approach, with access to fresh human colorectal tissue. This is then topically infected with transmitted founder strains of HIV, simulating a real-world HIV transmission event. Cyclic immunofluorescence and RNA scope *in situ* hybridization are then utilized to visualize HIV and its interactions with key HIV target cells. While this approach provides a unique opportunity to visualize HIV and its early interaction with the mucosal immune system, there is an additional challenge with regards to the analysis of the imaging data obtained. To this end, the aims of this thesis are two-fold:

1. To develop a computational pipeline for analyzing high parameter CyCIF images of early HIV transmission in human colorectal tissue.

2. To develop new computational tools to facilitate the analysis of high parameter microscopy images.

The first aim serves as the biological focus and motivation of this thesis. With the complexity of the data being produced by the host lab, as well as the increasing development high parameter imaging techniques, the development of novel computational tools for image analysis is necessitated. Chapter 2 will present a publication which introduces a novel algorithm for identifying and removing autofluorescence from fluorescent microscopy images. Such analysis is necessary in order to robustly identify both HIV virions as well as target cells within a fluorescence image. Secondly, Chapter 3 presents a publication which introduces a novel spatial analysis technique for identifying changes in co-localization between cell types. Finally, Chapter 4 will introduce an image analysis pipeline used in the analysis of the HIV CyCIF image dataset.

Table 1.3: Summary of analytical questions with clinical examples and the techniques used to answer these questions.

Analytical question	Clinical example	Analytical technique
How does cell composition change with disease context?	How does cell composition change with type-1 diabetes progression? [59, 60]	Measurements such as cell counts, cell proportions, or cell densities can be used to compare between different disease contexts Pearson's correlation of the above measurements can be used to identify the co-occurrence or anti-occurrence of cell types Cell types can be considered present or not present within an image if the cell count is greater than a given cut-off (e.g., 10 cells). A chi-square test can then be used to identify cell type co-occurrence
Does marker expression or co-expression change with diseased context?	How does islet marker expression change with type 1 disease progression? [59]	Heatmaps can be utilized for visualizing marker changes across images Markers can be considered present or not present within an image. A chi-square test can then be used to identify marker co-occurrence Pseudotime analysis such as SCORPIUS [117, 118] allow marker changes associated with cell dynamic processes to be investigated
Does cell or structural morphology change with diseased context?	Does islet morphology change with disease progression? [59]	Morphology measurements can be identified using image analysis software such as histoCAT [99], CellProfiler [107], and ImageJ [121]

Are there any interactions between specific cell types, and does this change with disease context?	Are tumor-immune interactions present and significant within tissue compared to immune-immune interactions? [55, 99]	Neighborhood analysis using histoCAT [99], or by setting a distance cut-off to define neighbors [55], can be used to identify cell interaction or avoidance, visualized with a heatmap Marked point process models using the R package "spatstat" can be used to determine cell co-localization or anti-co-localization [127, 128]
Do cells localize to histological structures and does this vary with disease context?	In breast cancer sections that exhibit compartmentalized structures, are there differences in marker expression with distance from the tumor-immune boundary? [55]	Within binned distances away from a histological boundary, differences in cell composition [59, 60] or marker expression [55] can be identified Marked point process models using the R package "spatstat" can be used to explore the distribution of cells as a function of distance from a histological boundary [127, 128]
What is the role of the cell microenvironment in a diseased setting?	In multiple sclerosis brain lesions, how does the environment influence variations in cell marker expression? [131]	Spatial variance component analysis [130] can be used to decompose the sources of variation of a marker into intrinsic effects, environmental effects, and cell-cell interactions

Bibliography

[1] UNAIDS. Unaids data 2021. https://www.unaids.org/en/resources/documents/2021/2021_unaids_data, 2021. Accessed: 2022-12-20.

[2] Gary Maartens, Connie Celum, and Sharon R Lewin. Hiv infection: epidemiology, pathogenesis, treatment, and prevention. *The Lancet*, 384(9939):258–271, 2014.

[3] Skye McGregor, Hamish McManus, and Richard Gray. Hiv, viral hepatitis and sexually transmissible infections in australia. 2016.

[4] HongKui Deng, Rong Liu, Wilfried Ellmeier, Sunny Choe, Derya Unutmaz, Michael Burkhart, Paola Di Marzio, Shoshana Marmon, Richard E Sutton, C Mark Hill, et al. Identification of a major co-receptor for primary isolates of hiv-1. *Nature*, 381 (6584):661, 1996.

[5] Conrad C Bleul, Michael Farzan, Hyeryun Choe, Cristina Parolin, Ian Clark-Lewis, Joseph Sodroski, and Timothy A Springer. The lymphocyte chemoattractant sdf-1 is a ligand for lestr/fusin and blocks hiv-1 entry. *Nature*, 382(6594):829, 1996.

[6] Conrad C Bleul, Lijun Wu, James A Hoxie, Timothy A Springer, and Charles R Mackay. The hiv coreceptors cxcr4 and ccr5 are differentially expressed and regulated on human t lymphocytes. *Proceedings of the National Academy of Sciences*, 94(5):1925–1930, 1997.

[7] H Schuitemaker, M Koot, NA Kootstra, M Wouter Dercksen, RE De Goede, RP Van Steenwijk, JM Lange, JK Schattenkerk, F Miedema, and M Tersmette. Biological phenotype of human immunodeficiency virus type 1 clones at different stages of infection: progression of disease is associated with a shift from monocytotropic to t-cell-tropic virus population. *Journal of virology*, 66(3):1354–1360, 1992.

[8] Teunis BH Geijtenbeek, Douglas S Kwon, Ruurd Torensma, Sandra J van Vliet, Gerard CF van Duijnhoven, Jeena Middel, Ine LMHA Cornelissen, Hans SLM Nottet, Vineet N KewalRamani, Dan R Littman, et al. Dc-sign, a dendritic cell–specific hiv-1-binding protein that enhances trans-infection of t cells. *Cell*, 100(5):587–597, 2000.

[9] Stuart G Turville, Jim Arthos, Kelli Mac Donald, Garry Lynch, Hassan Naif, Georgina Clark, Derek Hart, and Anthony L Cunningham. Hiv gp120 receptors on human dendritic cells. *Blood*, 98(8):2482–2488, 2001.

[10] Kevin B Gurney, Julie Elliott, Hoorig Nassanian, Carol Song, Elizabeth Soilleux, Ian McGowan, Peter A Anton, and Benhur Lee. Binding and transfer of human immunodeficiency virus by dc-sign+ cells in human rectal mucosa. *Journal of virology*, 79(9):5762–5773, 2005.

[11] Joey Lai, Oliver K Bernhard, Stuart G Turville, Andrew N Harman, John Wilkinson, and Anthony L Cunningham. Oligomerization of the macrophage mannose receptor enhances gp120-mediated binding of hiv-1. *Journal of Biological Chemistry*, 284(17):11027–11038, 2009.

[12] Najla Nasr, Joey Lai, Rachel A Botting, Sarah K Mercier, Andrew N Harman, Min Kim, Stuart Turville, Rob J Center, Teresa Domagala, Paul R Gorry, et al. Inhibition of two temporal phases of hiv-1 transfer from primary langerhans cells to t cells: the role of langerin. *The Journal of Immunology*, page 1400630, 2014.

[13] Nuria Izquierdo-Useros, Maier Lorizate, Maria C Puertas, Maria T Rodriguez-Plata, Nadine Zangger, Elina Erikson, Maria Pino, Itziar Erkizia, Bärbel Glass, Bonaventura Clotet, et al. Siglec-1 is a novel dendritic cell receptor that mediates hiv-1 trans-infection through recognition of viral membrane gangliosides. *PLoS biology*, 10(12):e1001448, 2012.

[14] James Arthos, Claudia Cicala, Elena Martinelli, Katilyn Macleod, Donald Van Ryk, Danlan Wei, Zhen Xiao, Timothy D Veenstra, Thomas P Conrad, Richard A Lempicki, et al. Hiv-1 envelope protein binds to and signals through integrin α 4 β 7, the gut mucosal homing receptor for peripheral t cells. *Nature immunology*, 9(3): 301, 2008.

[15] José Marcio Neves Jorge and Angelita Habr-Gama. Anatomy and embryology of the colon, rectum, and anus. In *The ASCRS textbook of colon and rectal surgery*, pages 1–22. Springer, 2007.

[16] Andrew N Harman, Min Kim, Najla Nasr, Kerrie J Sandgren, and Paul U Cameron. Tissue dendritic cells as portals for hiv entry. *Reviews in medical virology*, 23(5): 319–333, 2013.

[17] Rebecca F Baggaley, Dobromir Dimitrov, Branwen N Owen, Michael Pickles, Ailsa R Butler, Ben Masse, and Marie-Claude Boily. Heterosexual anal intercourse: a neglected risk factor for hiv? *American journal of reproductive immunology*, 69: 95–105, 2013.

[18] Othell Begay, Ninochka Jean-Pierre, Ciby J Abraham, Anne Chudolij, Samantha Seidor, Aixa Rodriguez, Brian E Ford, Marcus Henderson, David Katz, Thomas Zydowsky, et al. Identification of personal lubricants that can cause rectal epithelial cell damage and enhance hiv type 1 replication in vitro. *AIDS research and human retroviruses*, 27(9):1019–1024, 2011.

[19] Mariangela Cavarelli, Chiara Foglieni, Maria Rescigno, and Gabriella Scarlatti. R5 hiv-1 envelope attracts dendritic cells to cross the human intestinal epithelium and sample luminal virions via engagement of the ccr5. *EMBO molecular medicine*, 5 (5):776–794, 2013.

[20] Barbara L Shacklett and Peter A Anton. Hiv infection and gut mucosal immune function: updates on pathogenesis with implications for management and intervention. *Current infectious disease reports*, 12(1):19–27, 2010.

[21] Mary Pat Moyer and Howard E Gendelman. Hiv replication and persistence in human gastrointestinal cells cultured in vitro. *Journal of leukocyte biology*, 49(5): 499–504, 1991.

[22] AGNES-LAURENCE Chenine, EVA MATOUSKOVA, GISELLE SANCHEZ, JOSEF REISCHIG, LUBOSLAVA PAVLIKOVA, CAROLE LeCONTEL, JEAN-CLAUDE CHERMANN, and IVAN HIRSCH. Primary intestinal epithelial cells

can be infected with laboratory-adapted strain hiv type 1 ndk but not with clinical primary isolates. *AIDS research and human retroviruses*, 14(14):1235–1238, 1998.

[23] Grigorios Fotopoulos, Alexandre Harari, Pierre Michetti, Didier Trono, Giuseppe Pantaleo, and Jean-Pierre Kraehenbuhl. Transepithelial transport of hiv-1 by m cells is receptor-mediated. *Proceedings of the National Academy of Sciences*, 99 (14):9410–9414, 2002.

[24] Milena Bogunovic, Florent Ginhoux, Julie Helft, Limin Shang, Daigo Hashimoto, Melanie Greter, Kang Liu, Claudia Jakubzick, Molly A Ingersoll, Marylene Leboeuf, et al. Origin of the lamina propria dendritic cell network. *Immunity*, 31(3):513–525, 2009.

[25] GM Orloff, SL Orloff, MS Kennedy, PJ Maddon, and JS McDougal. Penetration of cd4 t cells by hiv-1. the cd4 receptor does not internalize with hiv, and cd4-related signal transduction events are not required for entry. *The Journal of immunology*, 146(8):2578–2587, 1991.

[26] Nicolas Chomont, Mohamed El-Far, Petronela Ancuta, Lydie Trautmann, Francesco A Procopio, Bader Yassine-Diab, Genevieve Boucher, Mohamed-Rachid Boulassel, Georges Ghattas, Jason M Brenchley, et al. Hiv reservoir size and persistence are driven by t cell survival and homeostatic proliferation. *Nature medicine*, 15(8):893, 2009.

[27] Peter A Anton, Julie Elliott, Michael A Poles, Ian M McGowan, Jose Matud, Lance E Hultin, Kathie Grovit-Ferbas, Charles R Mackay, Irvin SY Chen, and Janis V Giorgi. Enhanced levels of functional hiv-1 co-receptors on human mucosal t cells demonstrated using intestinal biopsy tissue. *Aids*, 14(12):1761–1765, 2000.

[28] Claudia Cicala, Elena Martinelli, Jonathan P McNally, Diana J Goode, Ravindra Gopaul, Joseph Hiatt, Katija Jelicic, Shyamasundaran Kottilil, Katilyn Macleod, Angeline O'Shea, et al. The integrin $\alpha 4\beta 7$ forms a complex with cell-surface cd4 and defines a t-cell subset that is highly susceptible to infection by hiv-1. *Proceedings of the National Academy of Sciences*, 106(49):20877–20882, 2009.

[29] Dror Kolodkin-Gal, Sandrine L Hulot, Birgit Korioth-Schmitz, Randi B Gombos, Yi Zheng, Joshua Owuor, Michelle A Lifton, Christian Ayeni, Robert M Najarian, Wendy W Yeh, et al. Efficiency of cell-free and cell-associated virus in mucosal transmission of hiv-1/siv. *Journal of virology*, pages JVI–03108, 2013.

[30] Patricia Ribeiro dos Santos, Magali Rancez, Jean-Luc Prétet, Alice Michel-Salzat, Valérie Messent, Anna Bogdanova, Anne Couëdel-Courteille, Evelyne Souil, Rémi Cheynier, and Cécile Butor. Rapid dissemination of siv follows multisite entry after rectal inoculation. *PloS one*, 6(5):e19493, 2011.

[31] Daniel J Stieh, Edgar Matias, Huanbin Xu, Angela J Fought, James L Blanchard, Preston A Marx, Ronald S Veazey, and Thomas J Hope. Th17 cells are preferentially infected very early after vaginal transmission of siv in macaques. *Cell host & microbe*, 19(4):529–540, 2016.

[32] Lyle R McKinnon, Billy Nyanga, Duncan Chege, Preston Izulla, Makobu Kimani, Sanja Huibner, Lawrence Gelmon, Katharine E Block, Claudia Cicala, A Omu Anzala, et al. Characterization of a human cervical cd4+ t cell subset coexpressing multiple markers of hiv susceptibility. *The Journal of Immunology*, 187(11):6032–6042, 2011.

[33] Annie Gosselin, Tomas Raul Wiche Salinas, Delphine Planas, Vanessa S Wacleche, Yuwei Zhang, Rémi Fromentin, Nicolas Chomont, Éric A Cohen, Barbara Shacklett, Vikram Mehraj, et al. Hiv persists in ccr6+ cd4+ t cells from colon and blood during antiretroviral therapy. *AIDS (London, England)*, 31(1):35, 2017.

[34] Amy E Baxter, Rebecca A Russell, Christopher JA Duncan, Michael D Moore, Christian B Willberg, Jose L Pablos, Andrés Finzi, Daniel E Kaufmann, Christina Ochsenbauer, John C Kappes, et al. Macrophage infection via selective capture of hiv-1-infected cd4+ t cells. *Cell host & microbe*, 16(6):711–721, 2014.

[35] Anna Bujko, Nader Atlasy, Ole JB Landsverk, Lisa Richter, Sheraz Yaqub, Rune Horneland, Ole Øyen, Einar Martin Aandahl, Lars Aabakken, Hendrik G Stunnenberg, et al. Transcriptional and functional profiling defines human small intestinal macrophage subsets. *Journal of Experimental Medicine*, 215(2):441–458, 2018.

[36] Matthew Collin, Naomi McGovern, and Muzlifah Haniffa. Human dendritic cell subsets. *Immunology*, 140(1):22–30, 2013.

[37] L Richter, Ole JB Landsverk, Nader Atlasy, Anna Bujko, Sheraz Yaqub, Rune Horneland, Ole Øyen, Einar Martin Aandahl, Knut Erik Aslaksen Lundin, Hendrik G Stunnenberg, et al. Transcriptional profiling reveals monocyte-related macrophages phenotypically resembling dc in human intestine. *Mucosal immunology*, 11(5):1512–1523, 2018.

[38] Andrew N Harman, Marianne Kraus, Chris R Bye, Karen Byth, Stuart G Turville, Owen Tang, Sarah K Mercier, Najla Nasr, Josh L Stern, Barry Slobedman, et al. Hiv-1–infected dendritic cells show 2 phases of gene expression changes, with lysosomal enzyme activity decreased during the second phase. *Blood*, 114(1):85–94, 2009.

[39] David McDonald, Li Wu, Stacy M Bohks, Vineet N KewalRamani, Derya Unutmaz, and Thomas J Hope. Recruitment of hiv and its receptors to dendritic cell-t cell junctions. *Science*, 300(5623):1295–1297, 2003.

[40] Laura Burleigh, Pierre-Yves Lozach, Cécile Schiffer, Isabelle Staropoli, Valérie Pezo, Françoise Porrot, Bruno Canque, Jean-Louis Virelizier, Fernando Arenzana-Seisdedos, and Ali Amara. Infection of dendritic cells (dcs), not dc-sign-mediated internalization of human immunodeficiency virus, is required for long-term transfer of virus to t cells. *Journal of virology*, 80(6):2949–2957, 2006.

[41] Ruizhong Shen, Lesley E Smythies, Ronald H Clements, Lea Novak, and Phillip D Smith. Dendritic cells transmit hiv-1 through human small intestinal mucosa. *Journal of leukocyte biology*, 87(4):663–670, 2010.

[42] Ruizhong Shen, John C Kappes, Lesley E Smythies, Holly E Richter, Lea Novak, and Phillip D Smith. Vaginal myeloid dendritic cells transmit founder hiv-1. *Journal of virology*, pages JVI–00766, 2014.

[43] Pavel Zrazhevskiy and Xiaohu Gao. Quantum dot imaging platform for single-cell molecular profiling. *Nature Publishing Group*, 4(1):1619–1619, 2013. doi: 10.1038/ncomms2635. URL https://doi.org/10.1038/ncomms2635.

[44] Bernd Bodenmiller. Multiplexed epitope-based tissue imaging for discovery and healthcare applications. *Cell Press*, 2(4):225–238, 2016. doi: 10.1016/j.cels.2016.03.008. URL https://doi.org/10.1016/j.cels.2016.03.008.

[45] Jia-Ren Lin, Mohammad Fallahi-Sichani, and Peter K. Sorger. Highly multiplexed imaging of single cells using a high-throughput cyclic immunofluorescence method. *Nature Publishing Group*, 6(1):8390–8390, 2015. doi: 10.1038/ncomms9390. URL https://doi.org/10.1038/ncomms9390.

[46] Jia-Ren Lin, Benjamin Izar, Shu Wang, Clarence Yapp, Shaolin Mei, Parin Shah, Sandro Santagata, and Peter K. Sorger. Highly multiplexed immunofluorescence imaging of human tissues and tumors using t-cycif and conventional optical microscopes. *eLife Sciences Publications*, 7(NA):NA–NA, 2018. doi: 10.7554/elife.31657. URL https://doi.org/10.7554/elife.31657.

[47] Takahiro Tsujikawa, Sushil Kumar, Rohan N. Borkar, Vahid Azimi, Guillaume Thibault, Young Hwan Chang, Ariel Balter, Rie Kawashima, Gina Choe, David Sauer, Edward El Rassi, Daniel R. Clayburgh, Molly Kulesz-Martin, Eric R. Lutz, Lei Zheng, Elizabeth M. Jaffee, Patrick Leyshock, Adam A. Margolin, Motomi Mori, Joe W. Gray, Paul W. Flint, and Lisa M. Coussens. Quantitative multiplex immunohistochemistry reveals myeloid-inflamed tumor-immune complexity associated with poor prognosis. *Cell Press*, 19(1):203–217, 2017. doi: 10.1016/j.celrep.2017.03.037. URL https://doi.org/10.1016/j.celrep.2017.03.037.

[48] Gabriele Gut, Markus D. Herrmann, and Lucas Pelkmans. Multiplexed protein maps link subcellular organization to cellular states. *American Association for the Advancement of Science*, 361(6401):NA–NA, 2018. doi: 10.1126/science.aar7042. URL https://doi.org/10.1126/science.aar7042.

[49] Yury Goltsev, Nikolay Samusik, Julia Kennedy-Darling, Salil S. Bhate, Matthew Hale, Gustavo Vazquez, Sarah Black, and Garry P. Nolan. Deep profiling of mouse splenic architecture with codex multiplexed imaging. *Cell Press*, 174(4):968–981, 2018. doi: 10.1016/j.cell.2018.07.010. URL https://doi.org/10.1016/j.cell.2018.07.010.

[50] Michael J. Gerdes, Christopher J. Sevinsky, Anup Sood, Sudeshna Adak, Musodiq Bello, Alexander Bordwell, Ali Can, Corwin Alex D, Sean Richard Dinn, Robert John Filkins, Denise Hollman, Vidya Pundalik Kamath, Sireesha Kaanumalle, Kevin Bernard Kenny, Melinda Larsen, Michael Lazare, Qing Li, Christina Lowes, Colin Craig McCulloch, Elizabeth McDonough, Michael Christopher Montalto, Zhengyu Pang, Jens Rittscher, Alberto Santamaria-Pang, Brion Daryl Sarachan, Maximilian Lewis Seel, Antti Seppo, Kashan Shaikh, Yunxia Sui, Jingyu Zhang, and Fiona Ginty. Highly multiplexed single-cell analysis of formalin-fixed,

paraffin-embedded cancer tissue. *Proceedings of the National Academy of Sciences*, 110(29):11982–11987, 2013. doi: 10.1073/pnas.1300136110. URL https://doi.org/10.1073/pnas.1300136110.

[51] Eliot T. McKinley, Yunxia Sui, Yousef Al-Kofahi, Bryan A. Millis, Matthew J. Tyska, Joseph T. Roland, Alberto Santamaria-Pang, Christina L. Ohland, Christian Jobin, Jeffrey L. Franklin, Ken S. Lau, Michael J. Gerdes, and Robert J. Coffey. Optimized multiplex immunofluorescence single-cell analysis reveals tuft cell heterogeneity. *American Society for Clinical Investigation*, 2(11):NA–NA, 2017. doi: 10.1172/jci.insight.93487. URL https://doi.org/10.1172/jci.insight.93487.

[52] Michael Y. Gerner, Wolfgang Kastenmüller, Ina Ifrim, Juraj Kabat, and Ronald N. Germain. Histo-cytometry: A method for highly multiplex quantitative tissue imaging analysis applied to dendritic cell subset microanatomy in lymph nodes. *Cell Press*, 37(2):364–376, 2012. doi: 10.1016/j.immuni.2012.07.011. URL https://doi.org/10.1016/j.immuni.2012.07.011.

[53] Charlotte Giesen, Hao A. O. Wang, Denis Schapiro, Nevena Zivanovic, Andrea Jacobs, Bodo Hattendorf, Peter J. Schüffler, Daniel Grolimund, Joachim M. Buhmann, Simone Brandt, Zsuzsanna Varga, Peter J. Wild, Detlef Günther, and Bernd Bodenmiller. Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nature Publishing Group*, 11(4):417–422, 2014. doi: 10.1038/nmeth.2869. URL https://doi.org/10.1038/nmeth.2869.

[54] Michael Angelo, Sean C. Bendall, Rachel Finck, Matthew Hale, Chuck Hitzman, Alexander D. Borowsky, Richard M. Levenson, John B. Lowe, Scot D. Liu, Shuchun Zhao, Yasodha Natkunam, and Garry P. Nolan. Multiplexed ion beam imaging of human breast tumors. *Nature Publishing Group*, 20(4):436–442, 2014. doi: 10.1038/nm.3488. URL https://doi.org/10.1038/nm.3488.

[55] Leeat Keren, Marc Bosse, Diana M. Marquez, Roshan Angoshtari, Samir Jain, Sushama Varma, Soo-Ryum Yang, Allison W. Kurian, David Van Valen, Robert B. West, Sean C. Bendall, and Michael Angelo. A structured tumor-immune microenvironment in triple negative breast cancer revealed by multiplexed ion beam imaging. *Cell Press*, 174(6):1373–1387, 2018. doi: 10.1016/j.cell.2018.08.039. URL https://doi.org/10.1016/j.cell.2018.08.039.

[56] Paymaneh D. Malihi, Michael Morikado, Lisa Welter, Sandy T. Liu, Eric T. Miller, Radu M. Cadaneanu, Beatrice S. Knudsen, Michael S. Lewis, Anders Carlsson, Carmen Ruiz Velasco, Anand Kolatkar, Mariam Rodriguez-Lee, Isla P. Garraway, James W. Hicks, and Peter Kuhn. Clonal diversity revealed by morphoproteomic and copy number profiles of single prostate cancer cells at diagnosis. *IOP Publishing*, 4(1):015003–NA, 2018. doi: 10.1088/2057-1739/aaa00b. URL https://doi.org/10.1088/2057-1739/aaa00b.

[57] Erik Gerdtsson, Milind Pore, Jana-Aletta Thiele, Anna S. Gerdtsson, Paymaneh D. Malihi, Rafael Nevarez, Anand Kolatkar, Carmen Ruiz Velasco, Sophia Wix, Mohan Singh, Anders Carlsson, Amado J. Zurita, Christopher J. Logothetis, Akil Merchant, James W. Hicks, and Peter Kuhn. Multiplex protein detection on circulating tumor cells from liquid biopsies using imaging mass cytometry. *IOP Publishing*, 4(1):

015002–NA, 2018. doi: 10.1088/2057-1739/aaa013. URL https://doi.org/10.1088/2057-1739/aaa013.

[58] Daniel Schulz, Vito R. T Zanotelli, Jana R. Fischer, Denis Schapiro, Stefanie Engler, Xiao-Kang Lun, Hartland W. Jackson, and Bernd Bodenmiller. Simultaneous multiplexed imaging of mrna and proteins with subcellular resolution in breast cancer tissue samples by mass cytometry. *Cell Press*, 6(1):531–531, 2017.

[59] Nicolas Damond, Stefanie Engler, Vito R. T Zanotelli, Denis Schapiro, Clive Wasserfall, Irina Kusmartseva, Harry S. Nick, Fabrizio Thorel, Pedro Luis Herrera, Mark A. Atkinson, and Bernd Bodenmiller. A map of human type 1 diabetes progression by imaging mass cytometry. *Cell Press*, 29(3):755–768, 2019. doi: 10.1016/j.cmet.2018.11.014. URL https://doi.org/10.1016/j.cmet.2018.11.014.

[60] Yue J. Wang, Daniel Traum, Jonathan Schug, Long Gao, Chengyang Liu, Mark A. Atkinson, Alvin C. Powers, Michael Feldman, Ali Naji, Kyong-Mi Chang, and Klaus H. Kaestner. Multiplexed in situ imaging mass cytometry analysis of the human endocrine pancreas and immune system in type 1 diabetes. *Cell Press*, 29 (3):769–783.e4, 2019. doi: 10.1016/j.cmet.2019.01.003. URL https://doi.org/10.1016/j.cmet.2019.01.003.

[61] Sebastian Brähler, Bernd H. Zinselmeyer, Saravanan Raju, Maximilian Nitschké, Hani Suleiman, Brian Saunders, Michael W. Johnson, Alexander M.C. Böhner, Susanne F. Viehmann, Derek J. Theisen, Nicole M. Kretzer, Carlos G. Briseño, Konstantin Zaitsev, Olga Ornatsky, Qing Chang, Javier A. Carrero, Jeffrey B. Kopp, Maxim N. Artyomov, Christian Kurts, Kenneth M. Murphy, Jeffrey H. Miner, and Andrey S. Shaw. Opposing roles of dendritic cell subsets in experimental gn. *American Society of Nephrology*, 29(1):138–154, 2017. doi: 10.1681/asn.2017030270. URL https://doi.org/10.1681/asn.2017030270.

[62] Na Li, Vincent van Unen, Tamim Abdelaal, Nannan Guo, Sofya A. Kasatskaya, Kristin Ladell, James E. McLaren, Evgeny S. Egorov, Mark Izraelson, Susana M. Chuva de Sousa Lopes, Thomas Höllt, Olga V. Britanova, Jeroen Eggermont, Noel F. C. C de Miranda, Dmitriy M. Chudakov, David Price, Boudewijn P. F. Lelieveldt, and Frits Koning. Nature immunology - memory cd4(+) t cells are generated in the human fetal intestine. *Nature Publishing Group*, 20(3):301–312, 2019. doi: 10.1038/s41590-018-0294-9. URL https://doi.org/10.1038/s41590-018-0294-9.

[63] Yuan Zhao, Mohamed Uduman, Jacqueline Hy Siu, Thomas J Tull, Jeremy D. Sanderson, Yu-Chang Bryan Wu, Julian Q. Zhou, Nedyalko Petrov, Richard Ellis, Katrina Todd, Konstantia-Maria Chavele, William Guesdon, Anna Vossenkämper, Wayel Jassem, David D'Cruz, David J. Fear, Susan D. John, Dagmar Scheel-Toellner, Claire Hopkins, Estefania Moreno, Natalie Woodman, Francesca D. Ciccarelli, Susanne Heck, Steven H. Kleinstein, Mats Bemark, and Jo Spencer. Spatiotemporal segregation of human marginal zone and memory b cell populations in lymphoid tissue. *Nature Publishing Group*, 9(1):3857–3857, 2018. doi: 10.1038/s41467-018-06089-1. URL https://doi.org/10.1038/s41467-018-06089-1.

[64] Anna Cleta Croce and Giovanni Bottiroli. Autofluorescence spectroscopy and imaging: a tool for biomedical research and diagnosis. *PagePress Publications*, 58(4):

2461–2461, 2014. doi: 10.4081/ejh.2014.2461. URL https://doi.org/10.4081/ejh.2014.2461.

[65] Richard M. Levenson, Alexander D. Borowsky, and Michael Angelo. Immunohistochemistry and mass spectrometry for highly multiplexed cellular molecular imaging. *Nature Publishing Group*, 95(4):397–405, 2015. doi: 10.1038/labinvest.2015.2. URL https://doi.org/10.1038/labinvest.2015.2.

[66] Patrick C. Sachs, Peter A. Mollica, and Robert D. Bruno. Tissue specific microenvironments: a key tool for tissue engineering and regenerative medicine. *BioMed Central*, 11(1):34–34, 2017. doi: 10.1186/s13036-017-0077-0. URL https://doi.org/10.1186/s13036-017-0077-0.

[67] Berend Snijder and Lucas Pelkmans. Origins of regulated cell-to-cell variability. *Nature Publishing Group*, 12(2):119–125, 2011. doi: 10.1038/nrm3044. URL https://doi.org/10.1038/nrm3044.

[68] B. Rani, Y. Cao, Andrea Malfettone, Ciprian Tomuleasa, Isabel Fabregat, and Gianluigi Giannelli. Role of the tissue microenvironment as a therapeutic target in hepatocellular carcinoma. *WJG Press*, 20(15):4128–4140, 2014. doi: 10.3748/wjg.v20.i15.4128. URL https://doi.org/10.3748/wjg.v20.i15.4128.

[69] Jan Brábek, Claudia Tanja Mierke, Daniel Rösel, Pavel Veselý, and Ben Fabry. The role of the tissue microenvironment in the regulation of cancer cell motility and invasion. *Signal Transduction Society*, 8(1):22–22, 2010. doi: 10.1186/1478-811x-8-22. URL https://doi.org/10.1186/1478-811x-8-22.

[70] Amanda J Oliver, Peter Lau, Ashleigh S Unsworth, Sherene Loi, Phillip K. Darcy, Michael H. Kershaw, and Clare Y. Slaney. Tissue-dependent tumor microenvironments and their impact on immunotherapy responses. *Frontiers Media S.A.*, 9 (NA):70–70, 2018. doi: 10.3389/fimmu.2018.00070. URL https://doi.org/10.3389/fimmu.2018.00070.

[71] Frederic Geissmann, Markus G. Manz, Steffen Jung, Michael H. Sieweke, Miriam Merad, and Klaus Ley. Development of monocytes, macrophages, and dendritic cells. *American Association for the Advancement of Science*, 327(5966):656–661, 2010. doi: 10.1126/science.1178331. URL https://doi.org/10.1126/science.1178331.

[72] Martin Guilliams, Charles-Antoine Dutertre, Charlotte L. Scott, Naomi McGovern, Dorine Sichien, Svetoslav Chakarov, Sofie Van Gassen, Jinmiao Chen, Michael Poidinger, Sofie De Prijck, Simon Tavernier, Ivy Low, Sergio Erdal Irac, Citra Nurfarah Zaini Mattar, Hermi Sumatoh, Gillian Low, Tam John Kit Chung, Dedrick Kok Hong Chan, Ker-Kan Tan, Tony Lim Kiat Hon, Even Fossum, Bjarne Bogen, Mahesh Choolani, Jerry Kok Yen Chan, Anis Larbi, Hervé Luche, Sandrine Henri, Yvan Saeys, Evan W. Newell, Bart N. Lambrecht, Bernard Malissen, and Florent Ginhoux. Unsupervised high-dimensional analysis aligns dendritic cells across tissues and species. *Cell Press*, 45(3):669–684, 2016. doi: 10.1016/j.immuni.2016.08.015. URL https://doi.org/10.1016/j.immuni.2016.08.015.

[73] Tomer Granot, Takashi Senda, Dustin J. Carpenter, Nobuhide Matsuoka, Joshua Weiner, Claire L. Gordon, Michelle Miron, Brahma V. Kumar, Adam Griesemer, Siu-hong Ho, Harvey Lerner, Joseph J.C. Thome, Thomas J. Connors, Boris Reizis, and Donna L. Farber. Dendritic cells display subset and tissue-specific maturation dynamics over human life. *Cell Press*, 46(3):504–515, 2017. doi: 10.1016/j.immuni.2017.02.019. URL https://doi.org/10.1016/j.immuni.2017.02.019.

[74] Michael Thomas Wong, David E. Ong, Frances Sheau Huei Lim, Karen Wei Weng Teng, Naomi McGovern, Sriram Narayanan, Wen Qi Ho, Daniela Cerny, Henry K.K. Tan, Rosslyn Anicete, Bien Keem Tan, Tony Kiat Hon Lim, Chung Yip Chan, Peng Chung Cheow, Ser Yee Lee, Angela Takano, Eng Huat Tan, John Kit Chung Tam, Ern Yu Tan, Jerry Kok Yen Chan, Katja Fink, Antonio Bertoletti, Florent Ginhoux, Maria A. Curotto de Lafaille, and Evan W. Newell. A high-dimensional atlas of human t cell diversity reveals tissue-specific trafficking and cytokine signatures. *Cell Press*, 45(2):442–456, 2016. doi: 10.1016/j.immuni.2016.07.007. URL https://doi.org/10.1016/j.immuni.2016.07.007.

[75] Taheri Sathaliyawala, Masaru Kubota, Naomi Yudanin, Damian Turner, Philip Edward Camp, Joseph J.C. Thome, Kara L. Bickham, Harvey Lerner, Michael J. Goldstein, Megan Sykes, Tomoaki Kato, and Donna L. Farber. Distribution and compartmentalization of human circulating and tissue-resident memory t cell subsets. *Cell Press*, 38(1):187–197, 2012. doi: 10.1016/j.immuni.2012.09.020. URL https://doi.org/10.1016/j.immuni.2012.09.020.

[76] Jens Geginat, Moira Paroni, Stefano Maglie, Johanna Sophie Alfen, Ilko Kastirr, Paola Gruarin, Marco De Simone, Massimiliano Pagani, and Sergio Abrignani. Plasticity of human cd4 t cell subsets. *Frontiers Media S.A.*, 5(NA):630–630, 2014. doi: 10.3389/fimmu.2014.00630. URL https://doi.org/10.3389/fimmu.2014.00630.

[77] David Bauché, Barbara Joyce-Shaikh, Renu Jain, Jeff Grein, Karin S. Ku, Wendy M. Blumenschein, Stephanie C. Ganal-Vonarburg, Douglas C. Wilson, Terrill K. McClanahan, Rene De Waal Malefyt, Andrew J. Macpherson, Lakshmanan Annamalai, Jennifer H. Yearley, and Daniel J. Cua. Lag3+ regulatory t cells restrain interleukin-23-producing cx3cr1+ gut-resident macrophages during group 3 innate lymphoid cell-driven colitis. *Cell Press*, 49(2):342–352.e5, 2018. doi: 10.1016/j.immuni.2018.07.007. URL https://doi.org/10.1016/j.immuni.2018.07.007.

[78] Martin Guilliams, Alexander Mildner, and Simon Yona. Developmental and functional heterogeneity of monocytes. *Cell Press*, 49(4):595–613, 2018. doi: 10.1016/j.immuni.2018.10.005. URL https://doi.org/10.1016/j.immuni.2018.10.005.

[79] Jeffrey A. Bluestone, Charles R. Mackay, John J. O'Shea, and Brigitta Stockinger. The functional plasticity of t cell subsets. *Nature Publishing Group*, 9(11):811–816, 2009. doi: 10.1038/nri2654. URL https://doi.org/10.1038/nri2654.

[80] Frances R. Balkwill, Melania Capasso, and Thorsten Hagemann. The tumor microenvironment at a glance. *Company of Biologists Ltd*, 125(23):5591–5596, 2012. doi: 10.1242/jcs.116392. URL https://doi.org/10.1242/jcs.116392.

[81] Michele De Palma, Daniela Biziato, and Tatiana V. Petrova. Microenvironmental regulation of tumour angiogenesis. *Nature Publishing Group*, 17(8):457–474, 2017. doi: 10.1038/nrc.2017.51. URL https://doi.org/10.1038/nrc.2017.51.

[82] Matthew Collin and Venetia Bigley. Human dendritic cell subsets: an update. *Wiley-Blackwell*, 154(1):3–20, 2018. doi: 10.1111/imm.12888. URL https://doi.org/10.1111/imm.12888.

[83] Clare Bycroft, Colin Freeman, Desislava Petkova, Gavin Band, Lloyd T. Elliott, Kevin Sharp, Allan Motyer, Damjan Vukcevic, Olivier Delaneau, Jared O'Connell, Adrian Cortes, Samantha Welsh, Alan Young, Mark Effingham, Gil McVean, Stephen Leslie, Naomi E. Allen, Peter Donnelly, and Jonathan Marchini. The uk biobank resource with deep phenotyping and genomic data. *Nature Publishing Group*, 562(7726):203–209, 2018. doi: 10.1038/s41586-018-0579-z. URL https://doi.org/10.1038/s41586-018-0579-z.

[84] Martha Campbell-Thompson, Clive Wasserfall, John S. Kaddis, Anastasia Albanese-O'Neill, Teodora Staeva, Concepcion Nierras, Jayne M Moraski, Patrick Rowe, Roberto Gianani, George S. Eisenbarth, James M. Crawford, Desmond A. Schatz, Alberto Pugliese, and Mark A. Atkinson. Network for pancreatic organ donors with diabetes (npod): developing a tissue biobank for type 1 diabetes. *John Wiley and Sons Ltd*, 28(7):608–617, 2012. doi: 10.1002/dmrr.2316. URL https://doi.org/10.1002/dmrr.2316.

[85] Werner Baschong, Rosmarie Suetterlin, and R. Hubert Laeng. Control of autofluorescence of archival formaldehyde-fixed, paraffin-embedded tissue in confocal laser scanning microscopy (clsm). *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society*, 49(12):1565–1571, 2001. doi: 10.1177/002215540104901210. URL https://doi.org/10.1177/002215540104901210.

[86] P. Del Castillo, M. L. Molero, J. M. Ferrer, and Juan C. Stockert. Autofluorescence and induced fluorescence in epon embedded tissue sections. *Histochemistry*, 85 (5):439–440, 1986. doi: 10.1007/bf00982676. URL https://doi.org/10.1007/bf00982676.

[87] Jonas Wizenty, Muhammad Imtiaz Ashraf, Nadine Rohwer, Martin Stockmann, Sascha Weiss, Matthias Biebl, Johann Pratschke, Felix Aigner, and Tilo Wuensch. Autofluorescence: A potential pitfall in immunofluorescence-based inflammation grading. *Journal of immunological methods*, 456(NA):28–37, 2018. doi: 10.1016/j.jim.2018.02.007. URL https://doi.org/10.1016/j.jim.2018.02.007.

[88] Nicholas Billinton and Andrew W. Knight. Seeing the wood through the trees: a review of techniques for distinguishing green fluorescent protein from endogenous autofluorescence. *Analytical biochemistry*, 291(2):175–197, 2001. doi: 10.1006/abio.2000.5006. URL https://doi.org/10.1006/abio.2000.5006.

[89] A. Sally Davis, Anke Richter, Steven Becker, Jenna E. Moyer, Aline Sandouk, Jeff Skinner, and Jeffery K. Taubenberger. Characterizing and diminishing autofluorescence in formalin-fixed paraffin-embedded human respiratory tissue. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society*, 62(6):405–423, 2014. doi: 10.1369/0022155414531549. URL https://doi.org/10.1369/0022155414531549.

[90] Michael Neumann and Detlef Gabel. Simple method for reduction of autofluorescence in fluorescence microscopy. *The journal of histochemistry and cytochemistry*

: *official journal of the Histochemistry Society*, 50(3):437–439, 2002. doi: 10.1177/002215540205000315. URL https://doi.org/10.1177/002215540205000315.

[91] Mary E. Dickinson, Gregory H. Bearman, S Tille, Rusty Lansford, and Scott E. Fraser. Multi-spectral imaging and linear unmixing add a whole new dimension to laser scanning fluorescence microscopy. *BioTechniques*, 31(6):1272–1278, 2001. doi: 10.2144/01316bt01. URL https://doi.org/10.2144/01316bt01.

[92] Franco Woolfe, Michael J. Gerdes, Musodiq Bello, Xiaodong Tao, and Ali Can. Autofluorescence removal by non-negative matrix factorization. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, 20(4): 1085–1093, 2010. doi: 10.1109/tip.2010.2079810. URL https://doi.org/10.1109/tip.2010.2079810.

[93] Zhengyu Pang, Eugene Barash, Alberto Santamaria-Pang, Christopher J. Sevinsky, Qing Li, and Fiona Ginty. Autofluorescence removal using a customized filter set. *Microscopy research and technique*, 76(10):1007–1015, 2013. doi: 10.1002/jemt.22261. URL https://doi.org/10.1002/jemt.22261.

[94] Stéphane Chevrier, Helena L. Crowell, Vito R. T Zanotelli, Stefanie Engler, Mark D. Robinson, and Bernd Bodenmiller. Compensation of signal spillover in suspension and imaging mass cytometry. *Cell Press*, 6(5):612–620.e5, 2018. doi: 10.1016/j.cels.2018.02.010. URL https://doi.org/10.1016/j.cels.2018.02.010.

[95] Fai Wang, John G. Flanagan, Nan Su, Li Chong Wang, Son Bui, Allissa Nielson, Xingyong Wu, Hong Thuy Vo, Xiao Jun, and Yuling Luo. Rnascope: A novel in situ rna analysis platform for formalin-fixed, paraffin-embedded tissues. *Association of Molecular Pathology*, 14(1):22–29, 2012. doi: 10.1016/j.jmoldx.2011.08.002. URL https://doi.org/10.1016/j.jmoldx.2011.08.002.

[96] Sean C. Bendall, Garry P. Nolan, Mario Roederer, and Pratip K. Chattopadhyay. A deep profiler's guide to cytometry. *Elsevier Limited*, 33(7):323–332, 2012. doi: 10.1016/j.it.2012.02.010. URL https://doi.org/10.1016/j.it.2012.02.010.

[97] Dmitry Bandura, Vladimir I. Baranov, Olga Ornatsky, Alexei Antonov, Robert Kinach, Xudong Lou, Serguei Pavlov, Sergey Vorobiev, John E. Dick, and Scott D. Tanner. Mass cytometry: technique for real time single cell multitarget immunoassay based on inductively coupled plasma time-of-flight mass spectrometry. *American Chemical Society*, 81(16):6813–6822, 2009. doi: 10.1021/ac901049w. URL https://doi.org/10.1021/ac901049w.

[98] Olga Ornatsky, Robert Kinach, Dmitry R. Bandura, Xudong Lou, Scott D. Tanner, Vladimir Baranov, Mark Nitz, and Mitchell A. Winnik. Development of analytical methods for multiplex bio-assay with inductively coupled plasma mass spectrometry. *Royal Society of Chemistry*, 23(4):463–469, 2008. doi: 10.1039/b710510j. URL https://doi.org/10.1039/b710510j.

[99] Denis Schapiro, Hartland W. Jackson, Swetha Raghuraman, Jana R. Fischer, Vito R. T Zanotelli, Daniel Schulz, Charlotte Giesen, Raul Catena, Zsuzsanna Varga, and Bernd Bodenmiller. histocat: analysis of cell phenotypes and interactions in multiplex image cytometry data. *Nature Publishing Group*, 14(9):873–876, 2017. doi: 10.1038/nmeth.4391. URL https://doi.org/10.1038/nmeth.4391.

[100] Peter J. Schüffler, Denis Schapiro, Charlotte Giesen, Hao A. O. Wang, Bernd Bodenmiller, and Joachim M. Buhmann. Automatic single cell segmentation on highly multiplexed tissue images. *Wiley-Liss Inc.*, 87(10):936–942, 2015. doi: 10.1002/cyto.a.22702. URL https://doi.org/10.1002/cyto.a.22702.

[101] Mélanie Durand, Thomas Walter, Tiphène Pirnay, Thomas Naessens, Paul Gueguen, Christel Goudot, Sonia Lameiras, Qing Chang, Nafiseh Talaei, Olga Ornatsky, Tatiana Vassilevskaia, Sylvain Baulande, Sebastian Amigorena, and Elodie Segura. Human lymphoid organ cdc2 and macrophages play complementary roles in t follicular helper responses. *Rockefeller University Press*, 216(7):1561–1581, 2019. doi: 10.1084/jem.20181994. URL https://doi.org/10.1084/jem.20181994.

[102] Jacob H. Levine, Erin F. Simonds, Sean C. Bendall, Kara L. Davis, El-ad David Amir, Michelle D. Tadmor, Oren Litvin, Harris G. Fienberg, Astraea Jager, Eli R. Zunder, Rachel Finck, Amanda Larson Gedman, Ina Radtke, James R. Downing, Dana Pe'er, and Garry P. Nolan. Data-driven phenotypic dissection of aml reveals progenitor-like cells that correlate with prognosis. *Cell Press*, 162(1):184–197, 2015. doi: 10.1016/j.cell.2015.05.047. URL https://doi.org/10.1016/j.cell.2015.05.047.

[103] Christoph Sommer, Christoph N. Straehle, Ullrich Köthe, and Fred A. Hamprecht. Isbi - ilastik: Interactive learning and segmentation toolkit, 2011. URL https://doi.org/10.1109/isbi.2011.5872394.

[104] Stuart Berg, Dominik Kutra, Thorben Kroeger, Christoph N. Straehle, Bernhard X. Kausler, Carsten Haubold, Martin Schiegg, Janez Ales, Thorsten Beier, Markus Rudy, Kemal Eren, Jaime I Cervantes, Buote Xu, Fynn Beuttenmueller, Adrian Wolny, Chong Zhang, Ullrich Koethe, Fred A. Hamprecht, and Anna Kreshuk. ilastik: interactive machine learning for (bio)image analysis. *Nature Publishing Group*, 16(12):1226–1232, 2019. doi: 10.1038/s41592-019-0582-9. URL https://doi.org/10.1038/s41592-019-0582-9.

[105] David Van Valen, Takamasa Kudo, Keara Michelle Lane, Derek N. Macklin, Nicolas Quach, Mialy M. DeFelice, Inbal Maayan, Yu Tanouchi, Euan A. Ashley, and Markus W. Covert. Deep learning automates the quantitative analysis of individual cells in live-cell imaging experiments. *Public Library of Science*, 12(11):e1005177–NA, 2016. doi: 10.1371/journal.pcbi.1005177. URL https://doi.org/10.1371/journal.pcbi.1005177.

[106] Juan C. Caicedo, Jonathan R. Roth, Allen Goodman, Tim Becker, Kyle W. Karhohs, Matthieu Broisin, Csaba Molnar, Claire McQuin, Shantanu Singh, Fabian J. Theis, and Anne E. Carpenter. Evaluation of deep learning strategies for nucleus segmentation in fluorescence images. *Wiley-Liss Inc.*, 95(9):952–965, 2019. doi: 10.1002/cyto.a.23863. URL https://doi.org/10.1002/cyto.a.23863.

[107] Anne E. Carpenter, Thouis R. Jones, Michael R. Lamprecht, Colin Clarke, In Han Kang, Ola Friman, David A. Guertin, Joo Han Chang, Robert A. Lindquist, Jason Moffat, Polina Golland, and David M. Sabatini. Cellprofiler: image analysis software for identifying and quantifying cell phenotypes. *BioMed Central*, 7(10):1–11, 2006. doi: 10.1186/gb-2006-7-10-r100. URL https://doi.org/10.1186/gb-2006-7-10-r100.

[108] Sofie Van Gassen, Britt Callebaut, Mary J. van Helden, Bart N. Lambrecht, Piet Demeester, Tom Dhaene, and Yvan Saeys. Flowsom: Using self-organizing maps for visualization and interpretation of cytometry data. *Wiley-Liss Inc.*, 87(7):636–645, 2015. doi: 10.1002/cyto.a.22625. URL https://doi.org/10.1002/cyto.a.22625.

[109] Luca Scrucca, Michael Fop, T. Brendan Murphy, and Adrian E. Raftery. mclust 5: Clustering, classification and density estimation using gaussian finite mixture models. *R Foundation for Statistical Computing*, 8(1):289–317, 2016.

[110] Nils Eling, Nicolas Damond, Tobias Hoch, and Bernd Bodenmiller. Cytomapper: an r/bioconductor package for visualisation of highly multiplexed imaging data. *Bioinformatics (Oxford, England)*, 36(24):5706–5708, 2020. doi: 10.1093/bioinformatics/btaa1061. URL https://doi.org/10.1093/bioinformatics/btaa1061.

[111] Sotirios P Fortis, Michael Sofopoulos, Nectaria N. Sotiriadou, Christoforos Haritos, Christoforos Vaxevanis, Eleftheria A. Anastasopoulou, Nicole Janssen, Niki Arnogiannaki, Alexandros Ardavanis, Graham Pawelec, Sonia A. Perez, and Constantin N. Baxevanis. Differential intratumoral distributions of cd8 and cd163 immune cells as prognostic biomarkers in breast cancer. *BioMed Central*, 5(1):39–39, 2017. doi: 10.1186/s40425-017-0248-z. URL https://doi.org/10.1186/s40425-017-0248-z.

[112] Mark A. Atkinson and George S. Eisenbarth. Type 1 diabetes: new perspectives on disease pathogenesis and treatment. *Elsevier Limited*, 358(9277):221–229, 2001. doi: 10.1016/s0140-6736(01)05415-0. URL https://doi.org/10.1016/s0140-6736(01)05415-0.

[113] Mark A. Atkinson, George S. Eisenbarth, and Aaron W. Michels. Type 1 diabetes. *Elsevier Limited*, 383(9911):69–82, 2013. doi: 10.1016/s0140-6736(13)60591-7. URL https://doi.org/10.1016/s0140-6736(13)60591-7.

[114] Kathleen M. Mahoney, Gordon J. Freeman, and David F. McDermott. The next immune-checkpoint inhibitors: Pd-1/pd-l1 blockade in melanoma. *Excerpta Medica*, 37(4):764–782, 2015. doi: 10.1016/j.clinthera.2015.02.018. URL https://doi.org/10.1016/j.clinthera.2015.02.018.

[115] Shawna M. Woollard and Georgette D. Kanmogne. Maraviroc: a review of its use in hiv infection and beyond. *Dove Medical Press Ltd.*, 9(NA):5447–5468, 2015. doi: 10.2147/dddt.s90580. URL https://doi.org/10.2147/dddt.s90580.

[116] Andrew C. Chan and Paul Carter. Therapeutic antibodies for autoimmunity and inflammation. *Nature Publishing Group*, 10(5):301–316, 2010. doi: 10.1038/nri2761. URL https://doi.org/10.1038/nri2761.

[117] Robrecht Cannoodt, Wouter Saelens, and Yvan Saeys. Computational methods for trajectory inference from single-cell transcriptomics. *Wiley-VCH Verlag*, 46(11):2496–2506, 2016. doi: 10.1002/eji.201646347. URL https://doi.org/10.1002/eji.201646347.

[118] Wouter Saelens, Robrecht Cannoodt, Helena Todorov, and Yvan Saeys. A comparison of single-cell trajectory inference methods: towards more accurate and robust tools, 2018. URL https://doi.org/10.1101/276907.

[119] Nicholas Navin, Jude Kendall, Jennifer Troge, Peter Andrews, Linda Rodgers, Jeanne McIndoo, Kerry Cook, Asya Stepansky, Dan Levy, Diane Esposito, Lakshmi Muthuswamy, Alexander Krasnitz, W. Richard McCombie, James W. Hicks, and Michael Wigler. Tumour evolution inferred by single-cell sequencing. *Nature Publishing Group*, 472(7341):90–94, 2011. doi: 10.1038/nature09807. URL https://doi.org/10.1038/nature09807.

[120] William C. Cho. Proteomics in translational cancer research: biomarker discovery for clinical applications. *Taylor and Francis Ltd.*, 11(2):131–133, 2014. doi: 10.1586/14789450.2014.899908. URL https://doi.org/10.1586/14789450.2014.899908.

[121] Caroline A Schneider, Wayne Rasband, and Kevin W. Eliceiri. Nih image to imagej: 25 years of image analysis. *Nature Publishing Group*, 9(7):671–675, 2012. doi: 10.1038/nmeth.2089. URL https://doi.org/10.1038/nmeth.2089.

[122] Vebjorn Ljosa, Peter D. Caie, Rob ter Horst, Katherine L. Sokolnicki, Emma Jenkins, Sandeep Daya, Mark E. Roberts, Thouis R. Jones, Shantanu Singh, Auguste Genovesio, Paul A. Clemons, Neil O. Carragher, and Anne E. Carpenter. Comparison of methods for image-based profiling of cellular morphological responses to small-molecule treatment. *SAGE Publications Inc.*, 18(10):1321–1329, 2013. doi: 10.1177/1087057113503553. URL https://doi.org/10.1177/1087057113503553.

[123] Michael Boutros, Florian Heigwer, and Christina Laufer. Microscopy-based high-content screening. *Cell Press*, 163(6):1314–1325, 2015. doi: 10.1016/j.cell.2015.11.007. URL https://doi.org/10.1016/j.cell.2015.11.007.

[124] Masahiro Tanaka, Raynard L. Bateman, Daniel Rauh, Eugeni A. Vaisberg, Shyam Ramachandani, Chao Zhang, Kirk C. Hansen, Alma L. Burlingame, Jay Trautman, Kevan M. Shokat, and Cynthia L. Adams. An unbiased cell morphology-based screen for new, biologically active small molecules. *Public Library of Science*, 3 (5):128–128, 2005. doi: 10.1371/journal.pbio.0030128. URL https://doi.org/10.1371/journal.pbio.0030128.

[125] Thibault Lagache, Nathalie Sauvonnet, Lydia Danglot, and Jean-Christophe Olivo-Marin. Statistical analysis of molecule colocalization in bioimaging. *Wiley-Liss Inc.*, 87(6):568–579, 2015. doi: 10.1002/cyto.a.22629. URL https://doi.org/10.1002/cyto.a.22629.

[126] A. Francesca Setiadi, Nelson C. Ray, Holbrook E Kohrt, Adam Kapelner, Valeria Carcamo-Cavazos, Edina B. Levic, Sina Yadegarynia, Chris M. van der Loos, Erich J. Schwartz, Susan Holmes, and Peter P. Lee. Quantitative, architectural analysis of immune cell subsets in tumor-draining lymph nodes from breast cancer patients and healthy lymph nodes. *Public Library of Science*, 5(8):e12420–NA, 2010. doi: 10.1371/journal.pone.0012420. URL https://doi.org/10.1371/journal.pone.0012420.

[127] Adrian Baddeley and Rolf Turner. spatstat: An r package for analyzing spatial point patterns. *Foundation for Open Access Statistic*, 12(6):1–42, 2005.

[128] Adrian Baddeley, Peter J. Diggle, Andrew Hardegen, Thomas Lawrence, Robin K. Milne, and Gopalan Nair. On tests of spatial pattern based on simulation envelopes.

Wiley, 84(3):477–489, 2014. doi: 10.1890/13-2042.1. URL https://doi.org/10.1890/13-2042.1.

[129] Dmitry I. Gabrilovich and Srinivas Nagaraj. Myeloid-derived suppressor cells as regulators of the immune system. *Nature Publishing Group*, 9(3):162–174, 2009. doi: 10.1038/nri2506. URL https://doi.org/10.1038/nri2506.

[130] Damien Arnol, Denis Schapiro, Bernd Bodenmiller, Julio Saez-Rodriguez, and Oliver Stegle. Modeling cell-cell interactions from spatial molecular data with spatial variance component analysis. *Cell Press*, 29(1):202–211.e6, 2019. doi: 10.1016/j.celrep.2019.08.077. URL https://doi.org/10.1016/j.celrep.2019.08.077.

[131] Calvin Park, Gerald Ponath, Maya Levine-Ritterman, Edward Bull, Eric C Swanson, Philip L. De Jager, Benjamin M. Segal, and David Pitt. The landscape of myeloid and astrocyte phenotypes in acute multiple sclerosis lesions. *BioMed Central*, 7(1):130–130, 2019. doi: 10.1186/s40478-019-0779-2. URL https://doi.org/10.1186/s40478-019-0779-2.

[132] Nico Battich, Thomas Stoeger, and Lucas Pelkmans. Control of transcript variability in single mammalian cells. *Cell Press*, 163(7):1596–1610, 2015. doi: 10.1016/j.cell.2015.11.018. URL https://doi.org/10.1016/j.cell.2015.11.018.

[133] Hartland W Jackson, Jana R Fischer, Vito RT Zanotelli, H Raza Ali, Robert Mechera, Savas D Soysal, Holger Moch, Simone Muenst, Zsuzsanna Varga, Walter P Weber, et al. The single-cell pathology landscape of breast cancer. *Nature*, 578(7796):615–620, 2020.

[134] Ellis Patrick, Nicolas P Canete, Sourish S Iyengar, Andrew N Harman, Greg T Sutherland, and Pengyi Yang. Spatial analysis for highly multiplexed imaging data to identify tissue microenvironments. *bioRxiv*, 2021.

[135] Andreas Heindl, Sidra Nawaz, and Yinyin Yuan. Mapping spatial heterogeneity in the tumor microenvironment: a new era for digital pathology. *Laboratory investigation*, 95(4):377, 2015.

Chapter 2

AFid: a tool for automated identification and exclusion of autofluorescent objects from microscopy images

Publications incorporated into this chapter:

Baharlou, H.*, **Canete, N.P.***, Bertram K.M., Sandgren, K.J., Cunningham, A.L., Harman, A.N, Patrick, E. (2020) AFid: a tool for automated identification and exclusion of autofluorescent objects from microscopy images *Bioinformatics, Volume 37, Issue 4, 15 February 2021, Pages 559–567*

*Co-first authors

Preface

The first challenge that was present in the analysis of the HIV CycIF data was the presence of autofluorescence in the images, with these signals being difficult to distinguish from real immunofluorescent staining. This prevented the accurate quantification of HIV and its target cells. This chapter consists of a publication in Bioinformatics, which describes 'Autofluorescence Identifier' (AFid), a post-acquisition algorithm for identifying and removing autofluorescence from images. This publication had arisen from work first conceived by my co-first author H. Baharlou, with a prototype implementation being featured in my Honours thesis (Canete Honours Thesis 2018). The work featured here builds significantly on the initial implementation, containing novel improvements to the algorithm, including the improved k-means clustering and custom dilation function. As co-first author, I had several intellectual contributions to the project, specifically with the development of the algorithm. Additionally, I wrote the codebase in ImageJ, MATLAB and R, as well as the code used for assessing algorithm performance.

Abstract

Motivation Autofluorescence is a long-standing problem that has hindered the analysis of images of tissues acquired by fluorescence microscopy. Current approaches to mitigate autofluorescence in tissue are lab-based and involve either chemical treatment of sections or specialized instrumentation and software to 'unmix' autofluorescent signals. Importantly, these approaches are pre-emptive and there are currently no methods to deal with autofluorescence in acquired fluorescence microscopy images.

Results To address this, we developed Autofluorescence Identifier (AFid). AFid identifies autofluorescent pixels as discrete objects in multi-channel images post-acquisition. These objects can then be tagged for exclusion from downstream analysis. We validated AFid using images of FFPE human colorectal tissue stained for common immune markers. Further, we demonstrate its utility for image analysis where its implementation allows the accurate measurement of HIV–Dendritic cell interactions in a colorectal explant model of HIV transmission. Therefore, AFid represents a major leap forward in the extraction of useful data from images plagued by autofluorescence by offering an approach that is easily incorporated into existing workflows and that can be used with various samples, staining panels and image acquisition methods. We have implemented AFid in ImageJ, Matlab and R to accommodate the diverse image analysis community.

Availability and implementation AFid software is available at https://ellispatrick.github.io/AFid

2.1 Introduction

Immunofluorescence microscopy (IF) is a powerful tool for simultaneously visualising the localization of multiple proteins *in situ*. Additionally, several methods have been developed that push the number of parameters visualized in a single section to well beyond traditional 3–4 colour IF [1–7]. This allows for the definition of multiple cell types, complex subsets and also cellular states *in situ*. Despite these advances the utility of IF, particularly for quantitative measurements, has been hampered by the longstanding issue of autofluorescence.

Autofluorescence is present in all tissues and has many sources including components of structural and connective tissues, cellular cytoplasmic contents and also fixatives used to preserve tissue [8–11]. Autofluorescent substances have their own excitation and emission profiles that can span the entire visible and even infra-red spectrum and therefore significantly overlap with standard microscope excitation/emission filter setups [10] (Supplementary Figure 2.1). This presents a major obstacle to image analysis, particularly any kind of automated analyses, as 'real' versus 'autofluorescent' regions of interest (ROIs) cannot be readily distinguished. An example of this is shown (Supplementary Figure 2.2) where the accurate quantification of CD3 labelling in human colon tissue is severely hampered by autofluorescent signals.

Several methods have been developed to address the issue of autofluorescence. The oldest and most widely used are chemical methods to quench autofluorescence. These include exposing samples to either UV radiation or a chemical solution prior to or during staining [8, 12–14]. Although these methods can be effective, there are several disadvantages including quenching of desired signal from endogenous reporters or fluorescent probes, and also that there is no general recipe with specific protocols required to quench specific types of autofluorescence [8, 13]. However, the primary limitation is that the quenching must take place prior to imaging, so if autofluorescence is detected after image acquisition it is too late to remove it. This can be frustrating as autofluorescence is highly variable between tissue sections.

Digital methods of autofluorescence mitigation have also been developed such as spectral unmixing and algorithmic subtraction of a background reference image acquired either prior to staining or during image acquisition [1, 15–17]. These methods are robust and have the capacity to resolve signal versus autofluorescence. As such they represent an important pre-processing step to ensure accurate image analysis. However, there are several limitations to these approaches. Spectral unmixing requires the use of specialized instrumentation and proprietary software which is a limiting factor in its widespread use [15]. Additionally, users must acquire an entire emission spectrum for each laser line and for every single field of the image which significantly increases image acquisition time and the size of the acquired data. Similarly, background subtraction using a customized filter setup requires a microscope with tuneable filters and expertise beyond that of most researchers [16]. The alternative background subtraction method requires that entire tissue sections are imaged at a pre-defined resolution prior to staining, whereby the user must perform intensity scaling and pixel-perfect registration, again representing a major hurdle for most researchers [1]. Overall, the advantage of the aforementioned methods is that theoretically they can be applied to any type of sample. However, the trade-off is that they are impractical for many users due to either equipment/software availability, required expertise or due to the not insignificant amount of extra work involved.

Taken together, there is an unmet need for the development of alternative methods for

excluding autofluorescence from microscopy images, that are user-friendly, open source and practical. Based on the long excitation/emission wavelengths of autofluorescence [10] (Supplementary Figure 2.1) and the observation that in many cases the majority of interfering autofluorescence is spatially distinct from signal of interest (examples shown in figures throughout this paper), we reasoned that we could develop a post-acquisition approach to identify and exclude autofluorescence thereby improving image analysis accuracy. To this end, we developed 'Autofluorescence Identifier' (AFid), an algorithm which is able to detect autofluorescent pixels as discrete objects within multi-channel IF images of tissue. AFid requires only the information from two fluorescent channels, where bright fluorescent ROIs are located and classified as 'real' or 'autofluorescent' based on measures of pixel correlation and moments of the distribution of pixel values. Identified autofluorescent objects can then be tagged for exclusion from downstream analysis. A key advantage of this method is that it is applied to images post-acquisition, which makes it highly practical approach as users do not need to alter their experimental procedures. However, an important caveat is that AFid is only suitable for use when its assumptions are met, i.e. that autofluorescence is mostly non-overlapping with real signal and appears across at least two fluorescent channels. In this paper we describe the steps of the AFid algorithm, validate its usage on FFPE human colorectal tissue and also demonstrate its utility for image analysis where its implementation allows the accurate measurement of HIV-Dendritic Cell (DC) interactions in a colorectal explant model of HIV transmission.

2.2 Materials and methods

2.2.1 Ethics for use of human tissue samples

This study was approved by the Western Sydney Local Area Health District (WSLHD) Human Research Ethics Committee (HREC); reference number (4192) AU RED HREC/15 WMEAD/11. Human colorectal and skin tissues used for this study were approved by this committee and all patients were consented prior to sample collection. Brain and Heart tissue images were donated data.

2.2.2 Immunofluorescence staining

Tissues were fixed in 4% paraformaldehyde (Electron Microscopy Sciences) for 18–24h at room temperature then immersed in 70% ethanol prior to paraffin embedding. 4 µm paraffin sections were adhered to glass slides (SuperFrost Plus, Menzel Glazer), baked at 60°C for 40min, dewaxed in xylene followed by 100% ethanol then air dried. All wash steps described herein were carried out by immersing slides in three successive Coplan Jars of Tris-buffered saline (Amresco, Cat: 0788) on a rotator for a total of 10min. Antigen retrieval was then performed using a pH9 antigen retrieval buffer (DAKO) in a decloaking chamber (Biocare) for 20min at 95°C. Slides were then washed in TBS. To acquire unlabelled background images (hyperref[fig:s2.2]Supplementary Figures 2.2 and 2.9), sections were stained with 1ug/ml DAPI (Roche) for 3min, mounted under coverslips with SlowFade-Diamond Antifade (Molecular Probes) and the whole section imaged on an Olympus VS120 microscope (see Image acquisition below). Coverslips were then floated away in TBS and sections on slides were blocked for 30min (0.1% saponin, 1% BSA, 10% donkey serum, diluted in TBS) at room temperature. Sections were then washed in TBS and incubated with primary antibodies overnight at 4°C. Antibodies for

primary detection include: Abcam—rabbit CD11c (EP1347Y), mouse CD3 (F7.2.38), rabbit CD8 (polyclonal, ab4055); DAKO—rabbit CD3 (polyclonal, A045229-2); Affinity Biologicals—sheep FXIIIA (polyclonal). Sections were then washed in TBS and incubated with secondary antibodies for 30min at room temperature. Donkey secondary antibodies (Molecular Probes) against rabbit, mouse or sheep were used and were conjugated to either Alexa Fluor 488 or 546. Sections were stained with DAPI (if not already performed in a previous step) and mounted with SlowFade-Diamond Antifade.

2.2.3 HIV explant infection

Healthy Inner foreskin explants were infected with either HIVBal or Transmitted/Founder HIV-1 Z3678M using an explant setup as previously described [18]. A TCID50 of 3500 (titrated on TZMBLs as previously described [19]) was used to infect all explants. Tissues were then fixed and paraffin embedded as described above.

2.2.4 RNAScope

Detection of HIV RNA was performed using the 'RNAscope 2.5HD Reagent Kit-RED' and following the manufacturer's protocol (Cat: 322360, ACD Bio) with custom probes (consisting of 85 zz pairs) against HIV-$_{\text{BaL}}$ (REF: 486631, ACD Bio) spanning base pairs 1144-8431 of HIV-1$_{\text{BaL}}$ sequence. Following the RNAscope protocol, sections were stained from the blocking step as detailed above.

2.2.5 Microscopy

Imaging was performed using an Olympus VS120 Slide Scanner with ORCA-FLASH 4.0 VS: Scientific CMOS camera. VS-ASW 2.9 Olympus software was used for acquisition of images and conversion of raw vsi files to tiff format for downstream processing. Objectives used are indicated in figure legends and include: 10× (UPLSAPO 10X/NA 0.4, WD 3.1/CG Thickness 0.17), 20× (UPLSAPO 20X/NA 0.75, WD 0.6/CG Thickness 0.17) and 40× (UPLSAPO 40X/NA 0.95, WD 0.18/CG Thickness 0.11–0.23). Channels used include: DAPI (Ex 387/11–25nm; Em: 440/40–25nm), FITC (Ex: 485/20–25nm; Em: 525/30–25nm), TRITC (Ex: 560/25–25nm; Em: 607/36–25nm) and Cy5 (Ex: 650/13–25nm; Em: 700/75–75nm). For 40× images, Z-stacks were acquired 3.5 µm above and below the plane of focus with 0.5 µm step sizes. Huygens Professional 18.10 (Scientific Volume Imaging, The Netherlands, `http://svi.nl`) CMLE algorithm, with SNR: 20 and 40 iterations, was used for deconvolution of Z-stacks. For images where the unstained background was acquired prior to staining, images were aligned using the ImageJ plugin multiStackReg vs1.45 with the DAPI channel serving as a reference for alignment.

2.2.6 Acquisition of autofluorescence spectra

Autofluorescence spectra of unstained tissue samples (Supplementary Figure 2.1) were acquired using an Olympus FV1000 laser scanning confocal microscope with a 20× objective. The excitation lasers lines 405, 473 and 559nm were used and emission spectra were acquired using a 20nm wide bandpass filter, shifted in 20nm intervals from 415–795, 490–790 and 575–795 nm respectively.

2.2.7 Generation of intersection mask

A mask of the intersection of the two channels was used for autofluorescence exclusion. This is termed the 'intersection mask'. The intersection mask contains only signals present in both channels and therefore contains the autofluorescent ROIs among other objects such as co-stained markers and dim background stromal fluorescence. The intersection mask was generated by the following procedure. Each channel was Gaussian blurred with a sigma of two. A Niblack threshold was then applied to each channel (threshold radius 30 pixels) to generate binary masks. The intersection ('AND' operation) of these masks was then taken and used for autofluorescence classification by clustering as detailed below.

2.2.8 Clustering for autofluorescence identification

Within the objects defined by the intersection mask we measured multiple features in each of the two channels on non-Gaussian blurred images. These features included standard deviation, kurtosis, as well as the inter-channel Pearson's correlation coefficient of corresponding pixels. These features were transformed by taking the natural log (standard deviation, skewness and kurtosis) or the inverse tanh transformation (correlation). All features were standardized by dividing by the standard deviation of the transformed feature values. k-means clustering was then performed on these features to identify a cluster of ROIs which are likely to be autofluorescent. The cluster with the highest average correlation value was defined as the cluster containing autofluorescent ROIs. A well-chosen number of clusters (k) is important for detecting a homogeneous cluster of autofluorescent ROIs. As such we developed an automated approach for optimal choice of k (high sensitivity and specificity). The procedure is as follows. (i) k-means is performed iteratively with 3–20 clusters. (ii) A two-tailed t-test is performed on the arctanh transformed correlation values of the two clusters with highest average correlation values. (iii) The test statistic values are then plotted against k, which produces an asymptotically decreasing function (Supplementary Figure 2.14). (iv) We developed an 'elbow method' approach to finding the optimal cluster number. A straight line is drawn connecting the statistic value for the lowest k, to that of the highest k. The perpendicular distance of each plotted point to the line is measured and the optimal k is estimated to correspond to the point with the greatest distance below the line. This method is illustrated in (Supplementary Figure 2.14). The intersection mask is then modified, keeping only the objects identified as autofluorescence.

2.2.9 Custom dilatation function to outline autofluorescent ROIs

After clustering and creating a mask of autofluorescent objects we then employed a custom dilation function to outline the full body of autofluorescent objects for exclusion. The essence of the algorithm is to evenly distribute points within an amorphous object and then to expand out from these points in all directions until a halting condition is met.

To distribute points the following approach was developed: (i) ROIs in the autofluorescence mask were skeletonized, reducing objects to a line of 1 pixel-width that follows the morphological gradient of the original object. (ii) End-node pixels for each object in the image were first identified, defined as having only one neighbour. If there were no end-nodes for an object, as in the case of an annulus, the top-left-most pixel was defined as the end-node. (iii) A skeleton tracing algorithm was employed that starts from the end nodes and moves throughout the skeleton, distributing centres for expansion every

20 pixels (illustrated in Supplementary Figure 2.4d). Tracing of pixels to neighbours occurred as long as the neighbouring pixel was in the skeleton and had not yet been traced by another point. Once these conditions were no longer met, tracing for a given object was halted. Expansion from distributed centres is illustrated using a schematic in Supplementary Figure 2.5. In brief: (i) Lines of length 60 pixels emanating from centres were drawn in all directions separated by an angle of theta where theta was defined by the law of cosines. For a 60 pixel line the equation for theta becomes $\cos\left(1-\left(\frac{1}{60}\right)^2/2\right)$. (ii) Pixel values of the Gaussian blurred image for each channel were measured beginning from the point of intersection of the line and perimeter of the object in the intersection mask, to the end of the line. (iii) The co-ordinates of the first point where pixel values increased were recorded for each line. (iv) A new outline of the object was created by combining these co-ordinates (Supplementary Figure 2.4e). (v) Pixel values of the new outline of the object were set to 0.

2.2.10 Assessment of custom dilation function robustness with varying parameters

This section specifically details the methods for generating data for Supplementary Figure 2.4g, which assesses both the utility of the custom dilation function for capturing autofluorescent pixels and its robustness against varying parameters. For this figure whole-slide scanned images of colorectal tissue before antibody labelling and after CD11c/CD3 antibody labelling were used (data from Figure 2.2a-c). AFid was run with or without use of the custom dilation function and using the CD11c and CD3 channels as inputs. This was performed for varying threshold radii and for varying cluster number k as shown in the graph. To assess coverage of autofluorescent pixels (true positive) and real signal pixels (false positive) upon parameter variation as shown in the graph, we generated ground-truth masks. To do this, we specifically used the CD3 channel due to the consistently high density of T cells across images, thus providing enough data to assess the false positive rate. We generated a mask of the unlabelled background image (autofluorescence only) and the CD3 channel after staining (CD3+ T cells and autofluorescence). Our aim was to use the background image to parse out the autofluorescent pixels contained within the CD3 channel image.

First, we performed a morphological watershed on the CD3 channel mask to separate touching T cells and autofluorescence. We then performed a binary reconstruct of the watershed CD3 mask with the background image mask. This generated an estimate mask of the autofluorescent pixels within the CD3 image which we used as a ground-truth for autofluorescence. We then generated the real-signal ground-truth mask by subtracting the autofluorescence ground-truth mask from the CD3 channel mask. The ground-truth masks were then manually inspected and objects were removed which did not clearly correspond to T cells in the real-signal mask or autofluorescence in the autofluorescence mask.

2.2.11 Algorithm performance assessment

The performance of our algorithm was tested using three different staining panels on human colonic tissue as shown in Figure 2.2. To benchmark performance assessment, we manually annotated regions of the intersection mask (see Generation of intersection mask) as belonging to 'real' or 'autofluorescent' signals. Delineation of the two types of signal was

achieved using the 'unstained background image' as a reference (see Immunofluorescence staining). In total 400 ROIs, 200 for each category, were annotated. The actual annotation was performed using the Cell Counter Plugin in ImageJ. Results were exported as a csv file, where each row indicated an individual ROI, its category and x, y co-ordinates.

The two fluorescent channels, intersection mask and spreadsheet of annotated ROI co-ordinates were fed in to R. k-means clustering with estimated k was then performed as described above. The true positive rate and false positive rate were thus determined as the proportion of ROIs in each category that resided in the 'autofluorescence cluster', which was the cluster with highest average correlation values (Supplementary Figures 2.5 and 2.14).

2.2.12 HIV spot segmentation

Spot counting was performed using a custom MATLAB script implementing the spot counting technique presented by Battich et al. (2013) [20]. First, a manual threshold was performed on the HIV RNA channel to identify areas that have HIV stain present. The *IdentifySpots2D* function by Battich et al. was then used to identify the number of spots, with the detection threshold set to a generous value of 0.01 and the number of deblending steps equal to 2. Finally, any spots identified were excluded if they were not present in the threshold mask obtained previously.

2.2.13 Singe cell segmentation

To perform single cell segmentation, a custom MATLAB script was used. Briefly, a Gaussian filter with a full-width at half maximum of 10 pixels was applied to the DAPI image to ensure that each nucleus has only one locally maximum pixel intensity. Further, the *imordfilt2* function is used to ensure that maxima are not less than 7 pixels apart. Watershed segmentation is performed using the watershed function to identify nuclear boundaries. Objects with diameters less than 10 pixels or greater than 50 pixels were discarded, and the nuclear objects are dilated by 6 pixels to estimate the cell body. The *regionprops* function was finally used to measure the mean pixel intensities of other image channels within each identified cell boundary, as well as the number of HIV RNA spots identified within each cell. The data was exported as a .csv and was analysed using FlowJo.

2.3 Results

2.3.1 Algorithm overview

The steps for the AFid algorithm are summarized in (Figure 2.1). Also the pseudocode for all steps of the algorithm is provided in Supplementary Material under Supplementary Algorithms 1-8. First, thresholds are applied to the two fluorescent channels and an 'intersection mask' is created to detect the ROIs that are fluorescing in both channels (Figure 2.1a, left). Second, we then measure ROIs in the 'intersection mask' for multiple textural features (Figure 2.1b, middle). To select these features, we make a key assumption that the fundamental topology of pixel intensities for an autofluorescent ROI is conserved across channels. This makes sense as sources of autofluorescence tend to have long excitation and emission profiles (Supplementary Figure 2.1). Furthermore, this

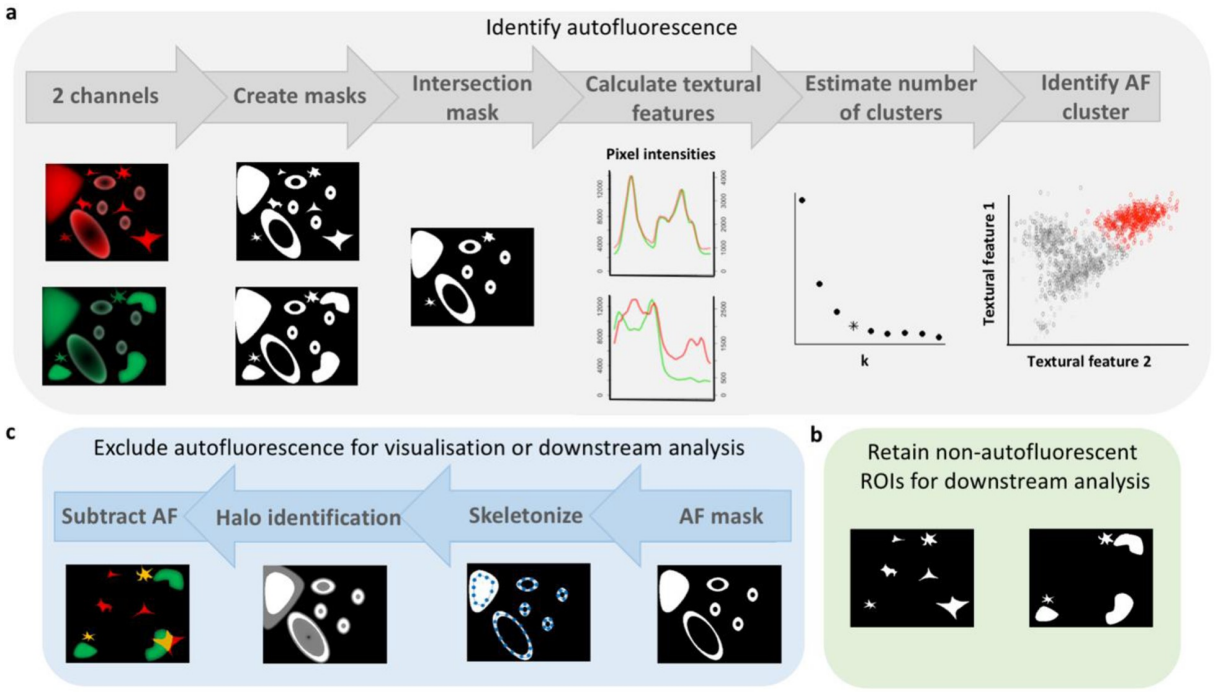


Figure 2.1: **Steps of the AFid algorithm. (a)** k-means clustering on a set of textual features of objects in an intersection mask of two channels. Autofluorescent ROIs can then be tagged for exclusion in downstream analysis **(b)**, or a custom dilation function can be employed to estimate the perimeter of autofluorescent ROIs, which are then excluded from the image (pixel values set to 0) **(c)**

assumption can be readily verified by examining the frequency distribution of pixel values of autofluorescent ROIs, which shows conservation between channels compared to that of cells labelled with fluorescently tagged antibodies (Supplementary Figure 2.3a). As such, any measure of pixel behaviour within an ROI will be linearly correlated across channels (Supplementary Figure 2.3b-d). Therefore, to identify autofluorescence we measure multiple features including pixel correlation (Fisher transformed Pearson's correlation coefficient) and moments of the distribution of pixel values, including Standard Deviation and Kurtosis. Third, ROIs can be clustered using the textural features as inputs to identify a distinct cluster with high correlation values that consists mainly of autofluorescent ROIs (Figure 2.1a, right). Here we have used a k-means clustering algorithm with automated choice of k. Finally, these autofluorescent ROIs can be excluded from downstream analysis or can be subtracted from the raw images for visualization (Figure 2.1b-c).

2.3.2 Custom dilation function to outline autofluorescent objects

For optimal visualization, and to aid in downstream analysis, we have also developed a novel algorithm which expands from autofluorescent objects to capture their full perimeter. Due to variations in intensity scale within an image and across different images, conventional thresholding algorithms rarely capture the entire perimeter of autofluorescent ROIs (Supplementary Figure 2.4a-c). This can represent a limitation for automated autofluorescence exclusion, as several threshold parameters need to be tested beforehand for each image, and then assessed by eye to determine appropriateness. To overcome this

limitation we developed a custom dilation function that works in tandem with thresholding to automatically outline the full body of autofluorescent ROIs, regardless of shape and intensity (Figure 2.1c). In brief this works by skeletonising ROIs and evenly distributing points throughout the skeleton (Supplementary Figure 2.4d). We then expand from these points until the gradient of pixel intensities from the ROI boundary outwards begins to increase, indicating the end of the object or the beginning of a neighbouring object (Supplementary Figure 2.4e). The specific details of how expansion occurs is described in the methods and also using the schematic in Supplementary Figure 2.5. Visually we can see that this method accurately captures the full perimeter of autofluorescent ROIs with minimal effect to neighbouring signals (Supplementary Figure 2.4f). Further, we quantified the percentage of autofluorescence captured by AFid with and without custom dilation for varying threshold radii and cluster numbers k (Supplementary Figure 2.4g). The results show that the custom dilation function identifies a higher percentage of autofluorescent pixels which is stable against variation of threshold radii and cluster number k, with minimal impact on neighbouring real signals (with the exception of small k values which are typically not selected by AFid). Accordingly, employment of the dilation function allows for efficient autofluorescence detection despite variation in choice of threshold radii. The final result is an image retaining only non-autofluorescent ROIs, which can then be used for visualization and downstream analysis (Figure 2.1c).

2.3.3 Validation of AFid

In order to establish both the efficacy and scope of utility for AFid we tested the algorithm with multiple types of input images. First, we tested whether the markers for detection in each channel could influence autofluorescence identification. To this end we defined three use-cases where input channels contained (i) non-co-expressed markers (CD11c and CD3), (ii) a marker expressed on autofluorescent cells (FXIIIA+ Macrophages) and (iii) co-expressed markers (CD3 and CD4). These three use-cases are shown for human colorectal tissue where sections were imaged before (Figure 2.2a, d, g) and after staining (Figure 2.2b, e, h). The unstained image was used for manual annotation of autofluorescent ROIs providing a ground truth for our classifier, and not used by AFid. We found that in all three use-cases the autofluorescence cluster was highly enriched for autofluorescent ROIs (mean=98.4%, SD=1.6%, n=9) (Supplementary Figure 2.6a). We also achieved good coverage identifying on average 96.0% (SD=4.0%, n=9) of all annotated autofluorescence, with a low false positive rate of 1.7% (SD=1.7%, n=9) (Supplementary Figure 2.6b). Pairwise plots for each use-case are shown (Supplementary Figures 7-9), demonstrating good separation of autofluorescence (yellow) from non-autofluorescence (grey) by k-means. The final result of AFid, after applying the custom dilation function is shown (Figure 2.2c, f, i), demonstrating near complete exclusion of autofluorescence in all three use-cases.

We next sought to test AFid on different tissue types and image acquisition conditions. We found that AFid successfully identified autofluorescence across various tissue-types, including heart, skin and brain tissues where autofluorescence is a well-known problem (Supplementary Figure 2.10). Additionally, AFid could specifically identify autofluorescent objects despite varying image quality, including variations in resolution, (Supplementary Figure 2.11) and on images both pre- and post-deconvolution (Supplementary Figure 2.12). Further, our algorithm is compatible with very large images that are inundated with autofluorescence, which allows large datasets with significant noise to be rescued for analysis (Supplementary Figure 2.13).

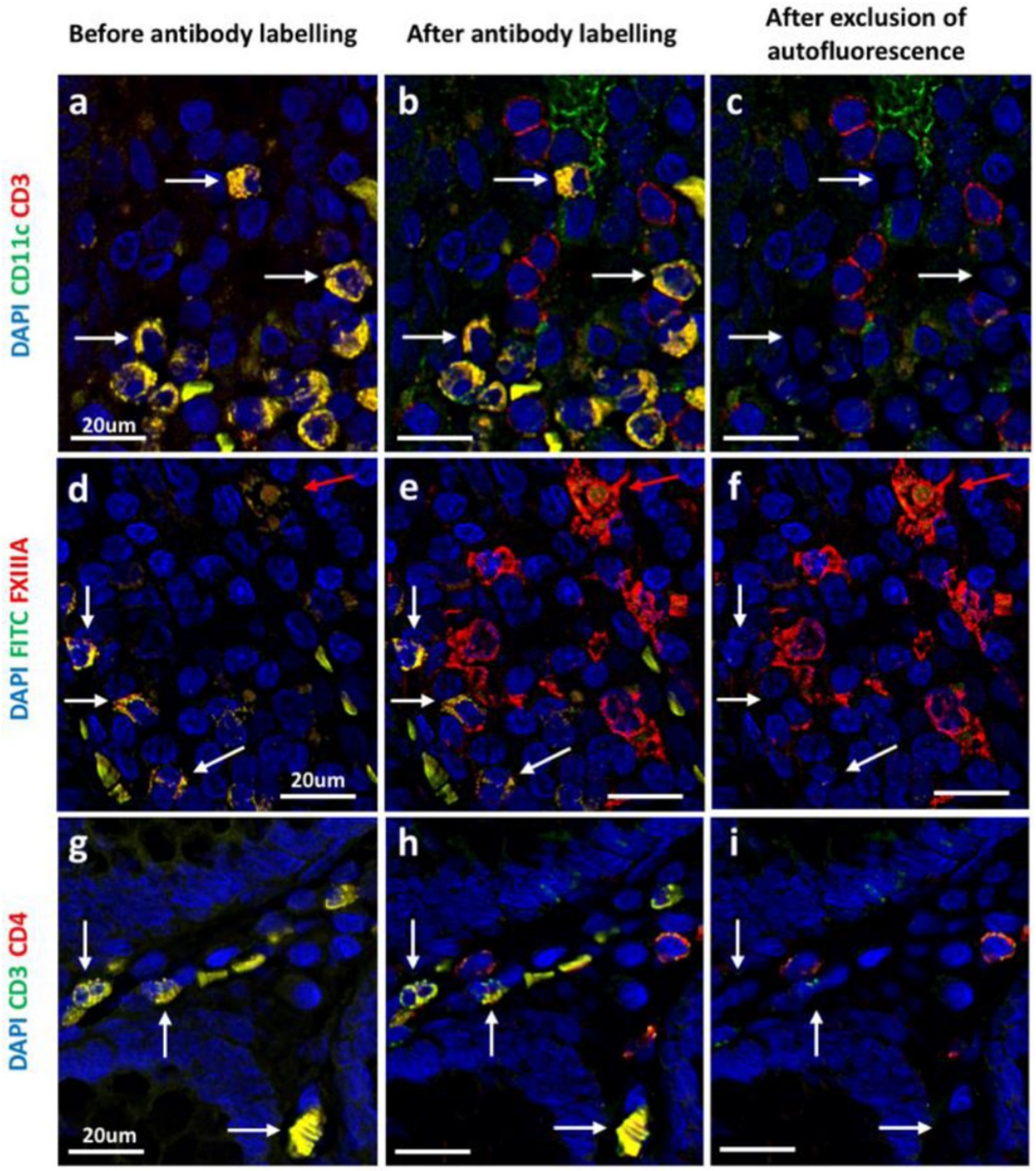


Figure 2.2: **Application of AFid to various staining panels.** Sections of fixed human colorectal tissue prior to **(a, d, g)** and after labelling **(b, e, h)** with antibodies targeting the indicated markers. **(c, f, i)** Labelled images after identification and exclusion of autofluorescencent objects using AFid. White arrows indicate some autofluorescent objects that have been excluded by the algorithm. The red arrow in the middle row indicates an autofluorescent macrophage which was not identified by AFid. In the middle row, 'FITC' is the FITC channel, which was imaged but not used to detect any markers. Images are representative of 3 donors for each staining panel

It is useful to outline instances where AFid is unable to identify and exclude autofluorescence. These instances typically correspond to when the two major assumptions of the AFid algorithm are not met. These assumptions are (i) that autofluorescence must appear, even if only faintly, across at least two acquired channels and (ii) that autofluorescence must be spatially separated from real signal. The white arrows in Supplementary Figure 2.14b, b indicate autofluorescence that was not identified by AFid as their spectra was restricted to the green channel and did not appear in the red channel. In particular these objects corresponded to a subset of autofluorescent cells that appear in the colonic epithelium of some donors and also a small percentage of muscle fibres in muscularis propria of the colon. Examples of when assumption 2 are not met are shown by the yellow arrows in Supplementary Figure 2.14b. Here the FXIIIA signal used to identify a macrophage is overlapping with autofluorescent muscle fibres in the muscularis propria and so AFid was unable to identify and remove these autofluorescent objects. For our samples, the frequency of such failed identifications was low and so had a negligible impact on our desired measurements. However, this may not be the case depending on the users samples and markers used for labelling. As such, these examples serve to illustrate the importance of users checking that their samples meet the assumptions of the AFid algorithm so as to ensure sufficient autofluorescence exclusion for accurate measurements.

2.3.4 AFid allows accurate assessment of HIV-dendritic cell interactions in an explant infection model

Finally, we show that the presence of autofluorescence and its exclusion can have a major impact on down-stream analysis. In our own studies, we are interested in the early HIV-target cell interactions that occur in human colorectal explants that we topically infect with HIV for up to 2h. However, these explants are prone to significant amounts of autofluorescence which impedes accurate analysis.

To assess the frequency and number of HIV+ DCs in an image we first ran AFid to generate a mask of autofluorescence within the image. We then used a spot counting algorithm [20] to segment individual HIV particles, and subsequently segmented individual cells using their nuclei to generate single cell data. The fluorescent channels and masks of both the autofluorescence and HIV particles were then converted to FCS format for single cell analysis in FlowJo. Gating on putative CD11c+ DCs (Figure 2.3a, top left) we then excluded autofluorescent cells, measured as the percentage of the cell body overlapping with the autofluorescence mask (Figure 2.3a, top right). After excluding cells containing autofluorescence we found that just 2% of DCs were HIV+ (contained at least 1 HIV particle) (Figure 2.3a, bottom right) versus 16% without excluding autofluorescent cells (Figure 2.3a, bottom left). This corresponded to a large difference in the total number of HIV+ DCs identified, 10 versus 168 cells (Figure 2.3c). We visually inspected the co-ordinates of putative HIV+ DCs after autofluorescence exclusion and were able to confirm that these cells corresponded to legitimate HIV-DC interactions as shown in the zoomed in images in Figure 2.3f. In contrast, false positive signals were almost all due to autofluorescent signals spanning multiple channels. Importantly, we found that it was not possible to simply circumvent autofluorescence by gating specifically on true HIV+ DCs. This was because they did not occupy a unique position on biaxial plots and therefore could not be gated on without also including autofluorescent cells (Figure 2.3b). Therefore, highlighting the necessity of a computational approach to tag false positive signals within a mixed population.

Finally, we show that the presence of autofluorescence can mask spatial relationships which are then revealed by its exclusion. DCs are highly migratory and thought to be an important early target cell for HIV transmission [21]. To assess whether DCs may be specifically localising to HIV early post-exposure we divided an image into 100×100 µm^2 quadrats and classified each as HIV- or HIV+ and then measured the density of CD11c labelling in each area to quantify expression. Prior to autofluorescence exclusion (setting pixel values to 0), the apparent CD11c expression did not significantly differ between HIV- and HIV+ areas (Figure 2.3d), whereas after exclusion CD11c expression was revealed as significantly higher in HIV+ areas compared to HIV- areas (Figure 2.3e). This was due to a large amount of measured CD11c expression being derived from autofluorescence (Figure 2.3f). Further, we found that CD11c, HIV and autofluorescent cells were differentially located. CD11c and HIV clustered toward the tip of the lamina propria where the majority of interactions took place (Figure 2.3f, zoomed images), whilst autofluorescent cells were particularly clustered toward the base of the lamina propria, thus skewing the results. Together these data demonstrate how the presence of autofluorescence can drastically skew measurements, and that its post-acquisition exclusion, here using AFid, can enable accurate image analysis.

2.4 Discussion

Here, we have presented AFid, a first of its kind method for identifying autofluorescence in multi-channel fluorescence microscopy images post-acquisition. We have shown that AFid is able to identify autofluorescence across tissue types, staining panels and image acquisition conditions. Importantly, we showed that the presence of autofluorescence can have a major impact on downstream analyses which can be mitigated through the use of AFid.

Our rationale for creating AFid lay in its necessity for accurately measuring HIV-cell interactions in our images of HIV-infected FFPE human colorectal tissues. In particular, our staining protocols were incompatible with chemical bleaching methods discussed in the introduction of this paper [8, 12–14], and we did not have the necessary resources to perform spectral unmixing [15] or background subtraction [1, 16, 17]. Noting that the majority of interfering autofluorescence in our samples were bright, spatially distinct objects appearing across multiple channels we reasoned that they could be identified and excluded from downstream analyses by way of a computational approach. Indeed, as demonstrated here the application of AFid was sufficient to enable accurate quantitative measurements of HIV-cell localization.

AFid has several distinct advantages. First, it is a post-acquisition algorithm. This is important because the user does not need to pre-emptively deal with autofluorescence. Additionally, it can be applied to existing microscopy datasets potentially mitigating the need for optimization and repeat experimentation. Indeed, this is exemplified in our own case, where bulk image datasets from HIV-infected explants were probed using RNAscope and multiple immune markers (partially presented in Figure 2.3, unpublished data) but quantification hindered by autofluorescence. In this case, even if further lab-based optimizations proved successful at mitigating the impact of autofluorescence for our specific tissues, repeating the experiments would require the use of precious samples and a significant investment of time and resources. The second key advantage of our algorithm is that its design allows for the detection of autofluorescence across varying tissues and image acquisition conditions, as demonstrated here. In particular, we assume that due to the

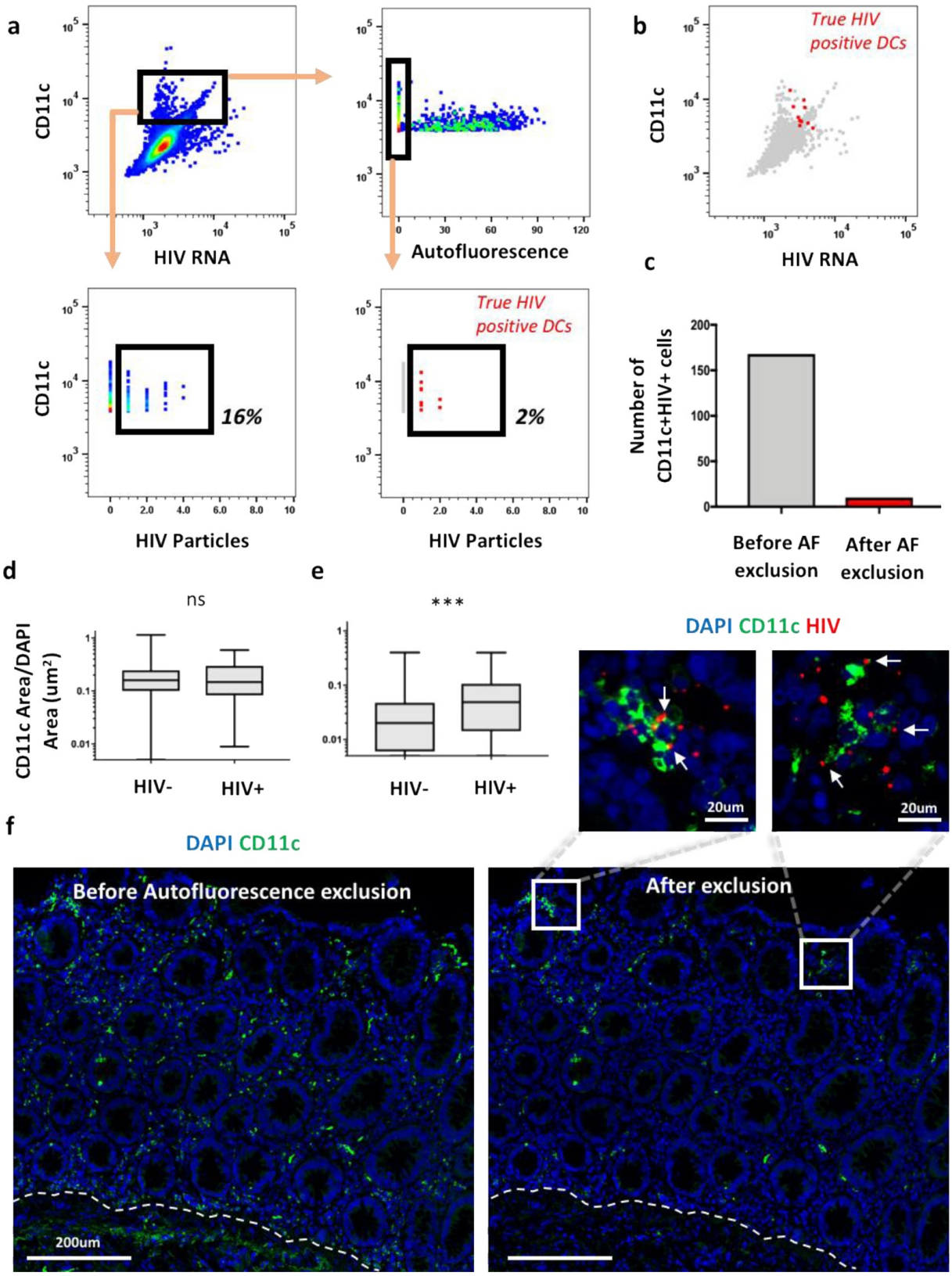


Figure 2.3: **AFid facilitates analysis of early HIV-target cell interaction.** (continued on next page)

Figure 2.3: Human colorectal explants were topically infected with HIVBal for 30 min, fixed, sectioned and then stained for CD11c, HIV RNA and DAPI. **(a)** Single cell segmentation and HIV spot segmentation were performed followed by FCS file generation for analysis using FlowJo. The figure shows a gating strategy for identifying CD11c+ cells that contain at least one HIV particle, either with, or without first gating-out cells containing autofluorescence. The autofluorescence parameter shows the percentage of the cell body overlapping with the autofluorescence mask generated by AFid. **(b)** Overlay of CD11c+HIV+ cells after gating out autofluorescence, showing that they cannot be specifically gated-on without first excluding autofluorescent cells. **(c)** Total number of CD11c+HIV+ cells before and after autofluorescence exclusion, as gated on in part a. **(d, e)** A whole tissue image from one donor was divided into 100×100 μm^2 quadrats, each classified as HIV- or HIV+, and CD11c labelling was measured before **(d)** and after **(e)** autofluorescence exclusion. CD11c expression was measured per μm^2 of DAPI. Quadrats with DAPI staining less than 1/10th their area (non-tissue areas) were excluded. Boxplots show the min, first quartile, median, third quartile and max values. **(f)** A cropped area of a whole-tissue image of HIV-infected colorectum before (left) and after (right) autofluorescence exclusion from the CD11c channel. Zoomed images of the boxes show interactions of HIV with CD11c+ cells (white arrows) in the image after autofluorescence exclusion. The broken white line indicates the base of the lamina propria. A two-tailed Mann Whitney test was performed in part a and b. ns, not significant; $***P$ 0.0002

long excitation/emission wavelengths of autofluorescence [10] (Supplementary Figure 2.1), textural features of autofluorescent objects will exhibit a conserved topology across channels (Supplementary Figure 2.3a), and therefore occupy a unique position in the feature space, relative to other objects (as shown in Supplementary Figures 7-9). Accordingly, the employment of clustering, here using k means with automated choice of k, is able to consistently isolate a cluster containing mainly autofluorescent objects, despite variations across tissues and image acquisition conditions.

Despite the advantages that have been discussed there are several limitations to our approach. The major limitation is that AFid cannot identify autofluorescence that is largely overlapping with real signal. This situation may occur if the object of interest is inherently autofluorescent or localized to a highly autofluorescent area of the tissue. This maybe an advantage or a disadvantage depending on the nature of the measurements to be performed. For example, AFid would aid in object counting measurements by retaining objects with mixed real-signal and autofluorescence, whilst excluding objects that purely contain autofluorescence (e.g. autofluorescent macrophages, Figure 2.2). However, it would not be suitable if precise fluorescence intensity measurements were required of a stained object that contained some autofluorescence. Another potential limitation is that the algorithm requires that autofluorescent signals are present, even if only faintly, across at least two acquired channels. In our experience this was true of the vast majority of bright interfering autofluorescence in the various tissues used for this study and fits in with the well-known broad spectra of autofluorescence [10] (Supplementary Figure 2.1). However, we did observe instances where this assumption did not hold. In particular a subset of autofluorescent cells within colonic epithelium as well as some muscle fibres within the muscularis propria of the colon, did not fluorescence across multiple channels and were therefore not identified by AFid (Supplementary Figure 2.14). It is difficult to evaluate the prevalence autofluorescence with narrow spectral profiles, as this will depend on many factors including variations in tissue type, preservation procedure and also

donor-to-donor variation [8–11]. As such, it is important for users to be acquainted with the distribution of autofluorescence in their samples so that they can assess the efficacy of AFid for autofluorescence exclusion. For the limiting cases discussed here, if they represent a genuine hindrance to image analysis, the optimization of chemical quenching [8, 12–14] or algorithmic background subtraction methods [1, 15–17] may be necessary. One final limitation that is worth mentioning is that as AFid is an object-based method of autofluorescence identification, it cannot be meaningfully compared to previous methods which deconvolute and subtract the contribution of autofluorescence in a pixel-by-pixel manner [1, 15–17]. For the latter methods, their success is dictated by the degree of autofluorescence exclusion from each pixel of the image. However, as AFid is an object-based method, its success is determined by the percentage of distinct autofluorescent objects that are identified and excluded from the image. As such, instead of comparing AFid to previous methods we benchmarked the algorithm against its stated purpose of autofluorescent object identification (Supplementary Figure 2.6). By this measure AFid achieves a very high sensitivity and specificity for images of colorectal tissue where the majority of autofluorescent objects meet the assumptions of the algorithm. Technically, autofluorescence appears to some degree in every pixel of an image of tissue. AFid represents a practical image processing algorithm which excludes objects that are likely to interfere with image analysis, rather than deconvoluting autofluorescence from every pixel of an image.

At present, AFid is limited to autofluorescence detection using two channels. Users can still identify autofluorescence across multiple channels by iteratively running AFid with multiple pairwise channel comparisons, saving a mask of the autofluorescence for each channel after each run. In future, it may be possible to improve the algorithm by simultaneously taking in information from more than two channels, potentially making autofluorescence detection more robust. However, this would require significant alterations to the current approach and would need to be justified by user demand. As AFid is a new approach to excluding autofluorescence from images we have included in this manuscript a table of frequently asked questions which outlines the kinds of things users should watch out for when applying AFid to their images (Supplementary Table 2.1).

In summary, here we have demonstrated that AFid has major utility for quantitative measurements in human gut tissue stained for common immune markers and have shown that it is able to identify autofluorescence in various other tissue types. With the rise of quantitative image analysis, particularly single cell cytometry pipelines as shown here, there is an increasing need for image processing algorithms to filter out artefacts and enable accurate measurements. Although the use of AFid in this paper was restricted to traditional 2–3 colour IF, it can also be integrated into more recent high-throughput assays such as CycIF and CODEX [2, 4, 6]. These methods employ iterative rounds of antibody staining and fluorescence removal using a bleaching solution. The bleaching solution also mitigates autofluorescence but can have an impact antigenicity meaning that markers sensitive to bleaching are used in the first round of staining with the trade-off that autofluorescence may still be bright enough to interfere with analysis. Our group has used AFid for this very situation when employing CycIF to examine multiple cell types interacting with HIV in human colorectal tissue (unpublished data). Accordingly, AFid provides a major leap forward in the extraction of useful data from images plagued by autofluorescence by offering an approach that is easily incorporated into existing workflows in ImageJ, Matlab and R, and that can be used with various samples, staining panels and image acquisition methods.

Funding

This work was funded by the Australian National Health and Medical Research Council (NHMRC) (APP1078697 and APP1181482) and the Australian Research Council (ARC) Discovery Early Career Researcher Award (DE200100944), both funded by the Australian Government

Conflict of Interest: none declared.

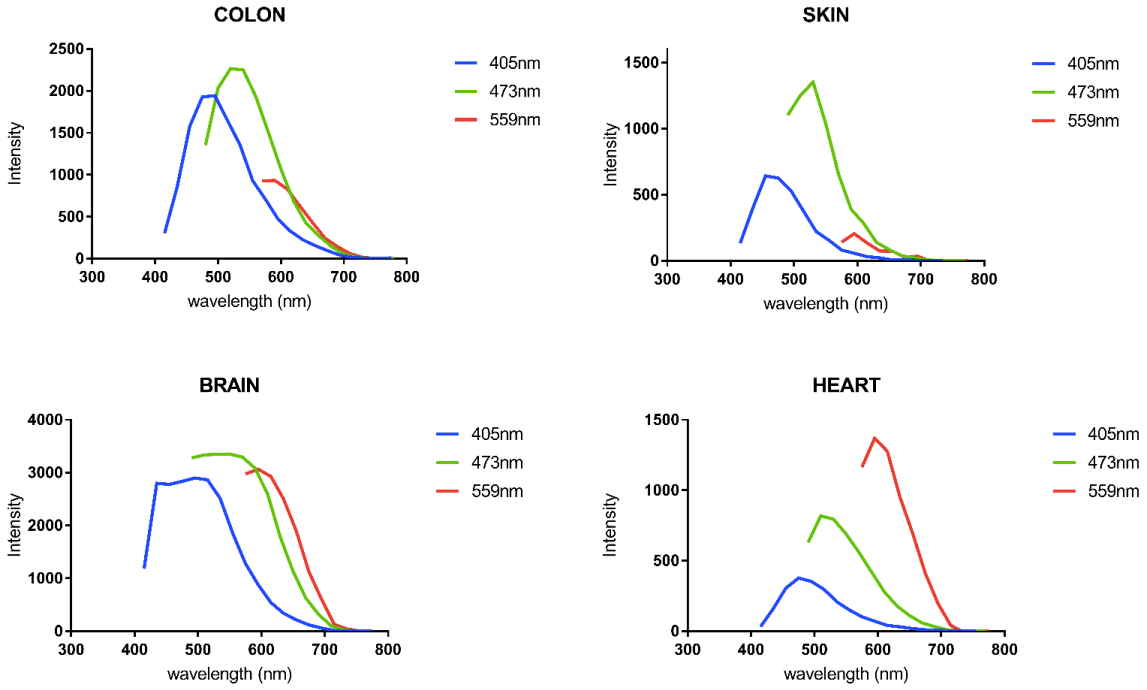
Data availability

The raw data and associated code needed to reproduce the figures in this paper are available at https://zenodo.org/record/3898373#.X04Yyvgzbo3.

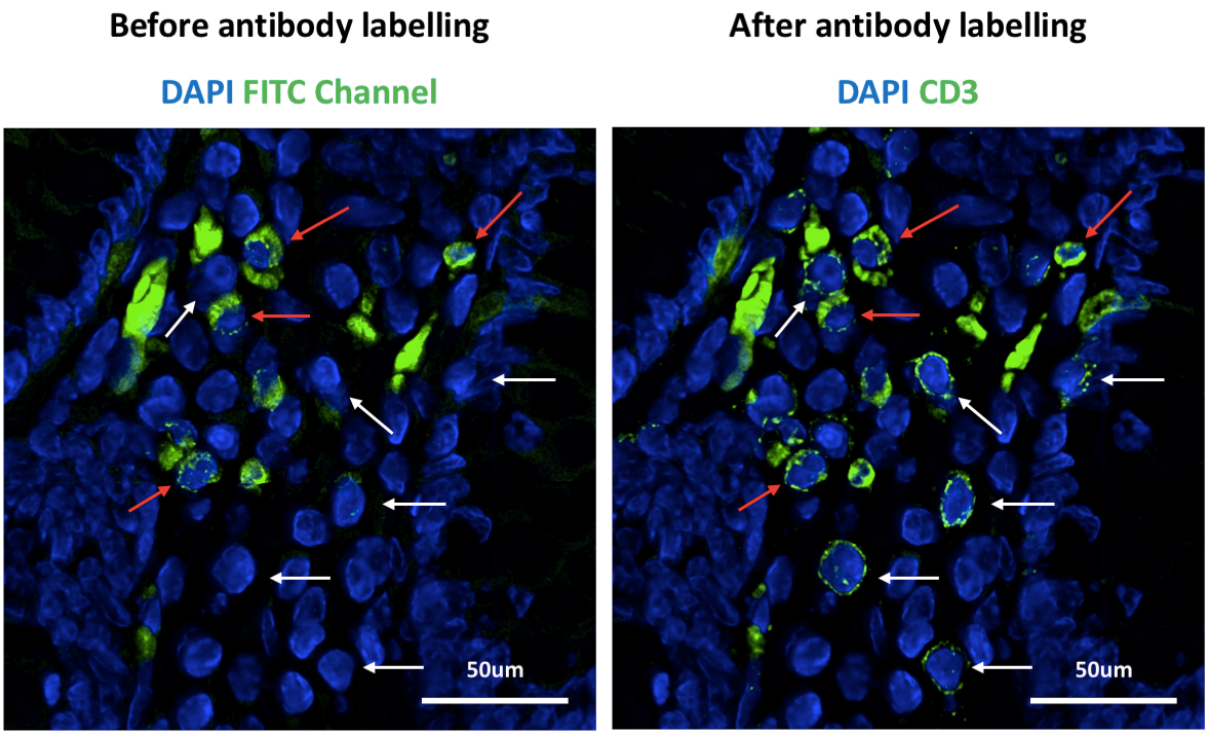
Acknowledgements

The authors acknowledge the Cell Imaging Core Facility at the Westmead Institute for their support and services. Also, Rashid F., member of the Chong Lab, for the rat heart images used for this study, and Bright F., for the human brain tissue images used for this study. Also a special thanks to Cantrill L. for reviewing this manuscript.

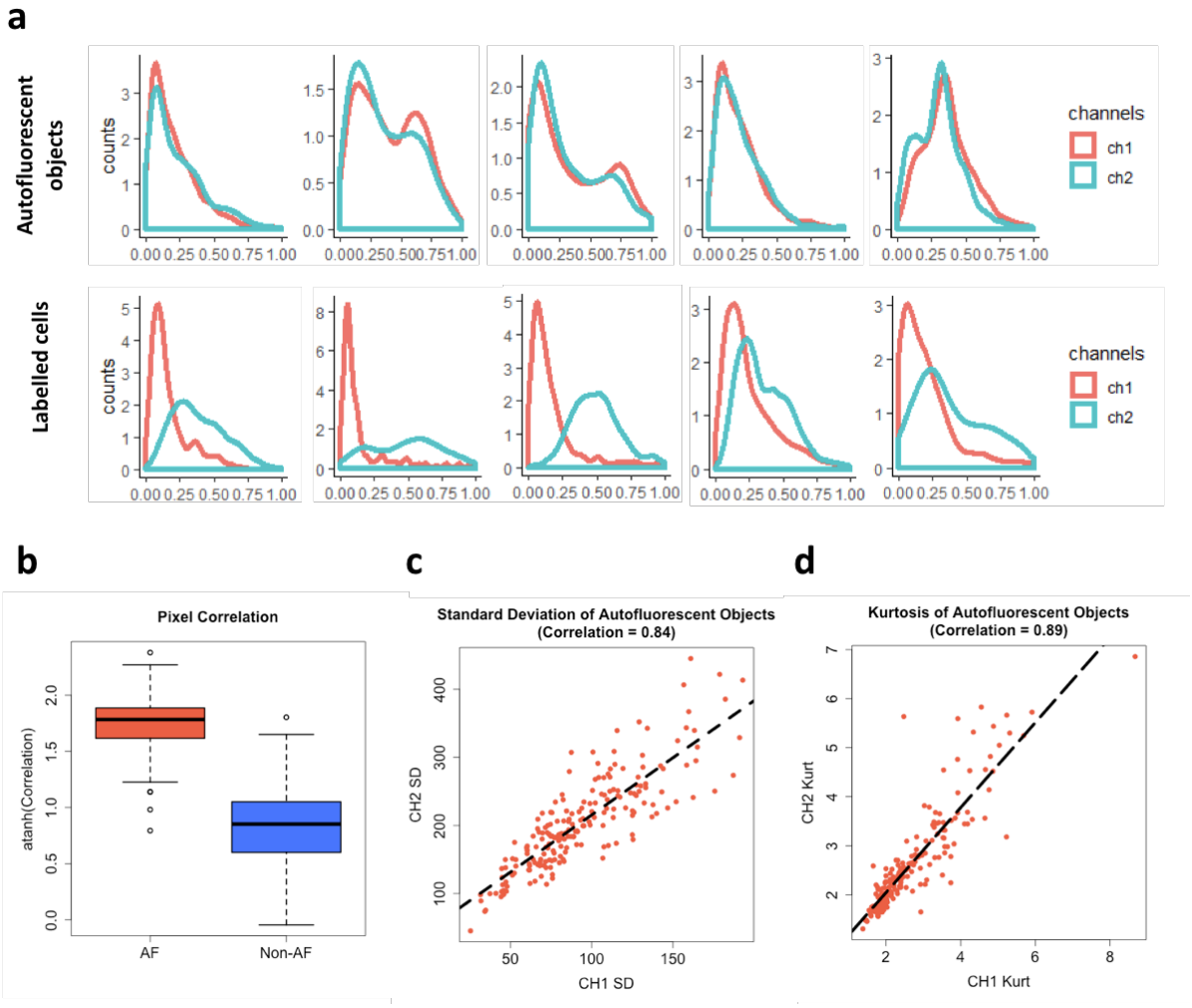
2.5 Supplementary Material



Supplementary Figure 2.1: **Excitation and emission spectra of autofluorescence in various tissues.** The intensity of pixels corresponding to autofluorescent structures measured at 20nm intervals upon excitation with laser lines 405nm, 473nm or 559nm in human colon, skin and brain tissues, as well as rat heart tissue. Results shown as the intensity of the autofluorescent object minus the intensity of the stromal background for each wavelength. Results are shown for a single image for each tissue type.

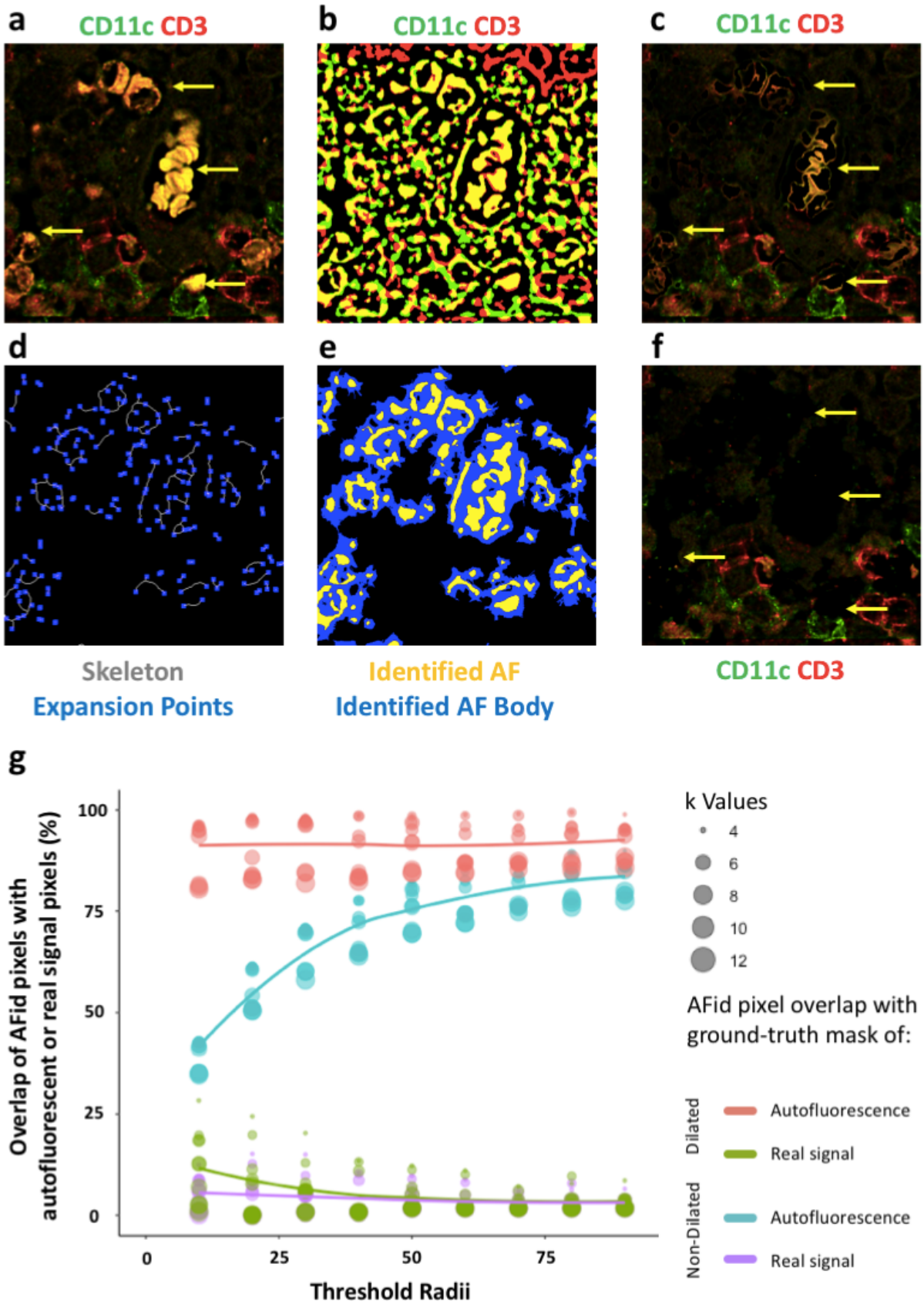


Supplementary Figure 2.2: **Autofluorescence inhibits assessment of CD3 labelling in the human colorectum.** Fixed human colorectal tissue sections imaged prior to (left) and after labelling with mouse anti human CD3 and donkey anti-mouse AF488 (right). Red arrows indicate some autofluorescent cells and white arrows indicate CD3+ cells. Images are representative of 6 unique donors where CD3 staining was performed.



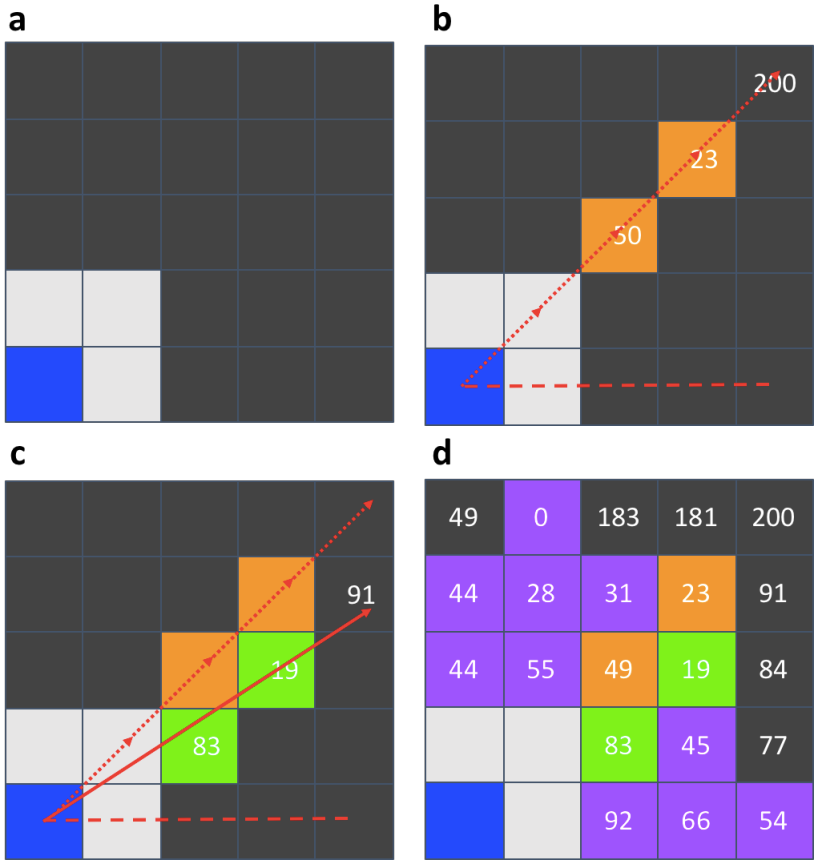
Supplementary Figure 2.3: **Features of autofluorescence are highly correlated between fluorescent channels.** Fixed human colorectal tissue sections were stained for mouse anti CD3 and rabbit anti CD4, detected using donkey anti-mouse AF488 and donkey anti-rabbit AF546 respectively. An intersection mask was created using the two fluorescent markers (Figure 2.1) and measurements performed on objects in the intersection mask. An unstained background image was used as a reference to manually annotate autofluorescent objects in the stained image. **a,** Density plots of pixel intensities in the CD3 channel (Red) and CD4 channel (Blue), for 5 randomly selected autofluorescent objects (top row) or 5 randomly selected labelled cells (bottom row). Pixel intensities for each object in each channel were normalized by subtracting the min pixel value and dividing by the max pixel value minus the min pixel value. **b,** The arctanh transformed Pearson's correlation coefficient values of autofluorescent objects vs non-autofluorescent objects within the intersection mask. The boxplots contain data from thousands of individual objects for each category. Boxplots show the min, first quartile, median, third quartile and max values. **c,d,** standard deviation (SD) and kurtosis (Kurt) measurements of autofluorescent objects in each channel (CH1 and CH2) used to create the intersection mask. A subsample of 200 autofluorescent objects, among thousands, is shown. A line of best fit is shown for both graphs. All graphs are representative of 13 total images used for this work.

This page has been left blank intentionally.



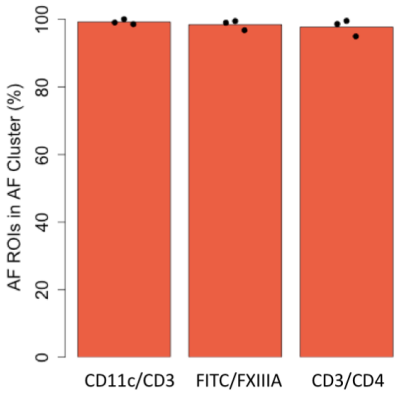
Supplementary Figure 2.4: **Custom dilation function to estimate the correct perimeter of autofluorescent ROIs.** (continued on next page)

Supplementary Figure 2.4: **a,** Fixed colorectal tissue sections were stained for rabbit CD11c and mouse CD3, followed by donkey anti rabbit AF488 and donkey anti mouse AF546. **b,** Fluorescent channels used to detect CD11c and CD3 were thresholded, binary masks created and the composite image displayed. Yellow indicates the overlapping area corresponding to the intersection mask. **c,** Image from part a with the pixels in the intersection mask from part b set to 0. **d,** Identified autofluorescent objects within the intersection mask in part b are skeletonised and points for outward expansion (blue) are distributed along the skeleton every 20 pixels. **e,** Thousands of equiangular lines are drawn outwards from the expansion centres identified in part d, each line propagating until it encounters a pixel brighter than the previous pixel, as measured in either the CD11c or CD3 channel. A mask of the identified autofluorescence body is thus generated for each fluorescent channel. **f,** Pixels corresponding to the identified AF body in each channel in part **e**, are set to 0. Yellow arrows indicate some autofluorescent objects. **g,** Measurement of overlap of autofluorescent pixels defined by AFid, with ground-truth masks of autofluorescence and real signal (CD3 antibody labelled) pixels. Measurements were performed with (red and green lines) and without (blue and purple) employing the custom dilation function and for varying cluster number k and varying threshold radii. Representative data of 3 images is shown. See methods for additional details.



Supplementary Figure 2.5: **Schematic illustrating programmatic steps of custom dilation function.** This figure illustrates how expansion occurs from each of the points in Supplementary Figure 2.4d, resulting in the expanded ROI in Supplementary Figure 2.4e. Each square represents an individual pixel. **a,** Light grey boxes denote pixels belonging to an autofluorescent object. The blue box is a single point from which expansion of the ROI can occur. Dark grey boxes denote the pixels which are not an autofluorescent object. **b,** A straight line is drawn outward from the point of expansion. Once the line reaches the edge of the autofluorescent object (light grey pixels) it will begin to measure the value of pixels that it intersects. The line will continue expanding outward so long as the next pixel value is lower or equal to the previous pixel value. All pixels which meet this condition are recorded (shown here by the orange boxes). **c,** The process in **(b)** is repeated with a line emanating from the expansion point (blue box) at a slightly different angle. This shift in angle is defined by the equation $\cos\left(1-\left(\frac{1}{60}\right)^2/2\right)$. This gives the angle that separates the endpoints of two 60 pixel lines by 1 pixel. Recorded pixels for the new line are shown by the green boxes. **d,** The process in **(a)** and **(b)** is repeated for multiple equiangular lines emanating from the expansion point. Additional recorded pixels are shown in purple. All of the coloured pixels now form the new expanded object, which originally consistent of just the 4 pixels in part **(a)**. This figure shows an example of expansion from just one expansion point within an object. Objects usually have multiple expansion points so when this process is carried out for all expansion points, the collection of recorded pixels forms the new outline of the object as demonstrated in Supplementary Figure 2.4e.

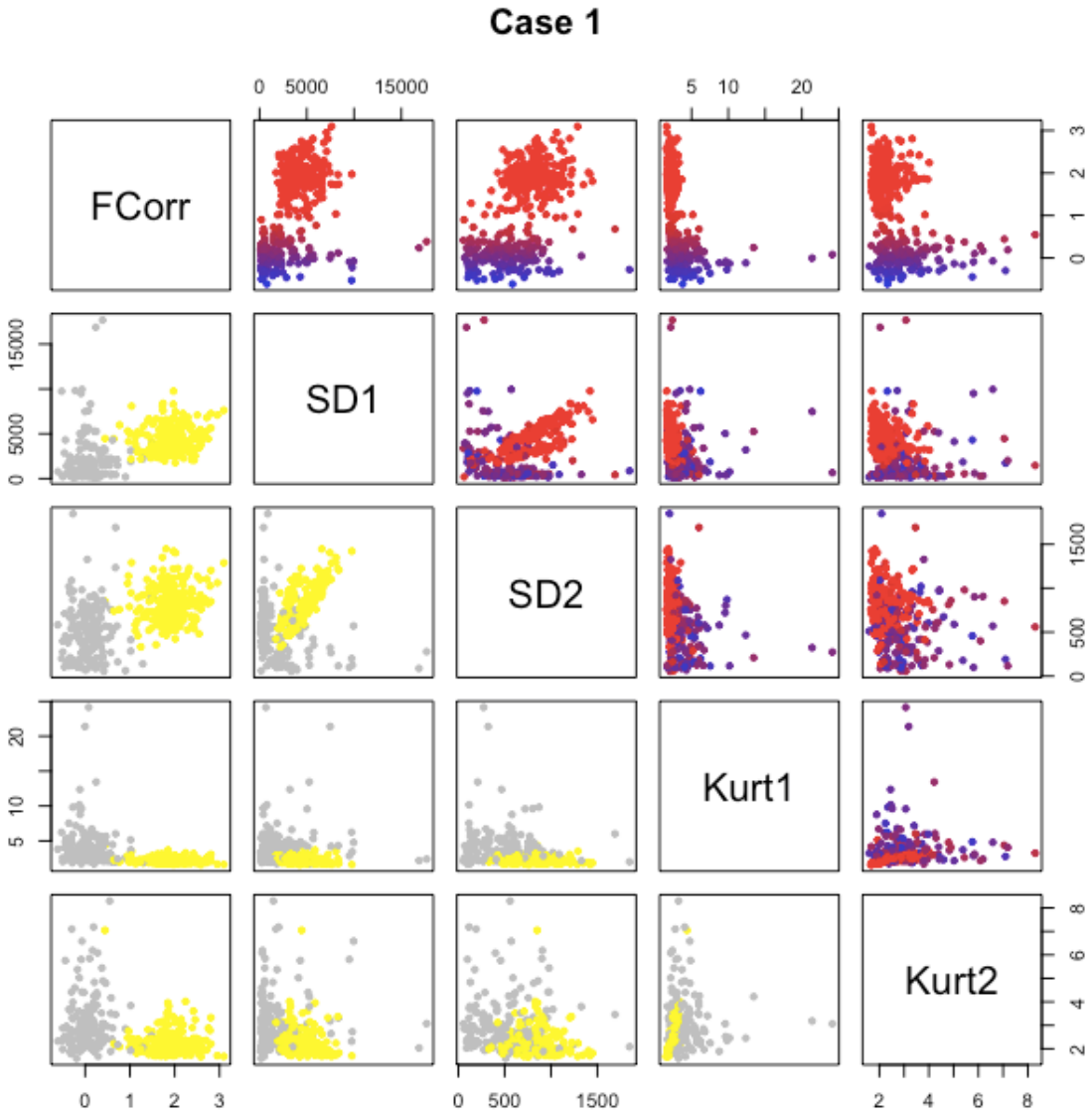
a



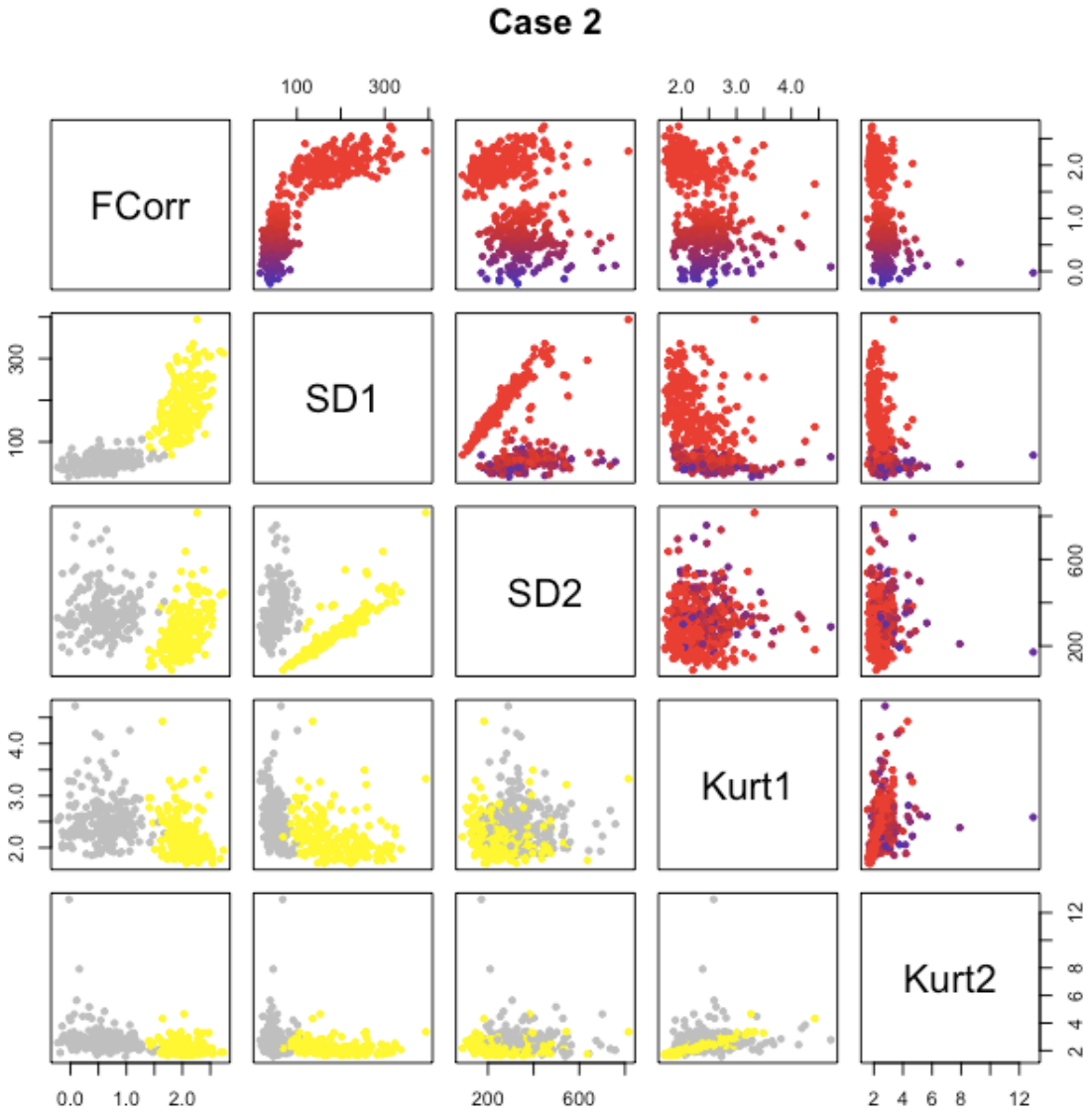
b

Case	Identified AF/Total AF	Identified Real Signal/Total Real Signal	Number of ROIs	Number of AF ROIs
CD11c/CD3	198/198	3/142	15904	1954
CD11c/CD3	182/199	0/193	34298	1373
CD11c/CD3	199/200	2/197	26287	1272
FITC/Mac	207/210	7/181	23456	3552
FITC/Mac	189/204	2/197	21552	2003
FITC/Mac	176/197	1/208	16442	1179
CD3/CD4	199/201	1/205	21355	2111
CD3/CD4	187/198	10/189	20310	1452
CD3/CD4	206/207	3/202	63997	3571

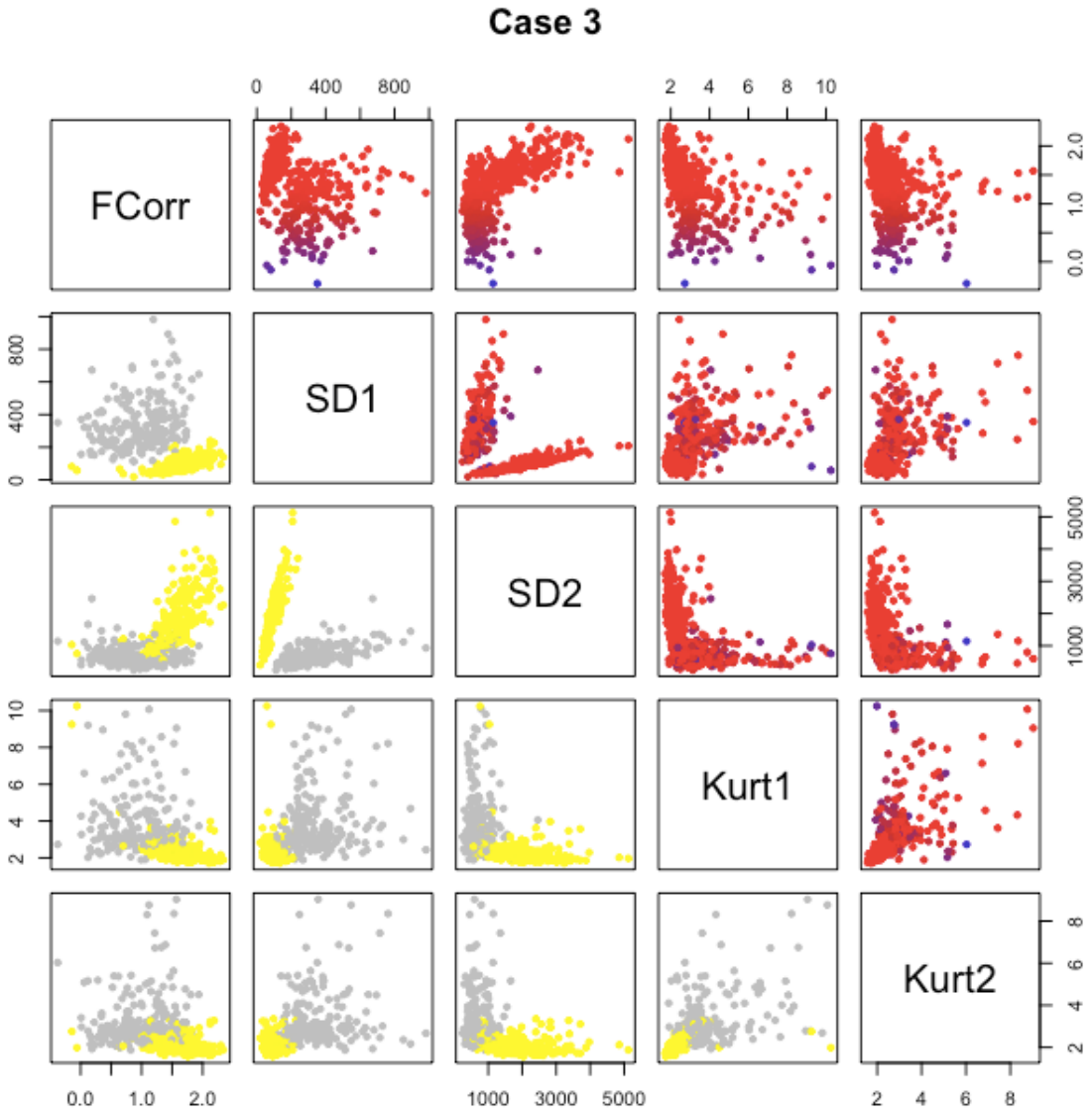
Supplementary Figure 2.6: **Specificity and sensitivity of autofluorescence exclusion for various use-cases in Figure 2.2.** Fixed human colorectal sections were imaged prior to, and after labelling with antibodies against markers for three separate panels, CD11c/CD3, FITC/FXIIIA and CD3/CD4. FITC indicates an unstained open channel that was imaged. An intersection mask was created using the two fluorescent channels for each panel (as in Figure 2.1). Textural features of objects within the intersection mask were then measured for each channel, including standard deviation, kurtosis, as well as the inter-channel Pearson's correlation coefficient of corresponding pixels. k-means clustering was then performed using these features and the cluster with the highest average correlation values was defined as the cluster containing autofluorescent ROIs. A ground truth for the classification of objects as autofluorescence or real signals (stemming from antibodies) was established by manually annotating a subset of up to 200 ROIs each, using the unlabelled background image as a reference. **a,** percentage of the 'autofluorescence cluster' comprised of autofluorescent ROIs (specificity), where the total number of ROIs in the cluster is defined as the sum of autofluorescent ROIs and ROIs stemming from real signal. Each data point represents counts performed on a unique donor for each panel. Mean values across the three donors are indicated above each column. **b,** table summarising the proportion of manually annotated autofluorescence or real signal assigned to the 'autofluorescence cluster' (sensitivity). The last two columns indicate the total number of ROIs and the total number of ROIs classified as autofluorescence respectively. Each row corresponds to results for unique donors for each use-case.



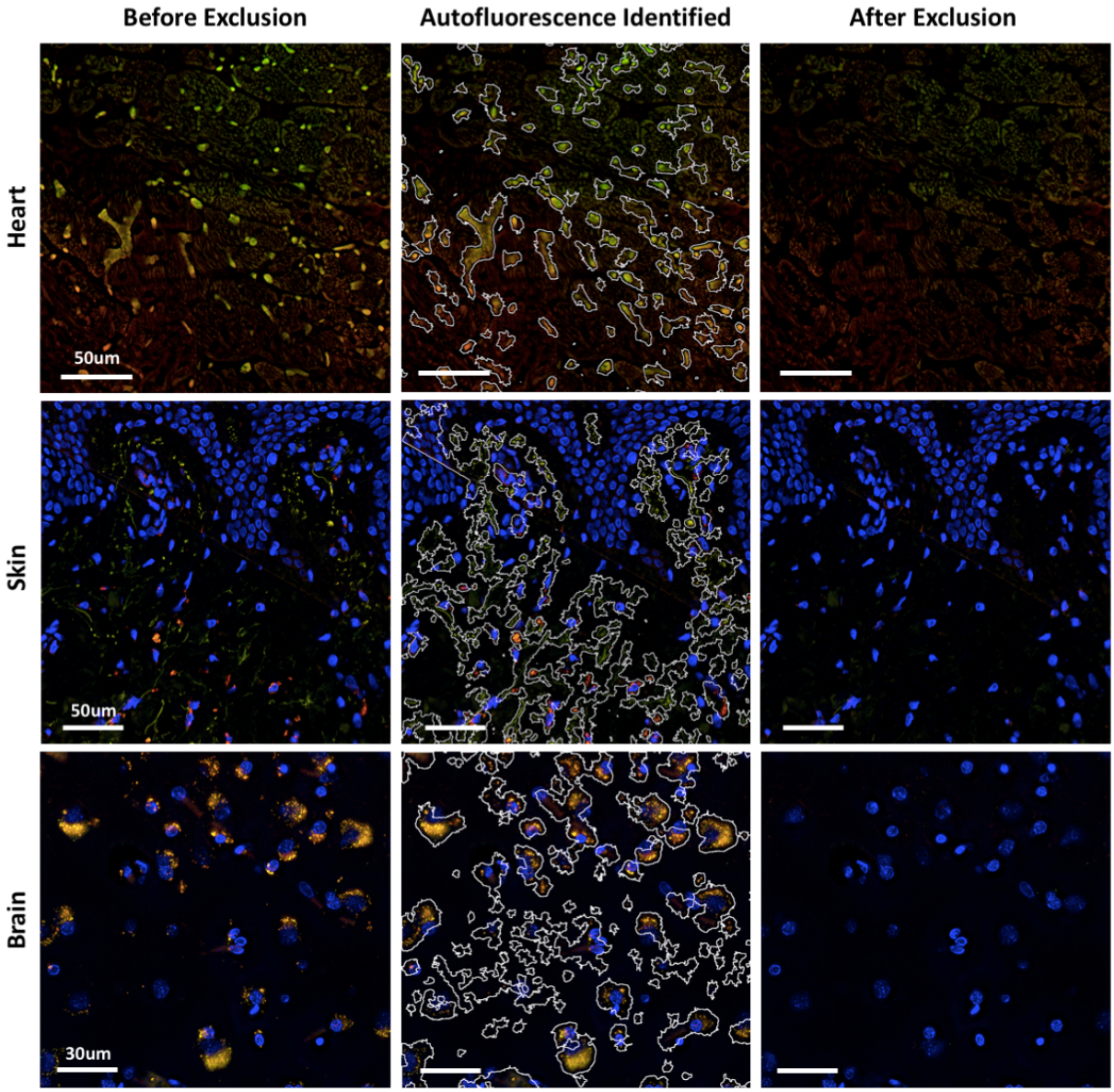
Supplementary Figure 2.7: **Pairwise plots of textural features used for k-means clustering of the non-co-expressed markers use-case.** ROIs from k-means clustering on the CD11c/CD3 use-case in Supplementary Figure 2.5 are shown. In the bottom half, ROIs in the autofluorescence cluster are coloured yellow, whilst non-autofluorescent ROIs (real signal + dim stromal background fluorescence) are coloured grey. The top half shows paired plots as a heatmap of correlation values. FCorr = Arctanh transformed Pearson's correlation coefficient values. SD1, SD2, Kurt1, Kurt2 = Standard deviation or Kurtosis values of ROIs in channels 1 or 2. The plot is representative of clustering performed on 3 independent donors.



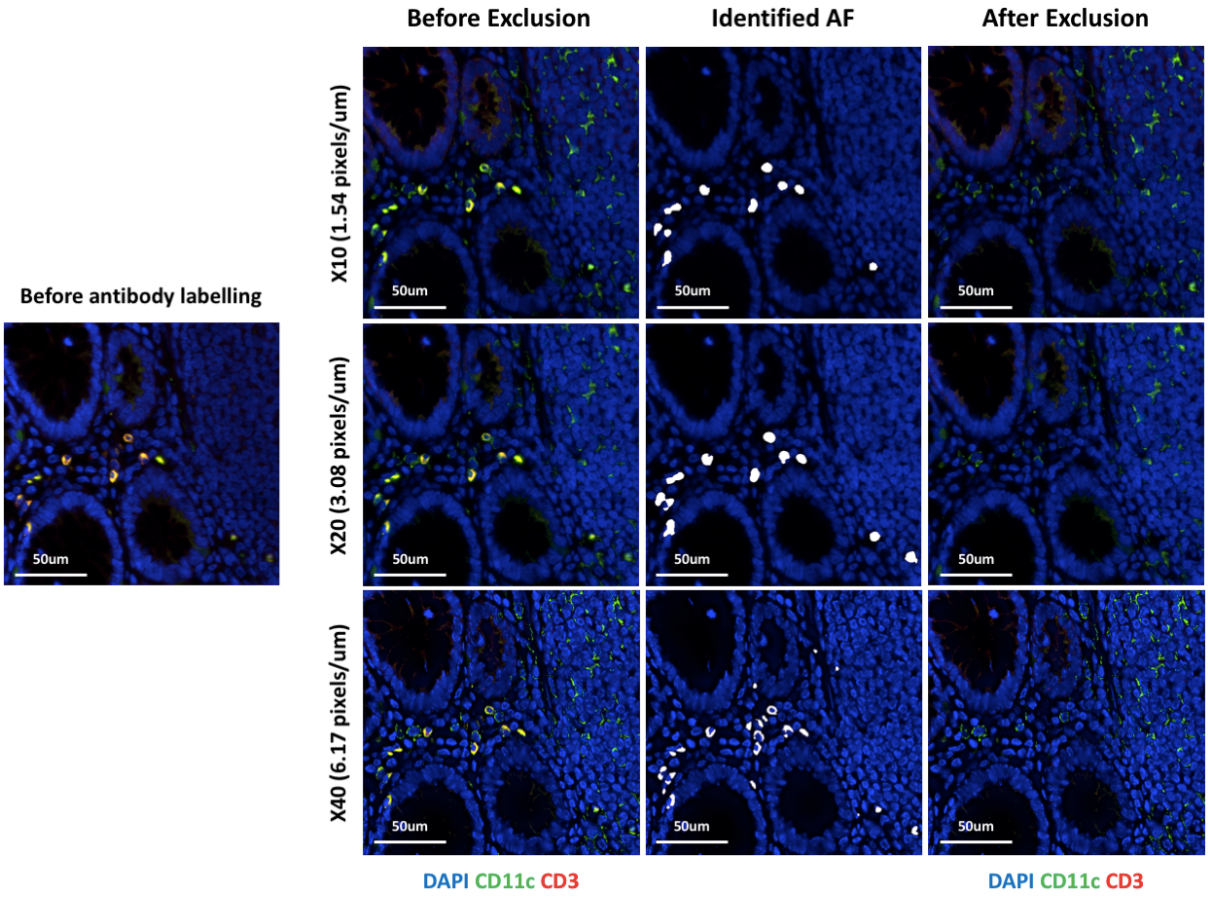
Supplementary Figure 2.8: **Pairwise plots of textural features used for k-means clustering of the autofluorescent cells use-case.** ROIs from k-means clustering on the FITC/FXIIIA usecase in Supplementary Figure 2.5 are shown. In the bottom half, ROIs in the autofluorescence cluster are coloured yellow, whilst non-autofluorescent ROIs (real signal + dim stromal background fluorescence) are coloured grey. The top half shows paired plots as a heatmap of correlation values. FCorr = Arctanh transformed Pearson's correlation coefficient values. SD1, SD2, Kurt1, Kurt2 = Standard deviation or Kurtosis values of ROIs in channels 1 or 2. The plot is representative of clustering performed on 3 independent donors.



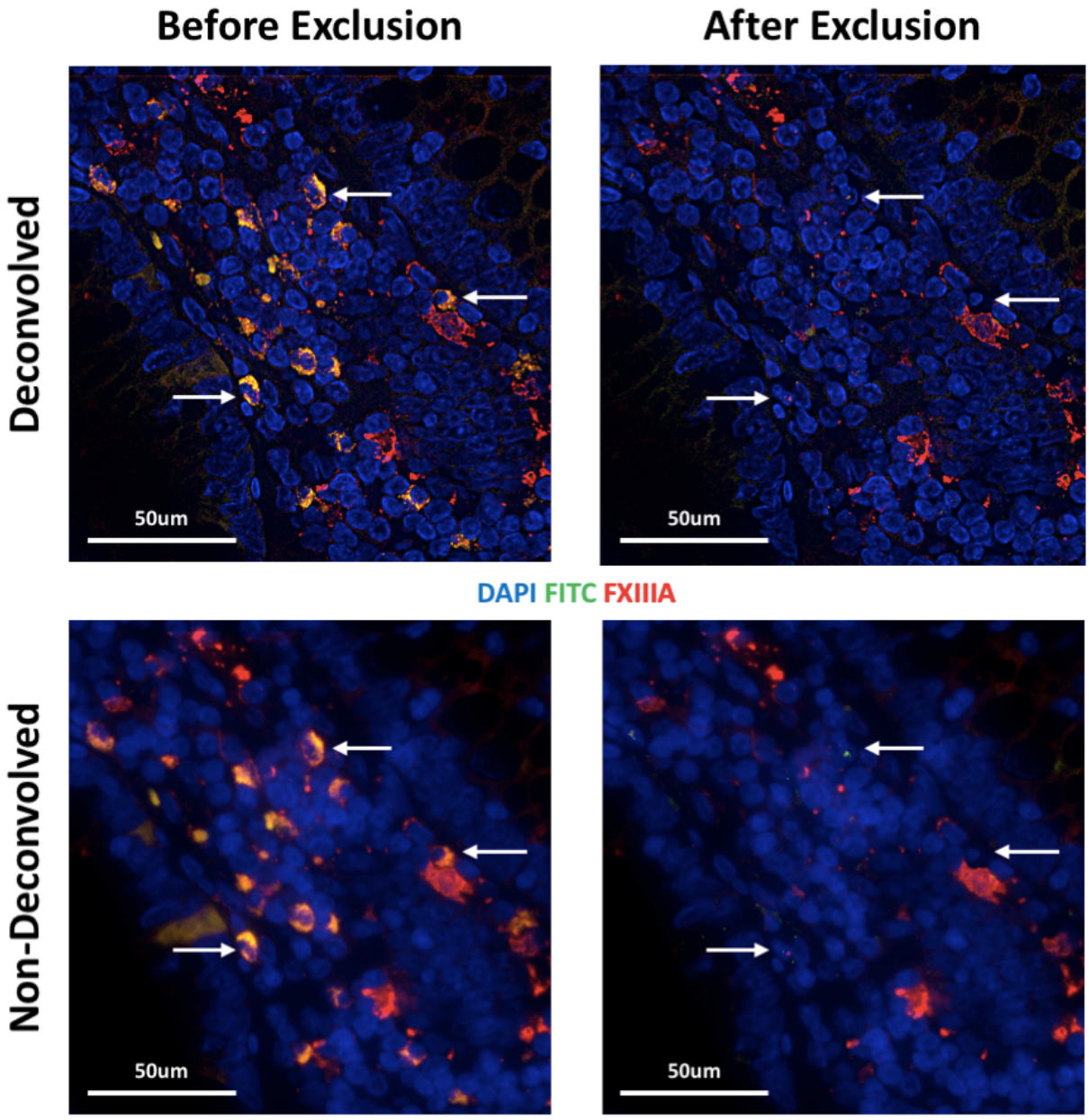
Supplementary Figure 2.9: **Pairwise plots of textural features used for k-means clustering of the co-expressed markers use-case.** ROIs from k-means clustering on the CD3/CD4 use-case in Supplementary Figure 2.5 are shown. In the bottom half, ROIs in the autofluorescence cluster are coloured yellow, whilst non-autofluorescent ROIs (real signal + dim stromal background fluorescence) are coloured grey. The top half shows paired plots as a heatmap of correlation values. FCorr = Arctanh transformed Pearson's correlation coefficient values. SD1, SD2, Kurt1, Kurt2 = Standard deviation or Kurtosis values of ROIs in channels 1 or 2. The plot is representative of clustering performed on 3 independent donors.



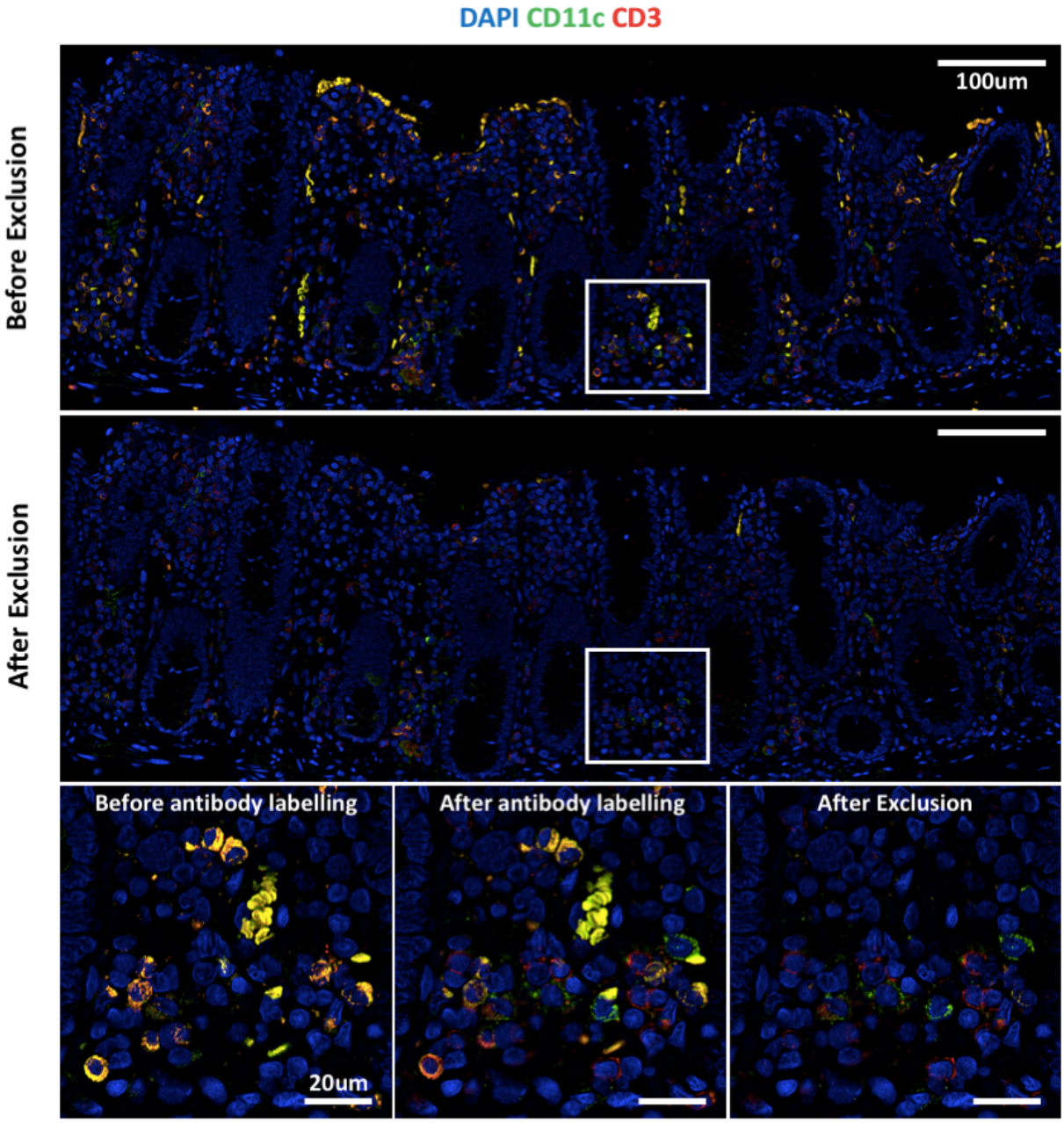
Supplementary Figure 2.10: **AFid performance across various tissue types.** Unlabelled sections of fixed rat heart, human abdominal skin and human brain tissue are shown, prior to autofluorescence exclusion (left panel), with the boundary of autofluorescent structures identified by the algorithm (middle panel), and after exclusion by setting pixel values to 0 (right panel). Note that the images the skin and brain samples were counterstained with DAPI for visualisation. The images are representative areas from one whole-tissue image for each tissue type that was used for autofluorescence identification and exclusion.



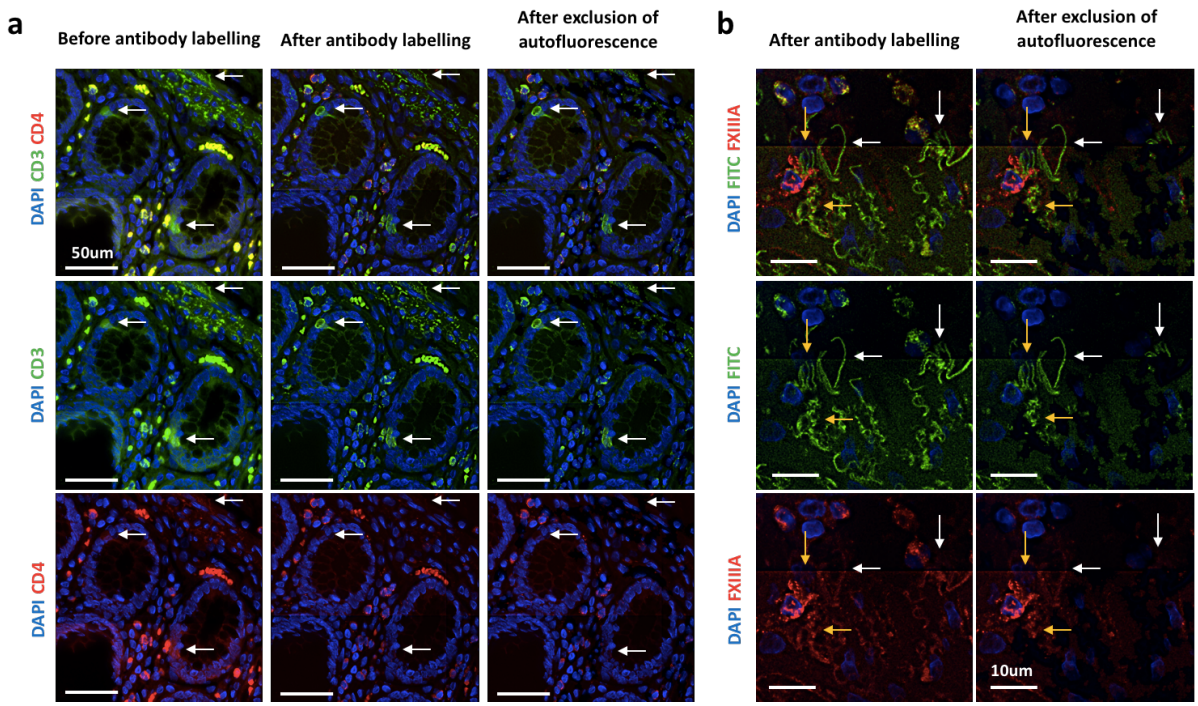
Supplementary Figure 2.11: **AFid performance at varying image resolutions.** Fixed colorectal tissue sections were imaged prior to labelling (left, unlabelled image) and after labelling for rabbit anti CD11c and mouse anti CD3, followed by donkey anti rabbit AF488 and donkey anti mouse AF546. Images of the same area were taken with x10, x20 and x40 objectives with an image resolution of 1.54, 3.08 and 6.17 pixels per µm respectively. Images before autofluorescence exclusion (left panel), with a mask of the identified autofluorescence overlaid (middle panel) and after exclusion by setting pixel values to 0 (right panel) are shown. Images are representative of 3 unique donors where CD11c/CD3 staining was carried out and imaged at various magnifications.



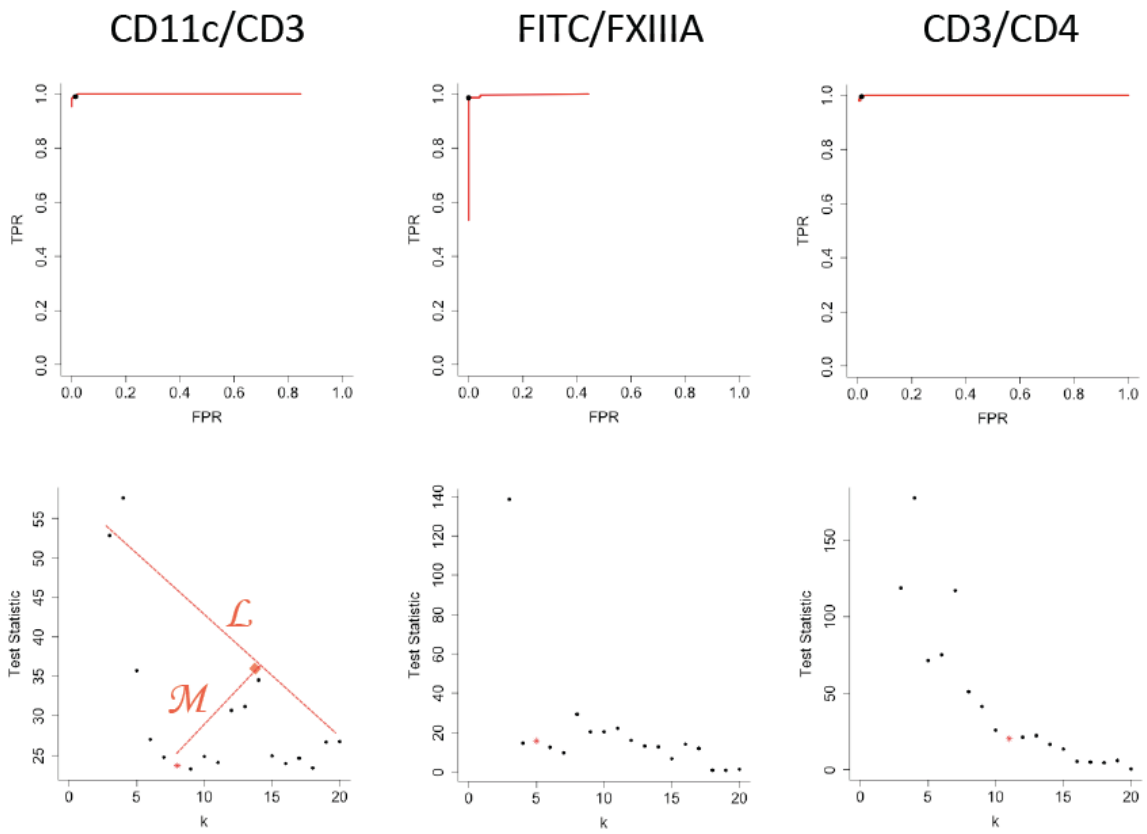
Supplementary Figure 2.12: **AFid performance before and after deconvolution.** Fixed colorectal tissue sections were labelled with a sheep anti FXIIIA antibody followed by donkey anti sheep AF546. 'FITC' is the FITC channel, which was imaged but not used to detect any markers. Images are shown of the same area before (bottom row) and after (top row) deconvolution using Huygens deconvolution software, CMLE algorithm. Images before (left panel) and after (right panel) autofluorescence exclusion are shown. White arrows indicate some autofluorescence identified by the algorithm. Images are representative of 3 unique donors stained for FXIIIA and processed using AFid before and after deconvolution.



Supplementary Figure 2.13: **AFid performance on large images.** Fixed colorectal tissue sections were imaged prior to and after labelling for rabbit anti CD11c and mouse anti CD3 antibodies, followed by donkey anti rabbit AF488 and donkey anti mouse AF546. A large area of tissue was imaged and the results before (top panel) and after (middle panel) autofluorescence exclusion are shown. Zoomed in images of the area outlined are shown with an additional image of the unlabelled section outlining the distribution of autofluorescence. Image is representative of 3 unique donors where CD11c/CD3 staining was carried out.



Supplementary Figure 2.14: **Instances of failed autofluorescence identification.** Sections of fixed human colorectal tissue prior to (**a**, left column) and after labelling (**a**, middle column and **b**, left column) with antibodies targeting the indicated markers. Labelled images after autofluorescence exclusion (setting pixel values to 0) are shown by the right most panel in **(a)** and **(b)**. White arrows indicate autofluorescent objects that were not identified by AFid due to the fluorescence only appearing the green channel (CD3 for part a, FITC for part **b**) and were therefore not excluded by the algorithm. Yellow arrows show autofluorescent muscle fibers that were not identified by AFid due to overlap with a FXIIIA+ macrophage. In the middle row of **(b)** 'FITC' is the FITC channel, which was imaged but not used to detect any markers. Note, the images in the leftmost column in **(a)** were not deconvolved as only a single z-plane was acquired to show the distribution of autofluorescence. As such the image appears less clear than the other images of the same area in **(a)**.



Supplementary Figure 2.15: **Estimation of optimal cluster number for k-means.** k-means clustering as described in Supplementary Figure 2.5 was performed iteratively for 3-20 clusters and the distribution of paired true positive rate (TPR) and false positive rate (FPR) values for each cluster number is indicated by the red line for each use-case (top row). A high TPR or FPR corresponds to a high proportion of the 'autofluorescence cluster' comprising manually annotated autofluorescence or real signals respectively. The bottom row shows plots of the cluster number (k) versus, test statistic of a t-test (two-tailed) comparing Pearson's correlation coefficient values of clusters with the highest ('autofluorescence cluster') and second highest average correlation values. An elbow method approach for estimation of optimal k is illustrated in the bottom left plot. A line (L) is drawn between the first and last plotted values. The line M indicates the plotted value that is below the line L, and has the greatest perpendicular distance to that line. The cluster number corresponding to this plotted point is estimated as the optimal cluster number. The points with optimal cluster number for each plot are indicated as a red *. The TPR/FPR of the optimal cluster number for each use-case is indicated by the black dot in the top row of plots. Data are representative of analysis performed on 3 unique donors for each use-case.

This page has been left blank intentionally.

Supplementary Table 2.1: **Table of frequently asked questions and their answers relating 0 to the use of AFid.**
*Details on how to implement parameter changes are in the user manual accompanying the plugin.

FREQUENTLY ASKED QUESTIONS	ANSWERS
Can I remove autofluorescence if it appears in only one channel?	No. However, if the autofluorescent signal appears in a second channel, even just faintly, AFid may still be able to identify the autofluorescent object.
I have acquired one channel to detect a marker of interest, as well as DAPI. Can these two channels be used with AFid?	DAPI usually cannot be used as a second comparison channel as it stains all cells and is usually very bright. As such it will either overlap with most autofluorescence in the tissue or will be far brighter than autofluorescence making it difficult to threshold on autofluorescent objects which is the required first step in autofluorescence classification.
Can I use AFid to remove stromal autofluorescence that is present throughout the whole image?	Some tissues exhibit strong stromal autofluorescence that is present throughout the entire tissue. As this signal is present throughout the whole tissue and is not in the form of discrete objects, it is unlikely that AFid will be able to successfully identify and exclude such autofluorescence. If this signal significantly hinders analysis then alternative methods should be explored such as chemical quenching, background subtraction or spectral unmixing.
Will AFid remove autofluorescent objects that are overlapping with my antibody/RNA Fish derived signal?	It depends. Since AFid is a clustering method, it is possible that some objects with overlapping autofluorescence and probe signal will be identified and excluded. Users can adjust the default parameters of AFid to avoid this scenario. Two possible adjustments: (1) Manually specify the number of clusters for k-means, choosing a greater number of clusters. This will result in a greater specificity at the expense of sensitivity.* (2) Specify a minimum Pearson's Correlation Coefficient value for an object to be classified as autofluorescence. If this is done with clustering then only objects within the cluster which are above the specified value will be identified as autofluorescence.*
Which channels should I use for Autofluorescence exclusion?	Neighbouring channels are usually best, but the algorithm does work for disparate channels (e.g FITC vs Cy5 channel). The major factor is whether autofluorescent signals span multiple channels. It is best to acquire all channels during acquisition, even channels for which no signal is being measured. These spare channels can be used for autofluorescence detection with AFid.
Is AFid compatible with high parameter cyclic microscopy protocols?	Yes. The bleaching of tissues between rounds of staining will usually reduce autofluorescence, enabling better measurement of desired of signal of interest. However bleach-sensitive epitopes maybe acquired in the first round of staining prior to bleaching. An example is CD4 which is sensitive to H2O2, a chemical used for bleaching. As such, AFid can be used to mitigate autofluorescence in early rounds of staining, prior to the use of bleaching solutions. Also, it is best to use AFid to compare channels acquired in one specific round of staining. This is because registration software rarely achieves pixel-perfect registration and this can affect the cross-channel comparisons needed for autofluorescence classification.

Supplementary Table 2.1: **continued from previous page**

FREQUENTLY ASKED QUESTIONS	ANSWERS
Why is AFid not capturing all autofluorescent objects that I wish to exclude?	There are multiple possible reasons for this. A few possibilities are listed here: (1) The objects are not present in the intersection mask used to classify objects. Users can adjust thresholding parameters in the AFid plugin or simply provide their own intersection mask for the program to then classify autofluorescent objects within this mask.* (2) The algorithm has split the autofluorescence cluster up into multiple clusters. There are two options to remedy this. One, manually set a lower cluster number k. Two, use a Pearson Correlation Coefficient cut-off value instead of clustering, specifying a low enough value to capture all autofluorescence (this will need to be done through a bit of trial and error).* (3) Image alignment issue. Microscopes sometimes have calibration issues which result in misalignment across channels of a given images. The best solution is to calibrate the microscope. It may be possible to rescue the images by using a post-acquisition alignment software. This is possible if there are enough similar landmarks between channels.
Why is AFid detecting and excluding too much of my desired signal of interest?	If using clustering then manually set a higher cluster number k. If using a Pearson's Correlation Coefficient cut-off value instead of clustering, specifying a higher value which will result in higher specificity for autofluorescence over probe signal.
How do I detect and exclude autofluorescence from more than 2 channels?	Run AFid on multiple channel pairs. For example (FITC vs Cy3, Cy3 vs Cy5). The user can then save the masks of identified autofluorescence in each the FITC, Cy3 and Cy5 channels.
I'm experiencing signal bleedthrough across channels. Will AFid remove these signals?	This could happen. Generally signal bleed through will cause a variety of issues in image processing and analysis. There are two possible methods to mitigate this issue: (1) Run AFid using two other channels where signal bleed-through is not present. (2) Recognize that bleed-through usually occurs when the signal in one channel is very bright. As such, when users create the intersection mask in AFid, they could measure the mean fluorescence intensity of all objects in the bright channel and remove objects which are above a user specified cut-off. In this way bright-bleed through signals will not be present in the mask which is used for autofluorescence classification and so will not be identified and excluded by AFid. We have used this method with good success when we occasionally see saturated bright signals by RNAscope due to a high concentration of an RNA target in one area of the image.
What image types are compatible with AFid?	The program was designed to be used with 2D 16 bit tiff images. If you have a stack of images you will need to split them into individual images prior to running AFid. 8 bit tiffs can be used but results may be suboptimal. If a 3D Z-stack has been captured then images should be maxprojected prior to running AFid.

Algorithm 1: AFid Main

Inputs : Image1 - first image
Image2 - second image
σ - width of Gaussian blur
minSize - minimum ROI area
maxSize - maximum ROI area
correlationCutoff - cutoff correlation for AF classification
cluster - boolean determining whether or not to perform clustering
kInput - number of clusters to use if cluster==TRUE
estimate - boolean determining whether or not to estimate the number of clusters
minK - minimum k if estimate==TRUE
maxK - maximum k if estimate==TRUE

Output: im1 - read in Image1

```
/* Read in images                                                              */
im1 = readImage(Image1)
im2 = readImage(Image2)

/* Gaussian blur images                                                        */
im1Blurred = GaussianBlur(im1, σ)
im2Blurred = GaussianBlur(im2, σ)

/* Create binary masks with local threshold                                    */
im1Mask = LocalThreshold(im1Blurred, method, radius)
im2Mask = LocalThreshold(im2Blurred, method, radius)

/* Create AND intersection mask                                                */
intersectionMask = (im1 AND im2)

/* Identify intersection ROIs                                                  */
roiFiltered = GetIntersectionROIs(intersectionMask, minSize, maxSize) // See Algorithm 2

/* Get measurements for classification                                         */
correlation, sd1, sd2, kurt1, kurt2 = GetMeasurements(im1, im2, roiFiltered) // See Algorithm 3

/* Classify ROIs as AF or Non-AF                                               */
roiAF = ClassifyROIs(correlationCutoff, cluster, kInput, estimate, minK, maxK, correlation, sd1, sd2, kurt1,
    kurt2) // See Algorithm 4

/* Remove AF ROIs                                                              */
im1AFRemoved = im1
im2AFRemoved = im2
for i in 1:length(roiAF) do
    if roiAF[i] == TRUE then
        im1AFRemoved[roiFiltered[i]] = 0 // Set pixel intensities to zero if contained in an AF ROI
        im2AFRemoved[roiFiltered[i]] = 0
    end
end

/* Perform expansion algorithm                                                 */
expansionMask1, expansionMask 2 = AFExpansion(roiFiltered, roiAF) // See Algorithm 6

/* Remove areas identified from expansion                                      */
im1AFRemoved(expansionMask1 == TRUE) = 0
im2AFRemoved(expansionMask2 == TRUE) = 0
```

Algorithm 2: Identify intersection ROIs

Inputs : intersectionMask - intersection of the two local threshold binary masks
minSize - minimum ROI area
maxSize - maximum ROI area
Output: roiFiltered - array of ROIs to consider for AF removal

```
1 Function GetInteresectionROIs(intersectionMask, minSize, maxSize)
      /* Get all ROI objects                                                    */
2     roiArray = GetRois(intersectionMask) // Each ROI in roiArray is an 2 × n matrix of each pixel
          co-ordinate of each ROI
3     roiFiltered = [] // Empty array of ROI
4     for i in 1:length(roiArray) do
5         if (ncol(roiArray[i]) ≤ maxSize) AND (ncol(roiArray[i]) ≥ minSize) then
6             roiFiltered.add(roiArray[i]) // Create new array which contains all the ROIs filtered by
                  size
7         end
8     end
9     return roiFiltered
```

Algorithm 3: Get measurements for classification

Inputs : cluster - boolean determining whether or not to perform clustering
roiFiltered - array of ROIs to consider for AF removal
im1 - read in Image1
im2 - read in Image2

Outputs: correlation - array of pixel correlations across the two images for each ROI
sd1 - array of pixel standard deviation for each ROI in im1
sd2 - array of pixel standard deviation for each ROI in im2
kurt1 - array of pixel kurtosis for each ROI in im1
kurt2 - array of pixel kurtosis for each ROI in im2

```
1  Function GetMeasurements(im1, im2, roiFiltered)
       /* Initialise arrays                                                    */
2      correlation = []
3      sd1 = []
4      sd2 = []
5      kurt1 = []
6      kurt2 = []

       /* Obtain measurements                                                  */
7      for i in 1:length(roiFiltered) do
8          roi = roiFiltered[i]
9          im1Pixels = im1[roi] // im1[ROI] gets the intensity value at each pixel coordinate given by
               ROI
10         im2Pixels = im2[roi]
11         correlation.add(measureCorrelation(im1Pixels, im2Pixels)) // Measure correlation of the pixel
               intensities between the two image channels
12         sd1.add(measureSD(im1Pixels)
13         sd2.add(measureSD(im2Pixels) // Measure standard deviation for each channel
14         kurt1.add(measureKurt(im1Pixels))
15         kurt2.add(measureKurt(im2Pixels)) // Measure kurtosis for each channel
16     end

17     return correlation, sd1, sd2, kurt1, kurt2
```

Algorithm 4: Classify ROIs as AF or non-AF

Inputs : correlationCutoff - cutoff correlation for AF classification
cluster - boolean determining whether or not to perform clustering
kInput - number of clusters to use if cluster==TRUE
estimate - boolean determining whether or not to estimate the number of clusters
minK - minimum k if estimate==TRUE
maxK - maximum k if estimate==TRUE
correlation - array of pixel correlations across the two images for each ROI
sd1 - array of pixel standard deviation for each ROI in im1
sd2 - array of pixel standard deviation for each ROI in im2
kurt1 - array of pixel kurtosis for each ROI in im1
kurt2 - array of pixel kurtosis for each ROI in im2
roiFiltered - array of ROIs to consider for AF removal
im1 - read in Image1
im2 - read in Image2

Outputs: roiAF - array of booleans specifying whether an ROI is classified as AF
im1AFRemoved - Image1 with AF ROI pixels set to 0
im2AFRemoved - Image2 with AF ROI pixels set to 0

```
1  Function ClassifyROIs(correlationCutoff, cluster, kInput, estimate, minK, maxK, correlation, sd1,
     sd2, kurt1, kurt2)
       /* Identify objects greater than correlation cutoff                                    */
2      correlationAF = (correlation ≥ correlationCutoff) // Logical array specifying whether an ROI is above
         the correlation cutoff

       /* Perform k-means clustering to identify AF objects                                  */
3      if doKMeans == TRUE then
4          if estimate == TRUE then
5              k = EstimateK(minK, maxK, atanh(correlation), sd1, sd2, kurt1, kurt2) // See Algorithm 5
6          else
7              k = kInput
8          end
9          clusters = kMeans(k, atanh(correlation), sd1, sd2, kurt1, kurt2)
10         afCluster = whichGreatestMean(correlation, clusters) // Calculates mean correlation for each
             cluster, then returns the cluster corresponding to the greatest correlation
11         kMeansAF = (clusters == afCluster) // Logical array specifying whether an ROI is in afCluster
12     end

       /* Identify AF ROIs                                                                   */
13     if doKMeans == TRUE then
14         roiAF = (correlationAF AND kMeansAF)
15     else
16         roiAF = correlationAF
17     end

18     return roiAF
```

Algorithm 5: Estimate k

Inputs : minK - minimum k
maxK - maximum k
Output: kBest - estimated k for clustering

```
1  Function EstimateK(minK, maxK)
       /* Measure t statistics                                                    */
2      tStats = [] // Empty array of t statistics
3      for i = minK:maxK do
4          clusters = kMeans(i, atanh(correlation), sd1, sd2, kurt1, kurt2)
5          afCluster = whichGreatestMean(correlation, clusters) // Identify cluster with highest mean
               correlation
6          afCluster2 = whichGreatestMean(correlation[clusters != afCluster], clusters[clusters != afCluster])
               // Identify cluster with second-highest mean correlation
7          tStats.add(measureTStatistic(atanh(correlation[clusters == afCluster]), atanh(correlation[clusters ==
               afCluster2]))) // Measure t statistic of the correlations between the top two clusters
8      end

       /* Perform elbow test                                                      */
9      x1 = 1
10     x2 = length(tStats)
11     y1 = tStats[1]
12     y2 = tStats[-1]
13     kBest = whichMax((y2-y1)*(x1:x2) - (x2-x1)*tStats + x2*y1 - y2*x1)]
14     kBest = kBest - minK + 1

15     return kBest
```

Algorithm 6: Expansion algorithm

Inputs : roiFiltered - array of ROIs to consider for AF removal
roiAF - array of booleans specifying whether an ROI is classified as AF
Output: -

```
1 Function AFExpansion(roiFiltered, roiAF)
      /* Perform skeletonisation                                                    */
2     afMask = createMask(roiFiltered[roiAF]) // Create a binary mask of the identified AF ROIs
3     afSkel = skeleton(afMask) // Perform skeletonisation on binary mask

      /* Identify expansion points                                                  */
4     expansionPoints = IdentifyExpansionPoints(roiSkeleton) // See Algorithm 7

      /* Perform expansion                                                          */
5     expansionMask1, expansionMask2 = Expansion(expansionPoints, im1Blurred, im2Blurred) // See
        Algorithm 8
```

Algorithm 7: Identify expansion points

Inputs : roiSkeleton - skeleton of AF ROIs
Output: expansionPoints - points of expansion

```
Function IdentifyExpansion(roiSkeletonPoints)
    /* Identify end nodes                                                    */
    [r,c] = size(afMask) // Obtain dimensions of the original image
    endNodes = newMask(r,c) // Create empty binary mask of the same size as afMask

    for i in 1:r do
        for j in 1:c do

            if afSkel(r,c) > 0 then
                rNeighbour = [i-1, i, i+1, i-1, i+1, i-1, i, i+1]
                cNeighbour = [j-1, j-1, j-1, c, c, c+1, c+1, c+1]
                neighbourSum = 0

                for k in 1:8 do
                    if afSkel(rNeighbour[k], cNeighbour[k]) then
                        neighbourSum += 1
                    end
                end

                if neighbourSum < 2 then
                    endNodes(i,j) = TRUE // If the number of neighbours is less than two, then that
                        point represents the end of a skeleton
                end
            end
        end
    end

    /* Identify points of expansion by tracing                               */
    currentTrace = endNodes; // Points being currently traced over
    totalTrace = newMask(r,c) // All points that have been traced over
    expansionPoints = endNodes
    traceCount = 0 // Number of tracing steps performed
    traceSensitivity = 20 // Number of tracing steps before an expansion point is added

    while TRUE do

        traceCount += 1
        newTrace = newMask(r,c)
        for i in 1:r do
            for j in 1:c do

                if currentTrace(i,j) == TRUE then
                    totalTrace(i,j) = TRUE
                    rNeighbour = [i-1, i, i+1, i-1, i+1, i-1, i, i+1]
                    cNeighbour = [j-1, j-1, j-1, c, c, c+1, c+1, c+1]

                    for k in 1:8 do
                        if (afSkel(rNeighbour[k],cNeighbour[k]) == TRUE) AND
                          (totalTrace(rNeighbour[k],cNeighbour[k]) == FALSE) AND
                          (currentTrace(rNeighbour[k],cNeighbour[k]) == FALSE) AND
                          (newTrace(rNeighbour[k],cNeighbour[k]) == FALSE) then
                            newTrace(rNeighbour[k], cNeighbour[k]) = TRUE

                            if mod(traceCount, traceSensitivity == 0 then
                                expansionPoints(rNeighbour[k], cNeighbour[k]) = TRUE
                            end
                        end
                    end
                end
            end
        end

        currentTrace = newTrace

        if sum(currentTrace) == 0 then
            break // Break if no new points can be traced
        end
    end

    return expansionPoints
```

Algorithm 8: Perform Expansion

Inputs : expansionPoints - points of expansion
im1Blurred - Image1 Blurred
im2Blurred - Image2 Blurred
Output: expansionMask1 - expansion mask for Image1
expansionMask2 - expansion mask for Image2

```
1  Function Expansion(expansionPoints, im1Blurred, im2Blurred)
2      expandDistance = 60
3      expandWidth = 1
4      theta = acos(1-(expandWidth/expandDist)2/2)
5      minSteps = 3
6      maxSteps = 30
7      expansionMask1 = newMask(r,c)
8      expansionMask2 = newMask(r,c)
9      for i in 1:r do
10         for j in 1:c do
11             if expansionPoints(i,j) == TRUE then
12                 thetaMultiple = 1
13                 thetaCurrent = theta
14                 while thetaCurrent < 2*pi do
15                     thetaMeasure = theta*thetaMultiple
16                     xStepSize = cos(thetaMeasure)
17                     yStepSize = sin(thetaMeasure)
18                     expandLength = 1
19                     expandSteps = 0
20                     done1 = FALSE
21                     done2 = FALSE
22                     while TRUE do
23                         xMeasure = floor(i + expandLength*xStepSize)
24                         yMeasure = floor(j + expandLength*yStepSize)
25                         xCompare = floor(i + (expandLength-1)*xStepSize)
26                         yCompare = floor(j + (expandLength-1)*yStepSize)
27                         if afMask(xMeasure,yMeasure) == TRUE then
28                             expandSteps += 1
29                             pixelDifference1 = im1Blurred(xMeasure,yMeasure) -
                                im1Blurred(xCompare,yCompare)
30                             pixelDifference2 = im2Blurred(xMeasure,yMeasure) -
                                im2Blurred(xCompare,yCompare)
31                             if (pixelDifference1 ≤ 0) OR ((expandSteps < minSteps) AND (done1 ==
                                FALSE)) then
32                                 expansionMask1(xMeasure,yMeasure) = TRUE
33                             else
34                                 done1 = TRUE
35                             if (pixelDifference2 ≤ 0) OR ((expandSteps < minSteps) AND (done2 ==
                                FALSE)) then
36                                 expansionMask2(xMeasure,yMeasure) = TRUE
37                             else
38                                 done2 = TRUE
39                         if (done1 == TRUE) AND (done2 == TRUE)) OR (expandSteps > maxSteps)
                            then
40                             break
41                         expandLength = extendLength+1
42                     thetaMultiple += 1
```

Bibliography

[1] Michael J. Gerdes, Christopher J. Sevinsky, Anup Sood, Sudeshna Adak, Musodiq Bello, Alexander Bordwell, Ali Can, Corwin Alex D, Sean Richard Dinn, Robert John Filkins, Denise Hollman, Vidya Pundalik Kamath, Sireesha Kaanumalle, Kevin Bernard Kenny, Melinda Larsen, Michael Lazare, Qing Li, Christina Lowes, Colin Craig McCulloch, Elizabeth McDonough, Michael Christopher Montalto, Zhengyu Pang, Jens Rittscher, Alberto Santamaria-Pang, Brion Daryl Sarachan, Maximilian Lewis Seel, Antti Seppo, Kashan Shaikh, Yunxia Sui, Jingyu Zhang, and Fiona Ginty. Highly multiplexed single-cell analysis of formalin-fixed, paraffin-embedded cancer tissue. *Proceedings of the National Academy of Sciences of the United States of America*, 110(29):11982–11987, 2013. doi: 10.1073/pnas.1300136110. URL https://doi.org/10.1073/pnas.1300136110.

[2] Yury Goltsev, Nikolay Samusik, Julia Kennedy-Darling, Salil S. Bhate, Matthew Hale, Gustavo Vazquez, Sarah Black, and Garry P. Nolan. Deep profiling of mouse splenic architecture with codex multiplexed imaging. *Cell*, 174(4):968–981, 2018. doi: 10.1016/j.cell.2018.07.010. URL https://doi.org/10.1016/j.cell.2018.07.010.

[3] Gabriele Gut, Markus D. Herrmann, and Lucas Pelkmans. Multiplexed protein maps link subcellular organization to cellular states. *Science (New York, N.Y.)*, 361(6401):NA–NA, 2018. doi: 10.1126/science.aar7042. URL https://doi.org/10.1126/science.aar7042.

[4] Jia-Ren Lin, Mohammad Fallahi-Sichani, and Peter K. Sorger. Highly multiplexed imaging of single cells using a high-throughput cyclic immunofluorescence method. *Nature communications*, 6(1):8390–8390, 2015. doi: 10.1038/ncomms9390. URL https://doi.org/10.1038/ncomms9390.

[5] Eliot T. McKinley, Yunxia Sui, Yousef Al-Kofahi, Bryan A. Millis, Matthew J. Tyska, Joseph T. Roland, Alberto Santamaria-Pang, Christina L. Ohland, Christian Jobin, Jeffrey L. Franklin, Ken S. Lau, Michael J. Gerdes, and Robert J. Coffey. Optimized multiplex immunofluorescence single-cell analysis reveals tuft cell heterogeneity. *JCI insight*, 2(11):NA–NA, 2017. doi: 10.1172/jci.insight.93487. URL https://doi.org/10.1172/jci.insight.93487.

[6] Jia-Ren Lin, Benjamin Izar, Shu Wang, Clarence Yapp, Shaolin Mei, Parin Shah, Sandro Santagata, and Peter K. Sorger. Highly multiplexed immunofluorescence imaging of human tissues and tumors using t-cycif and conventional optical microscopes. *eLife*, 7(NA):NA–NA, 2018. doi: 10.7554/elife.31657. URL https://doi.org/10.7554/elife.31657.

[7] Pavel Zrazhevskiy and Xiaohu Gao. Quantum dot imaging platform for single-cell molecular profiling. *Nature communications*, 4(1):1619–1619, 2013. doi: 10.1038/ncomms2635. URL https://doi.org/10.1038/ncomms2635.

[8] Werner Baschong, Rosmarie Suetterlin, and R. Hubert Laeng. Control of autofluorescence of archival formaldehyde-fixed, paraffin-embedded tissue in confocal laser scanning microscopy (clsm). *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society*, 49(12):1565–1571, 2001. doi: 10.1177/002215540104901210. URL https://doi.org/10.1177/002215540104901210.

[9] P. Del Castillo, M. L. Molero, J. M. Ferrer, and Juan C. Stockert. Autofluorescence and induced fluorescence in epon embedded tissue sections. *Histochemistry*, 85(5):439–440, 1986. doi: 10.1007/bf00982676. URL https://doi.org/10.1007/bf00982676.

[10] Anna Cleta Croce and Giovanni Bottiroli. Autofluorescence spectroscopy and imaging: a tool for biomedical research and diagnosis. *European journal of histochemistry : EJH*, 58(4):2461–2461, 2014. doi: 10.4081/ejh.2014.2461. URL https://doi.org/10.4081/ejh.2014.2461.

[11] Jonas Wizenty, Muhammad Imtiaz Ashraf, Nadine Rohwer, Martin Stockmann, Sascha Weiss, Matthias Biebl, Johann Pratschke, Felix Aigner, and Tilo Wuensch. Autofluorescence: A potential pitfall in immunofluorescence-based inflammation grading. *Journal of immunological methods*, 456(NA):28–37, 2018. doi: 10.1016/j.jim.2018.02.007. URL https://doi.org/10.1016/j.jim.2018.02.007.

[12] Nicholas Billinton and Andrew W. Knight. Seeing the wood through the trees: a review of techniques for distinguishing green fluorescent protein from endogenous autofluorescence. *Analytical biochemistry*, 291(2):175–197, 2001. doi: 10.1006/abio.2000.5006. URL https://doi.org/10.1006/abio.2000.5006.

[13] A. Sally Davis, Anke Richter, Steven Becker, Jenna E. Moyer, Aline Sandouk, Jeff Skinner, and Jeffery K. Taubenberger. Characterizing and diminishing autofluorescence in formalin-fixed paraffin-embedded human respiratory tissue. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society*, 62(6):405–423, 2014. doi: 10.1369/0022155414531549. URL https://doi.org/10.1369/0022155414531549.

[14] Michael Neumann and Detlef Gabel. Simple method for reduction of autofluorescence in fluorescence microscopy. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society*, 50(3):437–439, 2002. doi: 10.1177/002215540205000315. URL https://doi.org/10.1177/002215540205000315.

[15] Mary E. Dickinson, Gregory H. Bearman, S Tille, Rusty Lansford, and Scott E. Fraser. Multi-spectral imaging and linear unmixing add a whole new dimension to laser scanning fluorescence microscopy. *BioTechniques*, 31(6):1272–1278, 2001. doi: 10.2144/01316bt01. URL https://doi.org/10.2144/01316bt01.

[16] Zhengyu Pang, Eugene Barash, Alberto Santamaria-Pang, Christopher J. Sevinsky, Qing Li, and Fiona Ginty. Autofluorescence removal using a customized filter set. *Microscopy research and technique*, 76(10):1007–1015, 2013. doi: 10.1002/jemt.22261. URL https://doi.org/10.1002/jemt.22261.

[17] Franco Woolfe, Michael J. Gerdes, Musodiq Bello, Xiaodong Tao, and Ali Can. Autofluorescence removal by non-negative matrix factorization. *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, 20(4):1085–1093, 2010. doi: 10.1109/tip.2010.2079810. URL https://doi.org/10.1109/tip.2010.2079810.

[18] Min Kim, Naomi R. Truong, Virginia James, Lidija Bosnjak, Kerrie J Sandgren, Andrew N. Harman, Najla Nasr, Kirstie M. Bertram, Norman Olbourne, Shailandra

Sawleshwarkar, Kaylene McKinnon, Ralph C. Cohen, and Anthony L. Cunningham. Relay of herpes simplex virus between langerhans cells and dermal dendritic cells in human skin. *PLoS pathogens*, 11(4):e1004812–NA, 2015. doi: 10.1371/journal.ppat.1004812. URL https://doi.org/10.1371/journal.ppat.1004812.

[19] Andrew N. Harman, Joey Lai, Stuart Turville, Shamith A. Samarajiwa, Lachlan Robert Gray, Valerie Marsden, Sarah K. Mercier, Kate L. Jones, Najla Nasr, Arjun Rustagi, Helen E. Cumming, Heather Donaghy, Johnson Mak, Michael Gale, Melissa J Churchill, Paul J. Hertzog, and Anthony L. Cunningham. Hiv infection of dendritic cells subverts the ifn induction pathway via irf-1 and inhibits type 1 ifn production. *Blood*, 118(2):298–308, 2011. doi: 10.1182/blood-2010-07-297721. URL https://doi.org/10.1182/blood-2010-07-297721.

[20] Nico Battich, Thomas Stoeger, and Lucas Pelkmans. Control of transcript variability in single mammalian cells. *Cell*, 163(7):1596–1610, 2015. doi: 10.1016/j.cell.2015.11.018. URL https://doi.org/10.1016/j.cell.2015.11.018.

[21] Li Wu and Vineet N. KewalRamani. Dendritic-cell interactions with hiv: infection and viral dissemination. *Nature reviews. Immunology*, 6(11):859–868, 2006. doi: 10.1038/nri1960. URL https://doi.org/10.1038/nri1960.

Chapter 3

spicyR: spatial analysis of in situ cytometry data in R

Publications incorporated into this chapter:

Canete, N.P., Iyengar, S.S., Ormerod, J.T., Baharlou, H., Harman, A.N., Patrick, E. (2022) spicyR: spatial analysis of in situ cytometry data in R. *Bioinformatics, Volume 38, Issue 11, 1 June 2022, Pages 3099–3105*

Preface

One of the key steps in microscopy image analysis is spatial analysis, providing insight into the cell-cell interactions and cell migration events that underly the pathophysiology of disease. This chapter consists of a second publication in Bioinformatics, which describes the R package 'SPatial analysis of In Situ Cytometry data in R' (spicyR), a tool which facilitates inference on changes in spatial co-localization between cell types. spicyR was conceived to address the question "How does cell-type co-localization change with different experimental or clinical conditions?". This publication presents an easy-to-use tool which adds to the repertoire of spatial analysis techniques utilized in the analysis of high parameter images. To validate the algorithm, spicyR was applied to a range of simulations, as well as the type 1 diabetes dataset presented by Damond et al. (2019). As first author, I contributed to all aspects of the manuscript. I conceived the original approach for spicyR, wrote the package codebase, and performed the validation of the algorithm.

Abstract

Motivation High parameter histological techniques have allowed for the identification of a variety of distinct cell types within an image, providing a comprehensive overview of the tissue environment. This allows the complex cellular architecture and environment of diseased tissue to be explored. While spatial analysis techniques have revealed how cell–cell interactions are important within the disease pathology, there remains a gap in exploring changes in these interactions within the disease process. Specifically, there are currently few established methods for performing inference on cell-type co-localization changes across images, hindering an understanding of how cellular environments change with a disease pathology.

Results We have developed the spicyR R package to perform inference on changes in the spatial co-localization of types across groups of images. Application to simulated data demonstrates a high sensitivity and specificity. We demonstrate the utility of spicyR by applying it to a type 1 diabetes imaging mass cytometry dataset, revealing changes in cellular associations that were relevant to the disease progression. Ultimately, spicyR allows changes in cellular environments to be explored under different pathologies or disease states.

Availability and implementation R package is freely available at
http://bioconductor.org/packages/release/bioc/html/spicyR.html
and shiny app implementation at
http://shiny.maths.usyd.edu.au/spicyR/.

3.1 Introduction

Identifying changes in the spatial distribution of cells is vital for understanding the cellular processes that are present in diseased tissue. Multiplexed histological techniques have allowed the complex cellular architecture and environment of diseased tissue to be explored by enabling the simultaneous profiling of multiple cell types. Fluorescence-based methods, including co-detection by imaging (CODEX) [1], cyclic immunofluorescence [2] and iterative indirect immunofluorescence imaging (4i) [3], as well as mass cytometry imaging techniques, including imaging mass cytometry (IMC) [4] and multiplexed ion beam imaging by time of flight [5] allow up to 40 protein markers to be visualized with single-cell resolution. Additionally, the development of spatial-based transcriptomic techniques, such as High-Definition Spatial Transcriptomics [6] and sequential fluorescence *in situ* hybridization [7], allow tens of thousands of transcripts to be spatially resolved within an image at single-cell resolution. This large increase in the scale and dimensionality of the images being acquired has necessitated the development of analysis techniques capable of interrogating such complex data.

Established image analysis techniques have enabled the investigation of cell–cell interactions and cell migration within an image, facilitating the interrogation of high parameter imaging data in a single-cell manner. Standard pipelines start by identifying cells through single-cell segmentation followed by cell-type classification by clustering or manually gating marker expression [8–11]. From here, a differential analysis of cell composition or marker expression within the image dataset can be performed to identify high-level associations with a phenotype of interest. The spatial dimension afforded by imaging can furthermore allow the spatial context of these cells to be quantified in multiple ways. One way to interrogate the spatial organization of cells is to quantify the spatial attraction or avoidance between pairs of cell types. This often involves counting an association measure between cell types. Counting the number of touching cells of a pair of cell types provides a measure of spatial association [12]. Randomizing the labels of cells can then be used to identify how significant a pairwise interaction is within an image, as seen in an application to type 1 diabetes images [12]. This approach can be extended by calculating distances between cells and tabulating their nearest neighbours as used by Keren et al. to assess the spatial interactions involved in triple-negative breast cancer pathology [13]. Ripley's K-function [14, 15] can be used to assess how co-localization between cell types vary with distance by modelling cells within an image as a point process. Further to this, identification of spatial communities [16] has been performed through graph-based techniques, associating the spatial distribution of groups of cell types to disease outcomes. Finally, techniques, such as spatial variance component analysis [17], have been developed to identify the sources of variation of gene or protein markers attributed to cell–cell interactions. Overall, such techniques can identify spatial structure within high parameter images. This could then be attributed to the pathology of the disease states being studied.

A key gap that remains in the analysis of high parameter images is the identification of differential cell-type co-localization across groups — i.e. changes in the extent of pairwise cell-type co-localization across these groups. These groups could be a clinical outcome, such as disease stage or response to treatment or come from a perturbed experiment. Present strategies involve comparing association measures across groups. Damond et al. (2019) [18] compares the number of each pairwise interaction using a Mann–Whitney's U-test to show changes in cell-type co-localizations across different disease stages. Färkkilä et al. (2020) [19] compares the Z-scores of the pairwise cell interaction obtained from

bootstrapping across ovarian cancer clinical outcomes to identify changes in cell distribution. While such approaches are appropriate, they do not allow for the modelling of multiple images from multiple subjects. Additionally, these strategies ignore the variation observed in the quantification of cell-type co-localization.

In this manuscript, we present an R package 'SPatial analysis of In situ CYtometry data in R' (spicyR) and a corresponding web application to facilitate inference on changes in spatial co-localizations between cell types. spicyR aims to provide an easy-to-use approach to identify differential cell-type co-localization with respect to a disease or treatment and has the capacity to model information from multiple images per subject or account for images having a substantial difference in cell number. We demonstrate the performance of spicyR through simulation and apply the package to the diabetes imaging data presented in Damond et al. (2019) [18], revealing changes in cellular co-localization with type 1 diabetes progression.

3.2 Methods

Our R package, spicyR, provides the framework for performing inference on the changes in spatial co-localizations between pairs of cell types, which can be associated with a discrete or continuous outcome (Figure 3.1A). As described below, spicyR consists of three primary steps: (i) summarizing the degree of spatial co-localization between pairs of cell types for each image; (ii) modelling the variability in the co-localization summary statistics as a function of cell counts and (iii) testing for changes in co-localization associated with a response variable. The significance of this change is assessed using a linear model, or a mixed-effects model if there are multiple images belonging to a subject (Figure 3.1B). The R package is available on Bioconductor (`http://bioconductor.org/packages/release/bioc/html/spicyR.html`), and is also implemented as an interactive shiny application (`http://shiny.maths.usyd.edu.au/spicyR/`).

3.2.1 Construction of the L curve and co-localization score

Following single-cell segmentation and classification, images are modelled as a 'marked point process model', in which each cell is represented as a point in a 2D plane. Spatial co-localization between two cell types within an image can be quantified with a K-function (Ripley, 1976),

$$\widehat{K_{ij}}(r) = \frac{|W|}{n_i n_j} \sum_{n_i} \sum_{n_j} 1\{d_{ij} \leq r\} e_{ij}(r) \tag{3.1}$$

where $\widehat{K_{ij}}(r)$ summarized the degree of co-localization of cell type j with cell type i, n_i and n_j are the number of cells of type i and j, $|W|$ is the image area, d_{ij} is the distance between two cells and $e_{ij}(r)$ is an edge correcting factor. The K-function can be interpreted as the average number of cells of type j within a distance r away from each cell of type i. [1]

The L-function or L curve is a variance stabilized K-function given by the equation

$$\widehat{L_{ij}}(r) = \sqrt{\frac{\widehat{K_{ij}}(r)}{r}} \tag{3.2}$$

[1] It should be noted that the K-function is not symmetric - that is $\widehat{K_{ij}}(r) \neq \widehat{K_{ji}}(r)$.

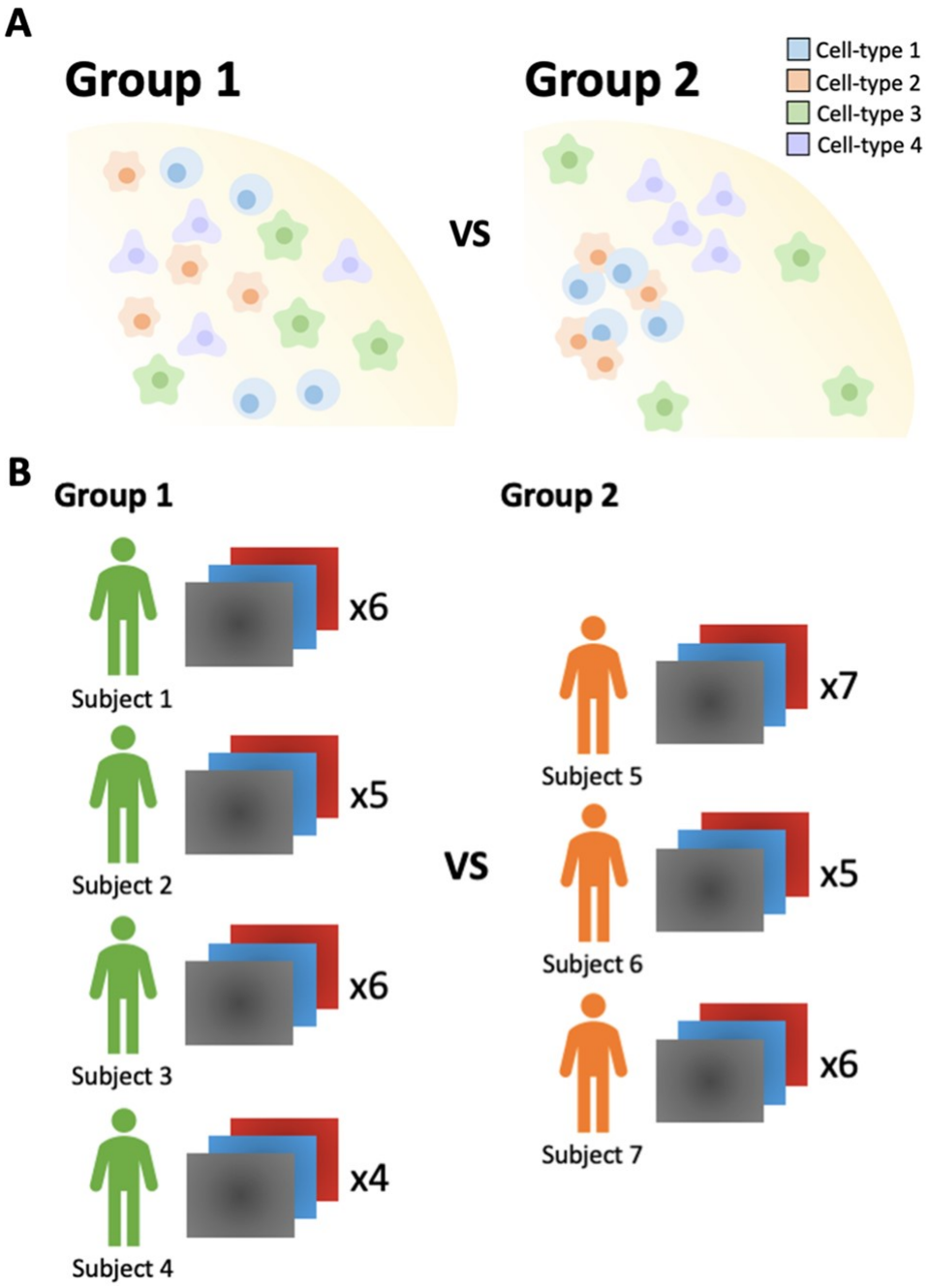


Figure 3.1: **Schematic of the experimental motivation of spicyR. (A)** Analytical question. From the images obtained, spicyR aims to identify differences in cell-type co-localizations between the two groups. **(B)** Example experimental setup. Here, we have four subjects from Group 1, and three donors from Group 2, each with a different number of high parameter images. When applying spicyR, the number of subjects per group does not have to be equal, the number of images per subject does not have to be equal, there can be one image for subject. Further to this, the number of groups can be larger than two or alternatively continuous outcome

[15]. Figure 3.2B shows an example observed L curve compared to the expected L curve obtained from points distributed randomly in a Poisson distribution. Curves above the Poisson line can be interpreted as showing greater attraction when compared to a random distribution, and curves below the Poisson line can be interpreted as avoidance of the two cell types.

To reduce these summary functions into a single co-localization score u, we take an area between curve measurement (Figure 3.2B), given by the equation

$$u = \sum_{r=r_{min}}^{r_{max}} \widehat{L_{ij.Experimental}}(r) - \widehat{L_{ij.Poission}}(r) \tag{3.3}$$

where the sum is taken over a discrete range of r between r_{min} and r_{max} (e.g. $r' = 10, 20, \ldots, r_{max}$). A special case of Equation 3.3 that is explored below is when there is only one value to sum over (i.e. $r_{min} = r_{max}$). Values of $u > 0$ represent greater attraction between cells when compared to random, while values < 0 represent avoidance when compared to random.

3.2.2 Weighting scheme used to account for varying cell counts

The relationship between cell count and variability of u is modelled with a shape constrained generalized additive model (GAM) fitted with SCAM [21]. We fit a GAM with a monotone decreasing constraint to the square of u with the counts of both cell types i and j as explanatory variables. The inverse of this fitted curve is used as a weighting scheme for each measurement in the model described in the following section — i.e. lower weights are applied to images with lower cell counts. These weights can be constructed per cell-type pair or using the scores from all cell types. In all the analyses in this manuscript the weights are constructed using all cell types concurrently.

3.2.3 Hypothesis testing with a linear model

To assess changes in cell-type co-localization between groups, we implement either a linear model, or a linear mixed-effects model with random intercepts using the lme4 package [22]. Here, u_{ij} is the co-localization score for subject i and image j, the treatment group or condition (x_i) is modelled as a fixed effect with coefficient β and if a subject has multiple images the intercept (α_i) is modelled as a random effect. Other covariates are included as W with coefficients Γ.

Linear model:

$$u_i = \alpha + \beta x_i + \Gamma W \tag{3.4}$$

Linear mixed-effects model:

$$u_{ij} = \alpha_i + \beta x_i + \Gamma W \tag{3.5}$$

For the mixed-effects model, P-values are calculated with Satterthwaite's approximations using the lmertest package [23].

3.2.4 Simulations to assess the performance of spicyR

To examine the performance of spicyR, simulations were generated using the spatstat package [24], as summarized below and in Supplementary Figure 3.2. There are a few key parameters in the simulation with their default values listed as follows.

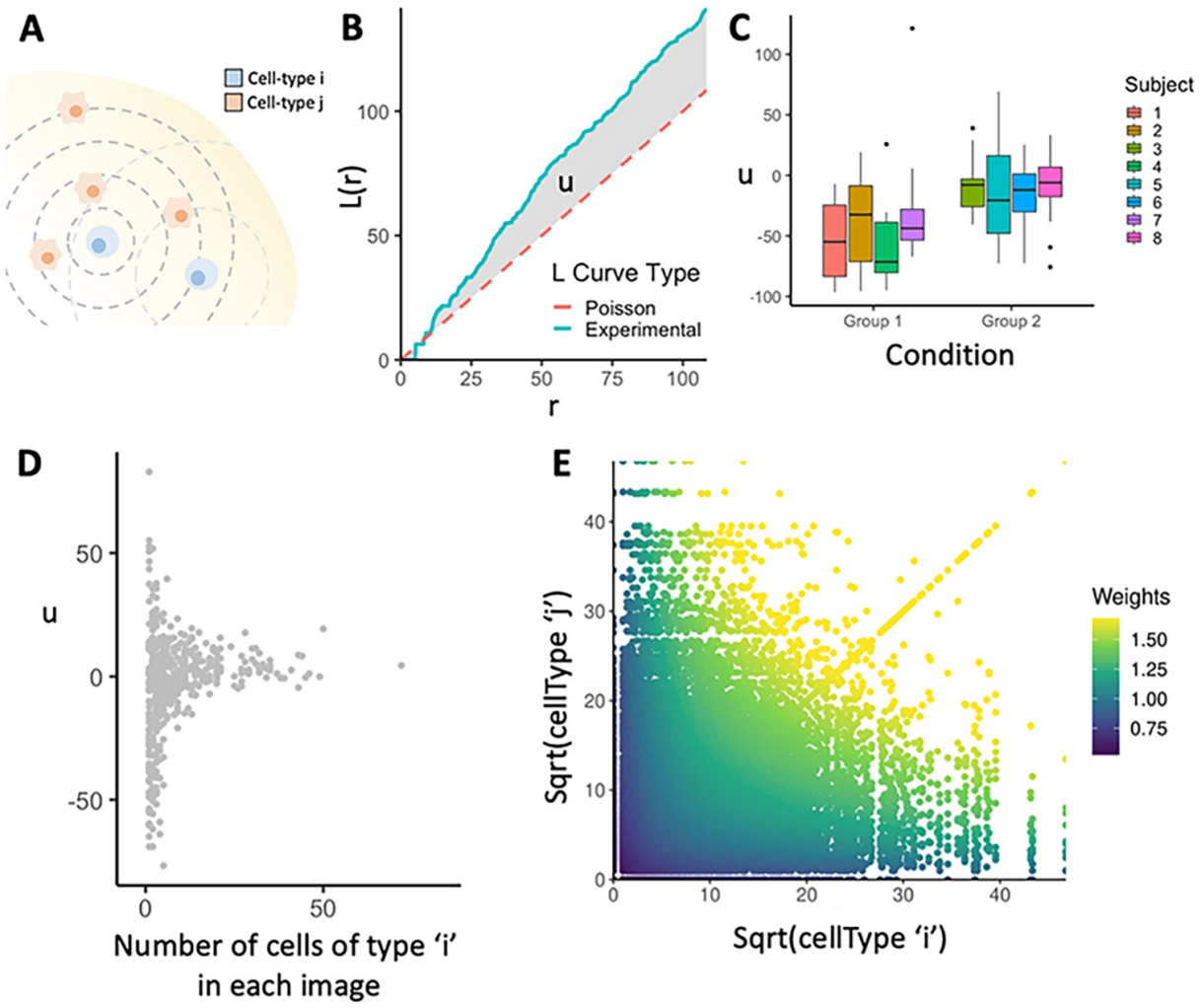


Figure 3.2: **Summary of the spicyR framework.** **(A)** Representation of two cell types in a 2D plane, which can be represented as points. Co-localization of cell type j around cell type i can be modelled by using Ripley's K-function, which summarizes a scaled average of the number of cell type j around cell type i for a range of radii, r (dashed lines). **(B)** Example of an L curve generated from an experiment (solid line) compared against a L curve (dashed line) expected if cell type i and j were distributed independent of each other. The shaded area represents the u statistic used to assess whether co-localization (positive value) or avoidance (negative value) is occurring. **(C)** A boxplot representation of the u statistics used to compare pairwise co-localization across different groups. **(D)** Plot of the u statistic versus the number of cells of type i. As cell count is decreased, the variance of the AUC statistic increases. A GAM is used to model the square of the statistic as a 2D function of the counts of cell types 'i' and 'j', and these are used as weights in the linear models used. **(E)** Plot representing the weighting regime when applied to the dataset from Schürch et al. (2020) [20]. The value of the weights increases as the pairwise cell counts increases

- The size of each image: 1000×1000 units.
- The number of subjects: 40, split into 2 groups of 20 each.
- The number of images per subject: three, each containing two cell types.
- The number of simulations: 500 without change (*null* simulation) and 500 with change in co-localization between groups (*difference* simulations).
- A range for the number of cells of each type per image: $cell\,count = 20, 40, \ldots, 380, 400$.
- Average representative distance of co-localization between cell type A and B: $sigma = 40$.
- Change in co-localization between cell types between groups: $delta = sigma/3$.

For each simulation, cells in each image were simulated by first generating cells from cell type A using a Poisson point process model with expected count equal to a random value from *cellCount*. The density of cell type A is then calculated using a disc kernel, where the size of the disc for each subject is drawn from a Poisson with expected value $sigma$ or $sigma + delta$ depending on which group that subject belongs to. In the *difference* simulations the average co-localization is equal to $sigma$ in the first group and $sigma + delta$ in the second group. In the *null* simulations the average co-localization in the first group of subjects is the same as that in the second group of subjects. The cells from cell type B are then generated using a Poisson point process model with expected count equal to another random value from *cellCount* multiplied by the density of cell type A. When applying spicyR to the simulations, we apply Equation 3.3 across a range of radii ($r' = 10, 20, \ldots, 100$) to obtain the co-localization statistic.

To evaluate the performance of spicyR, the P-value distributions from the tests on the *null* simulations were compared to the *difference* simulations. The percentage of P-values from the *null* simulations less than a significant threshold will provide a false positive rate and the percentage of P-values from the *difference* simulations less than a significance threshold will provide a true positive rate.

The simulations were first applied to identify if the use of the weighting regime as described above improved the sensitivity of spicyR. Next, simulations were performed to explore the sensitivity of spicyR to different average co-localization distances. This was achieved by repeating the above simulations, while also varying $sigma$ from 10 up to 100. spicyR is then applied by applying Equation 3.3 across a range of radii ($r' = 10, 20, \ldots, 100$), or by taking the difference between discrete L curve values for each r from 10 up to 100. Finally, the simulations were used to evaluate how spicyR performs for different cell counts with *cellCount* multiplied by a factor of 1, 5, 10 and 20.

3.2.5 Application to diabetes IMC data in Damond et al. 2019

As an implemented example of the framework, we apply spicyR to the data presented by Damond et al [18]. This study aimed to identify the spatial distribution of markers and cells in pancreatic islets, comparing three different stages of type 1 diabetes mellitus (T1DM) progression: non-diabetic, onset diabetes and long-duration diabetes, with four subjects per group. The single-cell image data were downloaded from Mendeley

(Version 2: https://data.mendeley.com/datasets/cydmwsfztj/2). This dataset consists of images of pancreatic islet cells across the three disease stages, obtained via IMC with 37 markers. The number of images per subject ranged from 64 to 81 (Total=845; Non-diabetic=274; Onset=290; Long-duration=281). Additionally, the dataset contained spreadsheets providing relevant patient information, as well as the cell coordinates, marker expression and cell classification.

Here, we aimed to apply spicyR to identify significant differential co-localizations across the different disease stages. The X and Y coordinates of each cell, which image the cell belonged to, and the cell type were obtained, alongside the patient ID and diabetes disease stage for each image. spicyR was then applied to the data, with patient ID being treated as the random effect and the disease stage (non-diabetic, onset and long-duration diabetes) being treated as the fixed effect. The co-localization statistic was obtained by applying Equation 3.3 across a range of radii ($r' = 10, 20, \ldots, 100$). The cell types studied here are the endocrine cell subsets (alpha, beta, gamma and delta) and the immune cell subsets [T helper (Th), T cytotoxic (Tc), neutrophils and macrophages].

3.2.6 Application to other datasets

The IMC dataset from Damond et al. has many images per patient and a categorical outcome with three categories. To demonstrate that spicyR can be applied to datasets with one image per patient, fit models which include covariates and binary outcomes and demonstrate variation in computational performance we applied spicyR to a subset of a dataset consisting of breast cancer patients assayed with IMC [16] available in the imcdatasets R package, and colorectal cancer patients assayed with CODEX [20].

3.3 Results

Here, we present spicyR, a framework for identifying changes in spatial association between pairs of cell types that could be associated with images from different clinical or experimental groups (Figure 3.1). As input, spicyR requires images that have undergone single-cell segmentation and cell-type classification. In addition to calculating measures of co-localization between pairs of cell types (Figure 3.2A-C), we provide functionality for empirically estimating the variability of the spatial associations between cell types within an image and for including multiple images per subject in the models.

We observed across multiple datasets (Figure 3.2D and Supplementary Figure 3.1) that the variability of the measure of spatial association between a pair of cell types, u, decrease as the number of the cells in an image increase. Quantifying the relationship between the variability of u and the number of cells in an image, provides an opportunity to propagate this information in the model fitting process. We model the relationship between u and cell count (Figure 3.2D) by fitting a shape constrained GAM to the square of u as a function of the counts in both cell types. The inverse of the values obtained from this fitted curve are then used as weights (Figure 3.2E) when testing for association between the co-localization of a pair of cell types and an outcome, with lower weights being applied to images with lower cell counts for a given cell type.

Simulated images were generated using the spatstat package to assess the performance of spicyR. First, we examined the benefits of including image weights in spicyR where we observed a distinct difference in the area under the receiver operating characteristic curves (AUC) with tests that included weights performing better than the non-weighted

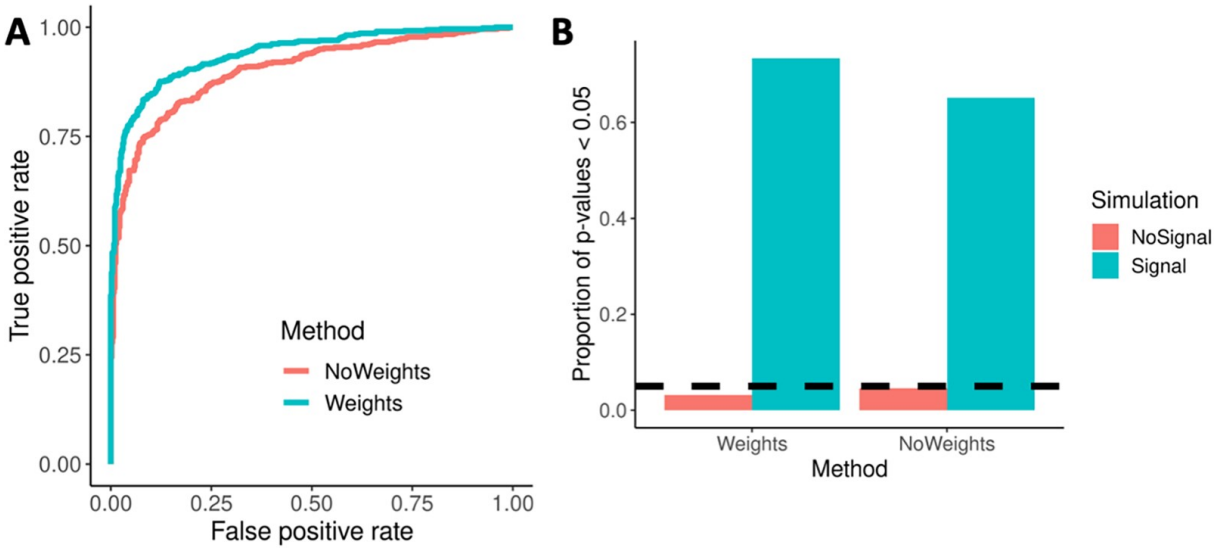


Figure 3.3: **Simulations demonstrating the performance of the spicyR framework. (A)** ROC curves obtained from simulations (AUC: no weights = 0.901; weights = 0.940). **(B)** Bar plots of the percentage of null simulations (false positives) and simulations with changes in co-localization (true positives) that were significant at a $P = 0.05$ level when spicyR was applied

tests (AUC values: no weights=0.901; weights=0.940) (Figure 3.3A-B). Next, we explore the role of cell counts on the performance of spicyR where the performance benefit from using weights was retained as the average number of cells in an image increased (Supplementary Figure 3.3A-B). We further assessed how cell count influences computation time (Supplementary Figure 3.3C) with a simulated image with 1000 cells taking 1s to calculate pairwise co-localizations. To complement this, we have included the computational times of three publicly available datasets in Supplementary Table 3.1. Finally, we examined the sensitivity of spicyR to the choice of radii for calculating the co-localization score. In all simulations, the performance of spicyR was superior when the choice of radius used to quantify co-localization was close to the distance, which was used to generate the cell-type relationships (Supplementary Figure 3.4A). However, running spicyR with a range of radii had the highest performance on average across all simulations (Supplementary Figure 3.4B). This indicates that when there is uncertainty around the distance which changes in co-localization are occurring, as in most cases, choosing a range of radii might be an effective strategy for detecting those changes. When the expected co-localization distance is known, the use of a single radius may be sufficient. Nevertheless, in most circumstances, it is recommended to use multiple radii to define the co-localization statistic.

Damond et al. aimed to identify spatial changes in marker and cell distribution in pancreatic islets of three T1DM disease groups: non-diabetic, onset diabetes and long-duration diabetes. A key finding of the study was that there was a temporal correlation between beta cell destruction, marked by marker loss and cell decreases, and an increased infiltration of Th and Tc cells in beta cell-rich pancreatic islets. Hence, Th and Tc cells were implicated in the destruction of beta cells characteristic of T1DM. We aimed to explore these results further, using spicyR to identify differential cell-type co-localizations between the non-diabetic group compared to the onset diabetes group. The cell types included were the endocrine cell subsets (alpha, beta, gamma and delta) and the immune cell subsets (Th, Tc, neutrophils and macrophages).

First, we applied spicyR by modelling cell-type co-localization with a simple linear model, which ignores patient information. Concordant with the findings in Damond et al., we observed that Th and Tc cells showed increased spatial co-localization with islet cells in the diseased groups, specifically towards beta cells (Figure 3.4A-B). The T cell co-localizations were more significant with beta cells (from beta to Tc: $P = 9.49 \times 10^{-7}$; from beta to Th: $P = 1.06 \times 10^{-6}$) compared to alpha cells (alpha to Tc: $P = 1.91 \times 10^{-3}$; alpha to Th: $P = 1.55 \times 10^{-4}$). This suggests a migration of Th and Tc cells towards beta cells (Figure 3.4B), specifically during the early disease stages, reiterating the results presented by Damond et al. It was also found that there was increased co-localization of Th cells in onset diabetics (from Th to Th: $P = 2.15 \times 10^{-13}$).

Next, we repeated this analysis, implementing a linear mixed-effects model, which explicitly modelled islets as coming from different individuals. In contrast to the previous model, immune interactions with beta cells did not appear to be significant (from beta to Tc: $P = 0.121$ from beta to Th: $P = 0.142$). This difference in result is attributed to the strong increase in co-localization only being present in two of the four onset diabetes patients (Supplementary Figure 3.5A-B). However, using the mixed-effects model there was a decrease in co-localization between delta cells and beta cells (from beta to delta: $P = 0.008$, Supplementary Table 2) suggesting a pattern to the way that beta cells degrade in the islet. The differences between the linear model and the linear mixed-effects model are highlighted in Figure 3.4D.

3.4 Discussion

Here, we have presented spicyR, a tool for identifying differential cell-type co-localizations across different groups. We have demonstrated its performance to identify cell co-localization changes through both simulated images, and with application to a type 1 diabetes IMC dataset. Simulations revealed that including weights that quantified the relationship between the number of cells in an image and the variability of a co-localization statistic increased the sensitivity and specificity of spicyR. Furthermore, when spicyR was applied to the diabetes dataset, the original results were reaffirmed and other key cell interactions present in diabetes progression were highlighted.

The spicyR package has many advantages compared to other differential co-localization strategies used in high parameter image analysis. The key advantage of the package is its ability to summarize changes in cell-type co-localizations across groups of images. If multiple images are obtained from multiple subjects, the mixed-effects model implemented allows variations within each subject to be modelled. By taking values of the L curve at multiple radii, we obtain a co-localization statistic that is easily interpretable and comparable across images. Furthermore, we implement a weighting scheme to account for variation in the co-localization statistic given the pairwise cell-type counts, which improves the predictive capabilities of spicyR. Finally, the package allows differential cell-type co-localizations to be identified across all pairwise cell types within the dataset, summarized in an interpretable heatmap. In this way, spicyR provides the framework for highlighting key cell–cell interactions that change across groups within a high parameter imaging dataset.

Despite the advantages discussed, there are several limitations to the implementation and evaluation of spicyR. Firstly, to benefit from the weighting regime, a moderate number of images are required to better model the relationship between cell count and co-localization score. Depending on the experimental approach, and the availability of bi-

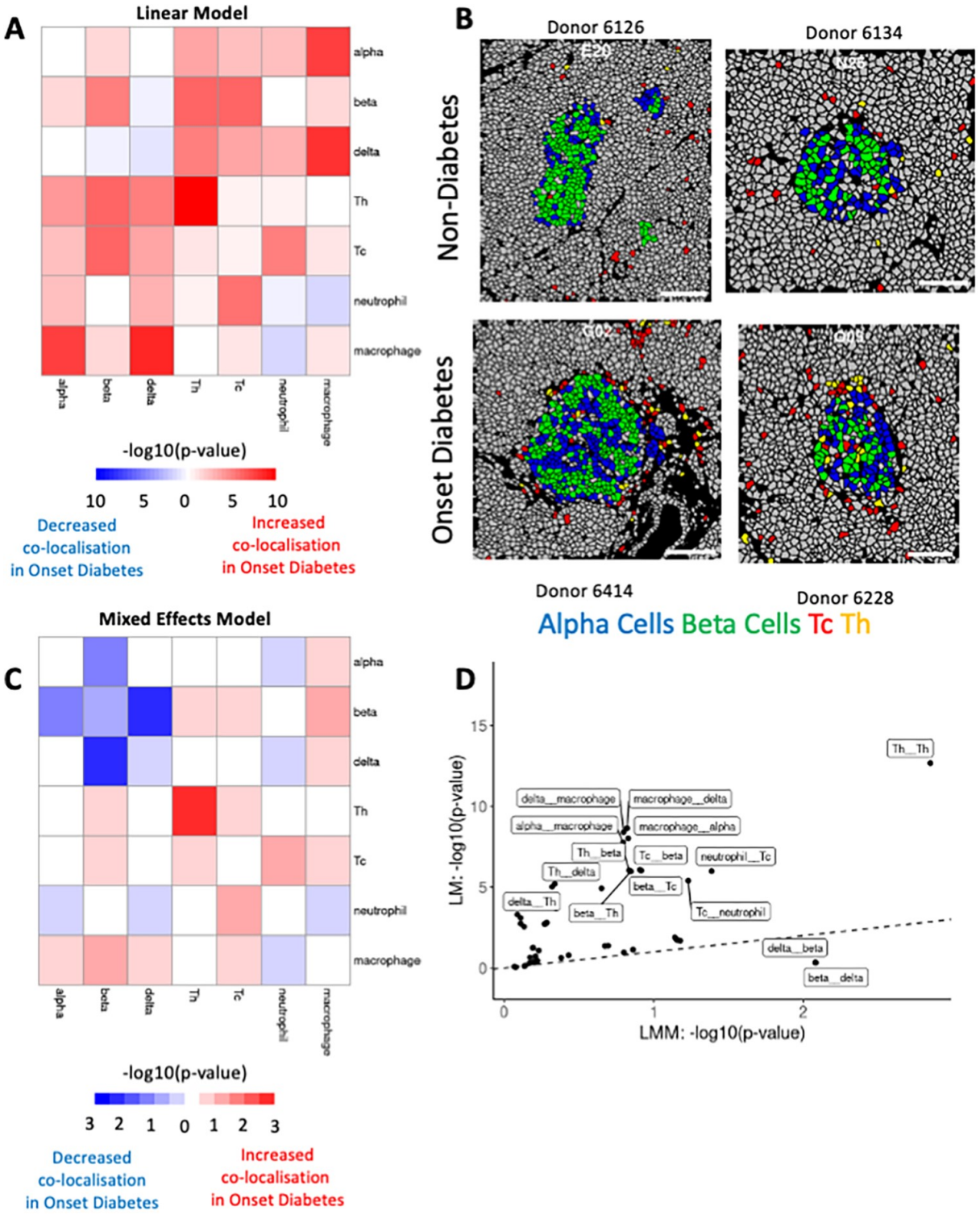


Figure 3.4: **Application of spicyR to the Damond et al. (2019) [18] type 1 diabetes IMC dataset.** (continued on next page)

Figure 3.4: The spicyR framework was used to compare changes in cell-type co-localizations between the non-diabetic group and the onset diabetes group. **(A)** Heatmap showing the log10 P-value from applying spicyR using a linear model. The y-axis represents cell type i (the 'from' cell type), and the x-axis represents cell type j (the 'to' cell type). Positive values represent increased co-localization in the onset diabetes group; negative values represent decreased co-localization in the onset diabetic group. **(B)** Masks showing representative images from the non-diabetes group and the onset diabetes group. Four cell types are highlighted: Alpha cells, Beta cells, Tc cells and Th cells. **(C)** Heatmap showing the log10 P-value from applying spicyR using a mixed-effects model. The y-axis represents cell type i (the 'from' cell type), and the x-axis represents cell type j (the 'to' cell type). Positive values (red) represent increased co-localization in the onset diabetes group; negative values (blue) represent decreased co-localization in the onset diabetic group. **(D)** Scatterplot showing log10 P-values from the application of spicyR with a linear model ignoring patient information and a linear mixed model including patient information

ological samples, this may not be feasible. To address this, we recommend calculating weights using information from all pairwise comparisons. Secondly, if there is confounding between cell number and the degree of change in co-localization between cell types, the weighting strategy might obscure these differences. Thirdly, there are trade-offs to using small versus large radii when applying spicyR. Large radii may provide a better overall summary of cell-type co-localization within an image, but can have decreased sensitivity, particularly if the co-localization occurs strongly only over short distances. Hence, users should choose an appropriate distance or range of distances based on the biological questions being studied. Finally, while computational simulations were performed, it is difficult to validate spicyR within biological scenarios. There may be further scope for producing biologically relevant ground truths with which spicyR can be tested against, as well as more complicated simulation studies to elicit the effectiveness of the package.

It is important to acknowledge that spicyR should not be used in isolation to other analysis techniques, with the package being a key step in a broader analysis pipeline for high parameter imaging data. Firstly, spicyR is contingent on single-cell segmentation and cell-type classification (unsupervised or manually gated) being sufficiently accurate to elicit biologically relevant results. It is also complementary to other tests, such as testing for changes in cell-type composition and cell marker expression [8], which are necessary for providing an overview of the biological process being studied. Furthermore, it is advised to explore images visually both before and after any spatial analysis to identify whether visual observations appear to be consistent with the results of spicyR. Tools, such as histoCAT [12] and the Bioconductor package cytomapper [25], provide useful exploratory tools for facilitating such exploratory data analysis. Ultimately, the package serves the role of highlighting changes in cell-type co-localization. This information can be crucial for synthesizing key biological insight from multiplexed imaging experiments. Overall, results from spicyR will complement observations obtained from other elements of an image analysis pipeline.

Funding

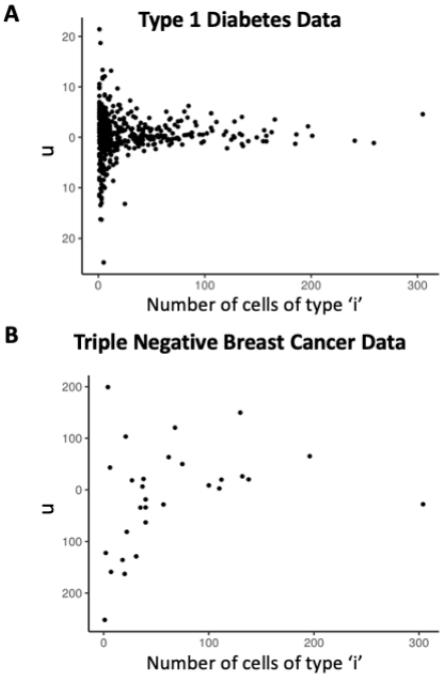
This work was partly supported by the University of Sydney and an Australian Research Council Discovery Early Career Researcher Award [DE200100944] funded by the Australian Government. The funders had no role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflict of Interest: none declared.

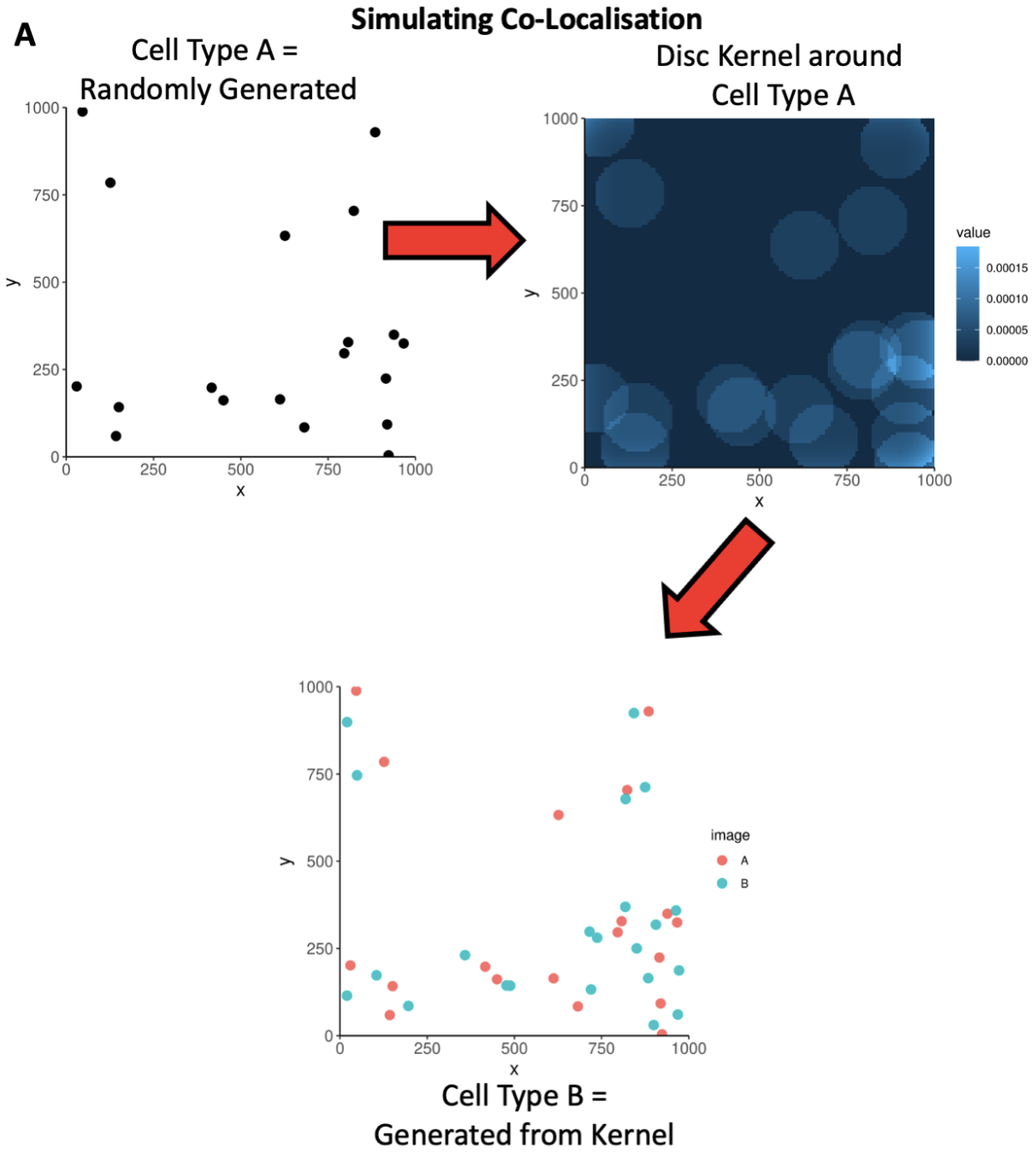
Data availability

The data underlying this article are available in https://github.com/nickcee/spicyRPaper.

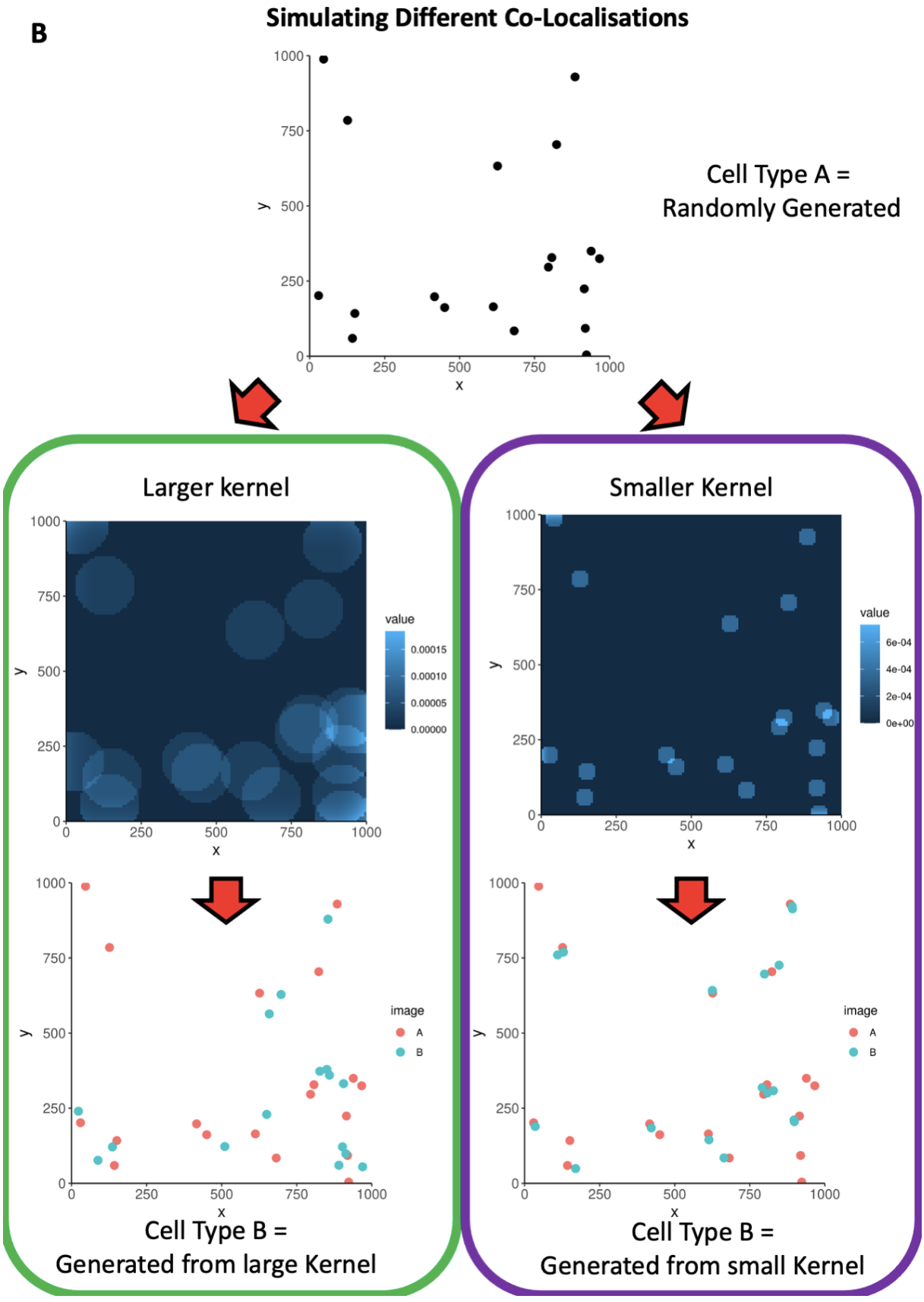
3.5 Supplementary Material



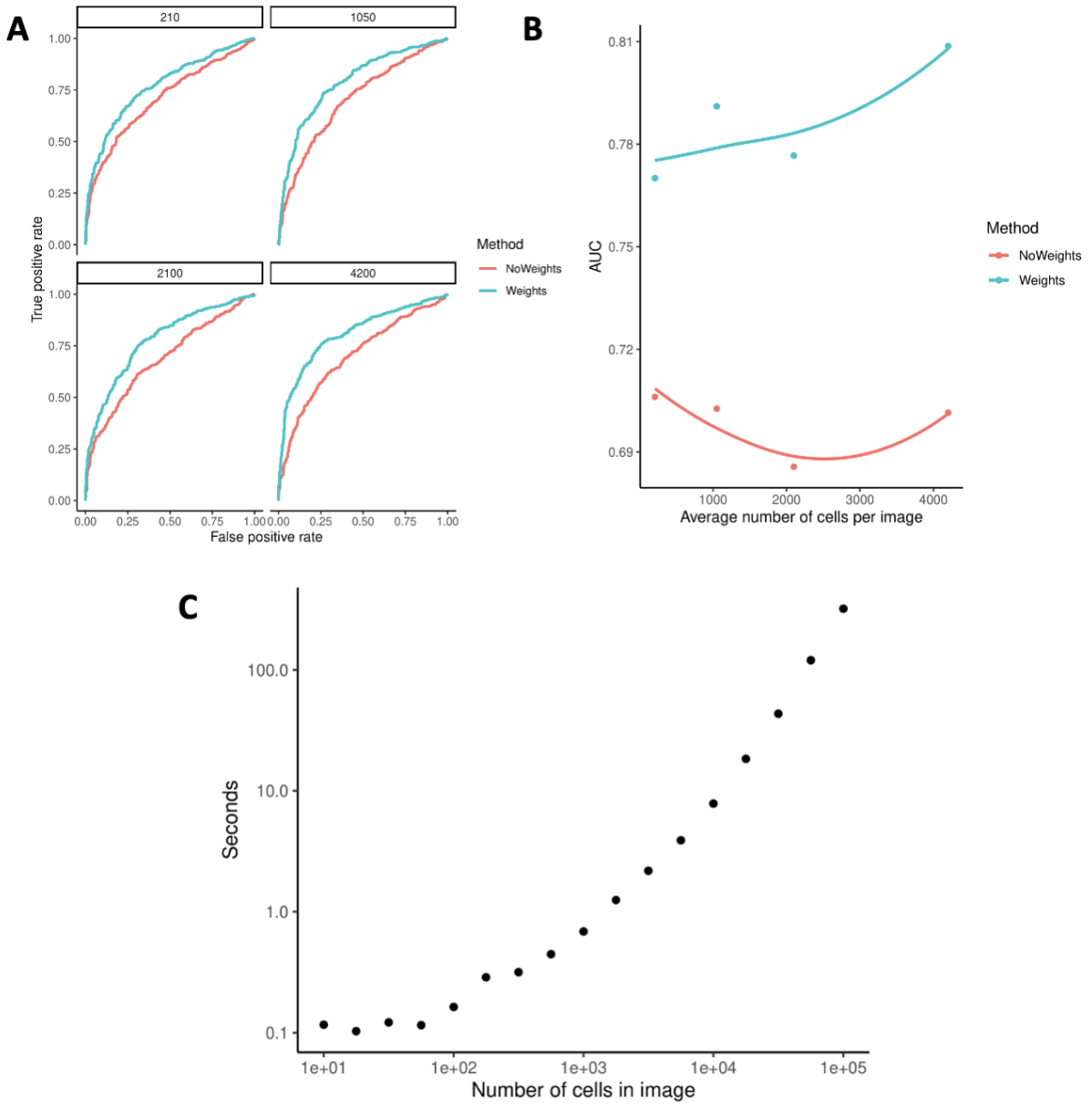
Supplementary Figure 3.1: **Relationship between the number of cells and variability of the quantification of localisation.** u as a function of the number of cells of type 'i' for each image. As the number of cell types is decreased, the number variation in the u statistic is increased. This is seen across three different datasets: **(A)** Imaging mass cytometry images of type 1 diabetes samples from Damond et al. 2019 **(B)** Multiplexed ion beam imaging by time-of-flight (MIBI-TOF) images of triple negative breast cancer samples from Keren et al. 2018.



Supplementary Figure 3.2: **Summary of how simulations are performed. (A)** Schematic of how co-localisation is simulated. Cells in cell type A is randomly generated with Poisson point process model. The density of cell type A is then calculated using a disc kernel, where the size of the disc is of a specified value. Cells from cell type B are then generated with a Poisson point process model within the kernel.

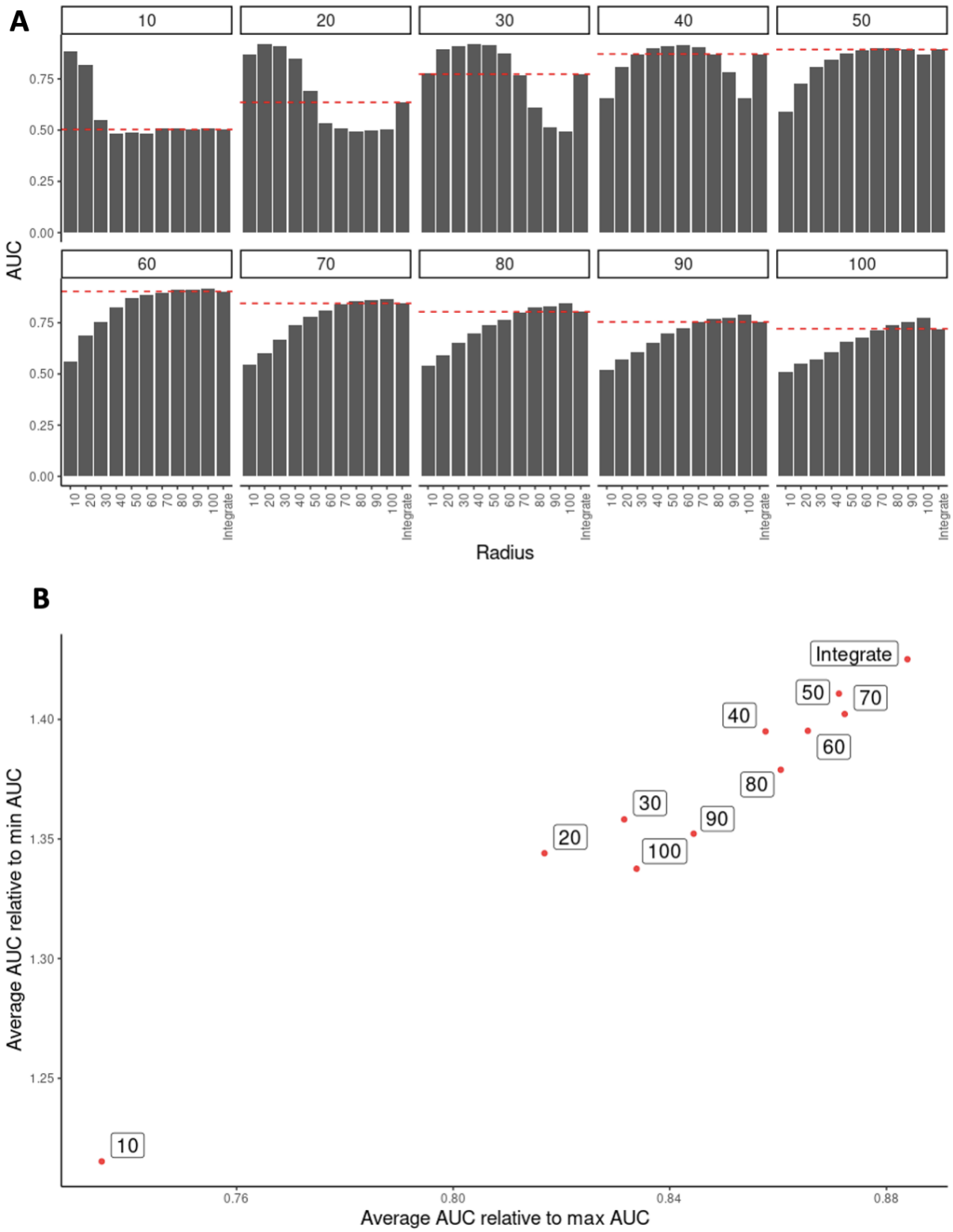


Supplementary Figure 3.2: **(B)** Different levels of co-localisation are generated using different sized kernels.



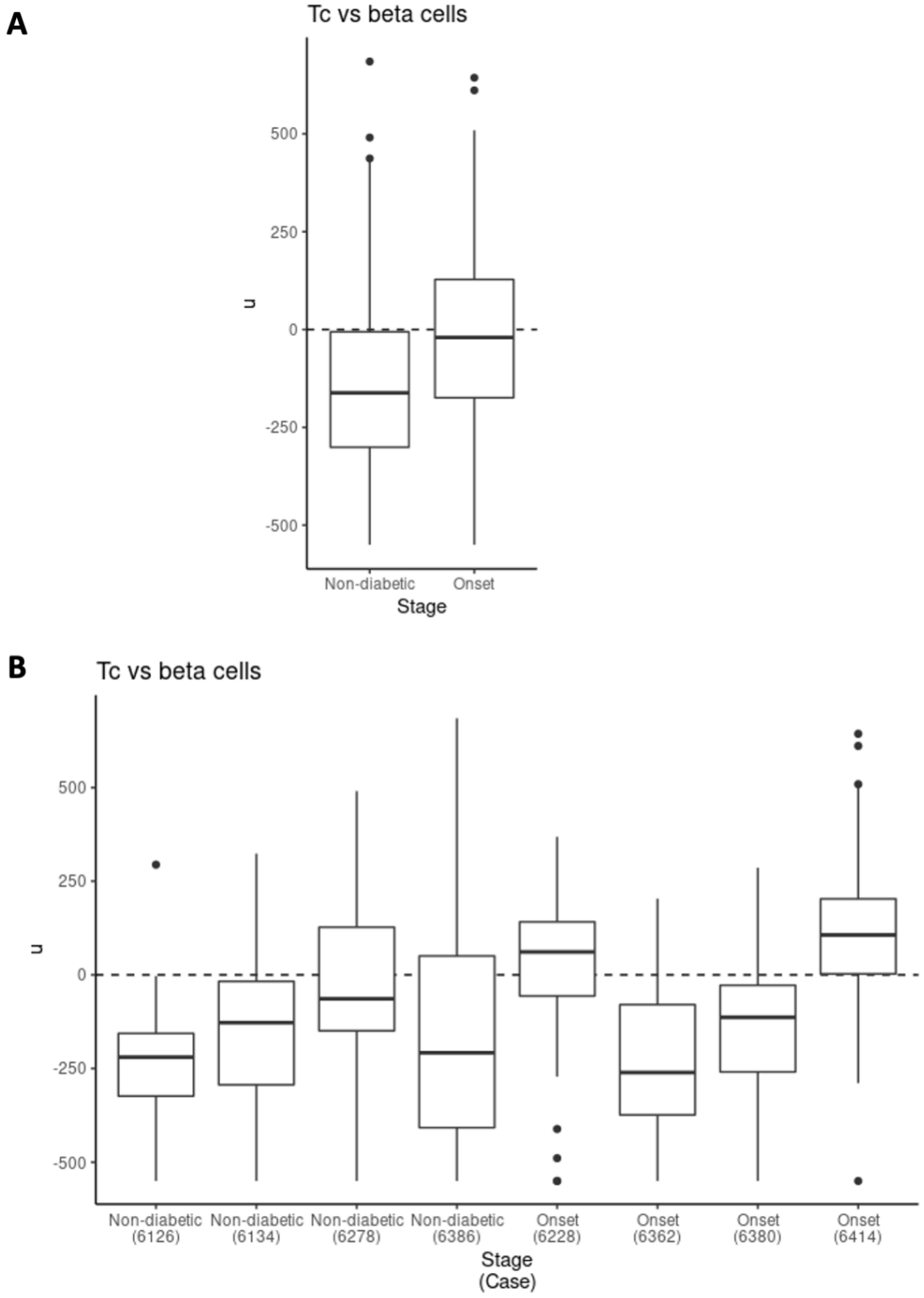
Supplementary Figure 3.3: **Simulations demonstrating the performance of the spicyR framework with varying cell counts.** Simulations were performed to demonstrate the performance of spicyR as the average number of cells of the two cell types being compared are increased. The simulation described in the Methods section is modified so that the number of cells in each image are multiplied by either 1, 5, 10 or 20 in respective simulations. **(A)** The receiver operating characteristic curve for each simulation with the average for each simulation, with the average number of cells in each image list labelled above. **(B)** Summary of the area under the curve of the receiver operating characteristic curve (AUC) for each simulation, as a function of average cell counts. **(C)** Simulations were performed to calculate the time taken to calculate L-functions for images with increasing cell numbers.

This page has been left blank intentionally.



Supplementary Figure 3.4: **Simulations demonstrating the performance of the spicyR framework with varying degrees of co-localisation.** (continued on next page)

Supplementary Figure 3.4: Simulations were performed to demonstrate the performance of spicyR as the colocalisation distance between cell types varies.The parameter lambda, which controls the degree of co-localisation of cell type A and cell type B, was varied from 10 to 100. **(A)** AUC values for simulations with varying co-localisation distances, given by the heading. The x-axis represents the distance at which a co-localisation score was calculated, or if it was calculated through integrating ('Integrate'). The red line signifies the AUC obtained when integrating to measure the co-localisation score. **(B)** Values in **(A)** were averaged across all simulations give the average AUC relative to the best performing radius in each simulation (x-axis) and the average AUC relative to the worst performing radius in each simulation (y-axis).



Supplementary Figure 3.5: **Cytotoxic T cells avoid beta cells in non-diabetic islets.** **(A)** Boxplots quantifying the level of co-localisation, u, of Tc and beta cells in islets from nondiabetic and onset diabetes patients. **(B)** The islets are stratified by patient. The difference in co-localisation in **(A)** is primarily driven by two patients in **(B)**.

Supplementary Table 3.1: **Calculating the L curves for each image is the most time-consuming component of spicyR.** This relationship is observed when applying spicyR to three different datasets; **a)** diabetes patients assayed with IMC by Damond et al. (2019) [18], **b)** breast cancer patients assayed with IMC by Jackson et al. (2020) [16] and **c)** colorectal cancer patients assayed with CODEX by Schürch et al. (2020) [20].

Dataset	Technology	Number of Images	Average number of cells per image	Number of subjects	Number of cell types	Time with one core (minutes)	Time with eight cores (minutes)
Damond et al. (2019)	IMC	845	2100	12	16	63	13.5
Jackson et al. (2020)	IMC	100	2851	100	27	6.4	3.1
Schurch et al. (2020)	CODEX	140	1680	35	13	2.7	1.4

Supplementary Table 3.2: **Results from applying spicyR to compare Onset diabetes with Non-diabetes data.** Fitting a mixed effects model with spicyR identifies four pairwise relationships with a nominal P-value less than 0.05. Reported from the model are the Intercept (the colocalisation score for the Non-diabetics), the Coefficient (the change in co-localisation in Onset diabetes relative to Non-diabetics), the P-value for the coefficient of change, the Adjusted P-value accounting for multiple testing (FDR correction), the first cell type (From) compared to the second cell type (To).

Intercept	Coefficient	P-value	Adjusted P-value	From	To
-117.34	273.45	0.0014	0.061	Th	Th
590.63	-184.00	0.0083	0.14	Beta	Delta
590.42	-182.44	0.0084	0.14	Delta	Beta
-67.16	74.88	0.042	0.31	Neutrophil	Tc

Bibliography

[1] Yury Goltsev, Nikolay Samusik, Julia Kennedy-Darling, Salil S. Bhate, Matthew Hale, Gustavo Vazquez, Sarah Black, and Garry P. Nolan. Deep profiling of mouse splenic architecture with codex multiplexed imaging. *Cell*, 174(4):968–981, 2018. doi: 10.1016/j.cell.2018.07.010. URL https://doi.org/10.1016/j.cell.2018.07.010.

[2] Jia-Ren Lin, Mohammad Fallahi-Sichani, and Peter K. Sorger. Highly multiplexed imaging of single cells using a high-throughput cyclic immunofluorescence method. *Nature communications*, 6(1):8390–8390, 2015. doi: 10.1038/ncomms9390. URL https://doi.org/10.1038/ncomms9390.

[3] Gabriele Gut, Markus D. Herrmann, and Lucas Pelkmans. Multiplexed protein maps link subcellular organization to cellular states. *Science (New York, N.Y.)*, 361(6401): NA–NA, 2018. doi: 10.1126/science.aar7042. URL https://doi.org/10.1126/science.aar7042.

[4] Charlotte Giesen, Hao A. O. Wang, Denis Schapiro, Nevena Zivanovic, Andrea Jacobs, Bodo Hattendorf, Peter J. Schüffler, Daniel Grolimund, Joachim M. Buhmann, Simone Brandt, Zsuzsanna Varga, Peter J. Wild, Detlef Günther, and Bernd Bodenmiller. Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nature methods*, 11(4):417–422, 2014. doi: 10.1038/nmeth.2869. URL https://doi.org/10.1038/nmeth.2869.

[5] Michael Angelo, Sean C. Bendall, Rachel Finck, Matthew Hale, Chuck Hitzman, Alexander D. Borowsky, Richard M. Levenson, John B. Lowe, Scot D. Liu, Shuchun Zhao, Yasodha Natkunam, and Garry P. Nolan. Multiplexed ion beam imaging of human breast tumors. *Nature medicine*, 20(4):436–442, 2014. doi: 10.1038/nm.3488. URL https://doi.org/10.1038/nm.3488.

[6] Sanja Vickovic, Gökcen Eraslan, Fredrik Salmén, Johanna Klughammer, Linnea Stenbeck, Denis Schapiro, Tarmo Äijö, Richard Bonneau, Ludvig Bergenstråhle, José Fernández Navarro, Joshua Gould, Gabriel K. Griffin, Åke Borg, Mostafa Ronaghi, Jonas Frisén, Joakim Lundeberg, Aviv Regev, and Patrik L. Ståhl. High-definition spatial transcriptomics for in situ tissue profiling. *Nature methods*, 16(10): 987–990, 2019. doi: 10.1038/s41592-019-0548-y. URL https://doi.org/10.1038/s41592-019-0548-y.

[7] Eric Lubeck, Ahmet F. Coskun, Timur Zhiyentayev, Mubhij Ahmad, and Long Cai. Single-cell in situ rna profiling by sequential hybridization. *Nature methods*, 11(4): 360–361, 2014. doi: 10.1038/nmeth.2892. URL https://doi.org/10.1038/nmeth.2892.

[8] Heeva Baharlou, Nicolas P Canete, Anthony L. Cunningham, Andrew N. Harman, and Ellis Patrick. Mass cytometry imaging for the study of human diseases-applications and data analysis strategies. *Frontiers in immunology*, 10(NA):2657–2657, 2019. doi: 10.3389/fimmu.2019.02657. URL https://doi.org/10.3389/fimmu.2019.02657.

[9] Anne E. Carpenter, Thouis R. Jones, Michael R. Lamprecht, Colin Clarke, In Han Kang, Ola Friman, David A. Guertin, Joo Han Chang, Robert A. Lindquist, Jason Moffat, Polina Golland, and David M. Sabatini. Cellprofiler: image analysis

software for identifying and quantifying cell phenotypes. *Genome biology*, 7(10): 1–11, 2006. doi: 10.1186/gb-2006-7-10-r100. URL https://doi.org/10.1186/gb-2006-7-10-r100.

[10] Christoph Sommer, Christoph N. Straehle, Ullrich Köthe, and Fred A. Hamprecht. Isbi - ilastik: Interactive learning and segmentation toolkit, 2011. URL https://doi.org/10.1109/isbi.2011.5872394.

[11] David Van Valen, Takamasa Kudo, Keara Michelle Lane, Derek N. Macklin, Nicolas Quach, Mialy M. DeFelice, Inbal Maayan, Yu Tanouchi, Euan A. Ashley, and Markus W. Covert. Deep learning automates the quantitative analysis of individual cells in live-cell imaging experiments. *PLoS computational biology*, 12(11):e1005177–NA, 2016. doi: 10.1371/journal.pcbi.1005177. URL https://doi.org/10.1371/journal.pcbi.1005177.

[12] Denis Schapiro, Hartland W. Jackson, Swetha Raghuraman, Jana R. Fischer, Vito R. T Zanotelli, Daniel Schulz, Charlotte Giesen, Raul Catena, Zsuzsanna Varga, and Bernd Bodenmiller. histocat: analysis of cell phenotypes and interactions in multiplex image cytometry data. *Nature methods*, 14(9):873–876, 2017. doi: 10.1038/nmeth.4391. URL https://doi.org/10.1038/nmeth.4391.

[13] Leeat Keren, Marc Bosse, Diana M. Marquez, Roshan Angoshtari, Samir Jain, Sushama Varma, Soo-Ryum Yang, Allison W. Kurian, David Van Valen, Robert B. West, Sean C. Bendall, and Michael Angelo. A structured tumor-immune microenvironment in triple negative breast cancer revealed by multiplexed ion beam imaging. *Cell*, 174(6):1373–1387, 2018. doi: 10.1016/j.cell.2018.08.039. URL https://doi.org/10.1016/j.cell.2018.08.039.

[14] Adrian Baddeley, Ege Rubak, and Rolf Turner. *Spatial Point Patterns*, volume NA of *NA*. 2015. doi: 10.1201/b19708. URL https://doi.org/10.1201/b19708.

[15] Brian D Ripley. The second-order analysis of stationary point processes. *Journal of applied probability*, 13(2):255–266, 1976.

[16] Hartland W. Jackson, Jana R. Fischer, Vito R. T Zanotelli, H. Raza Ali, Robert Mechera, Savas D. Soysal, Holger Moch, Simone Muenst, Zsuzsanna Varga, Walter P. Weber, and Bernd Bodenmiller. The single-cell pathology landscape of breast cancer. *Nature*, 578(7796):615–620, 2020. doi: 10.1038/s41586-019-1876-x. URL https://doi.org/10.1038/s41586-019-1876-x.

[17] Damien Arnol, Denis Schapiro, Bernd Bodenmiller, Julio Saez-Rodriguez, and Oliver Stegle. Modeling cell-cell interactions from spatial molecular data with spatial variance component analysis. *Cell reports*, 29(1):202–211.e6, 2019. doi: 10.1016/j.celrep.2019.08.077. URL https://doi.org/10.1016/j.celrep.2019.08.077.

[18] Nicolas Damond, Stefanie Engler, Vito R. T Zanotelli, Denis Schapiro, Clive Wasserfall, Irina Kusmartseva, Harry S. Nick, Fabrizio Thorel, Pedro Luis Herrera, Mark A. Atkinson, and Bernd Bodenmiller. A map of human type 1 diabetes progression by imaging mass cytometry. *Cell metabolism*, 29(3):755–768, 2019. doi: 10.1016/j.cmet.2018.11.014. URL https://doi.org/10.1016/j.cmet.2018.11.014.

[19] Anniina Färkkilä, Doga Gulhan, Julia Casado, Connor A. Jacobson, Huy V. Nguyen, Bose Kochupurakkal, Zoltan Maliga, Clarence Yapp, Yu-An Chen, Denis Schapiro, Yinghui Zhou, Julie R. Graham, Bruce J. Dezube, Pamela N. Munster, Sandro Santagata, Elizabeth P. Garcia, Scott J. Rodig, Ana Lako, Dipanjan Chowdhury, Geoffrey I. Shapiro, Ursula A. Matulonis, Peter J. Park, Sampsa Hautaniemi, Peter K. Sorger, Elizabeth M. Swisher, Alan D. D'Andrea, and Panagiotis A. Konstantinopoulos. Immunogenomic profiling determines responses to combined parp and pd-1 inhibition in ovarian cancer. *Nature communications*, 11(1):1459–1459, 2020. doi: 10.1038/s41467-020-15315-8. URL https://doi.org/10.1038/s41467-020-15315-8.

[20] Christian M Schürch, Salil S Bhate, Graham L Barlow, Darci J Phillips, Luca Noti, Inti Zlobec, Pauline Chu, Sarah Black, Janos Demeter, David R McIlwain, Shigemi Kinoshita, Nikolay Samusik, Yury Goltsev, and Garry P. Nolan. Coordinated cellular neighborhoods orchestrate antitumoral immunity at the colorectal cancer invasive front. *Cell*, 182(5):1341–1359.e19, 2020. doi: 10.1016/j.cell.2020.07.005. URL https://doi.org/10.1016/j.cell.2020.07.005.

[21] Natalya Pya and Simon N. Wood. Shape constrained additive models. *Statistics and Computing*, 25(3):543–559, 2014. doi: 10.1007/s11222-013-9448-7. URL https://doi.org/10.1007/s11222-013-9448-7.

[22] Douglas M. Bates, Martin Mächler, Benjamin M. Bolker, and Steven C. Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67 (1):1–48, 2015.

[23] Alexandra Kuznetsova, Per B. Brockhoff, and Rune Haubo Bojesen Christensen. lmertest package: Tests in linear mixed effects models. *Journal of Statistical Software*, 82(13):1–26, 2017.

[24] Adrian Baddeley and Rolf Turner. spatstat: An r package for analyzing spatial point patterns. *Journal of Statistical Software*, 12(6):1–42, 2005.

[25] Nils Eling, Nicolas Damond, Tobias Hoch, and Bernd Bodenmiller. Cytomapper: an r/bioconductor package for visualisation of highly multiplexed imaging data. *Bioinformatics (Oxford, England)*, 36(24):5706–5708, 2020. doi: 10.1093/bioinformatics/btaa1061. URL https://doi.org/10.1093/bioinformatics/btaa1061.

Chapter 4

An *in situ* analysis pipeline for initial host-pathogen interactions reveals signatures of human colorectal HIV transmission

Publications incorporated into this chapter:

Baharlou, H, **Canete, N.P**, Vine, E.E.. et al. (2022) An *in situ* analysis pipeline for initial host-pathogen interactions reveals signatures of human colorectal HIV transmission. *Cell Rep. 2022 Sep 20;40(12):111385.*

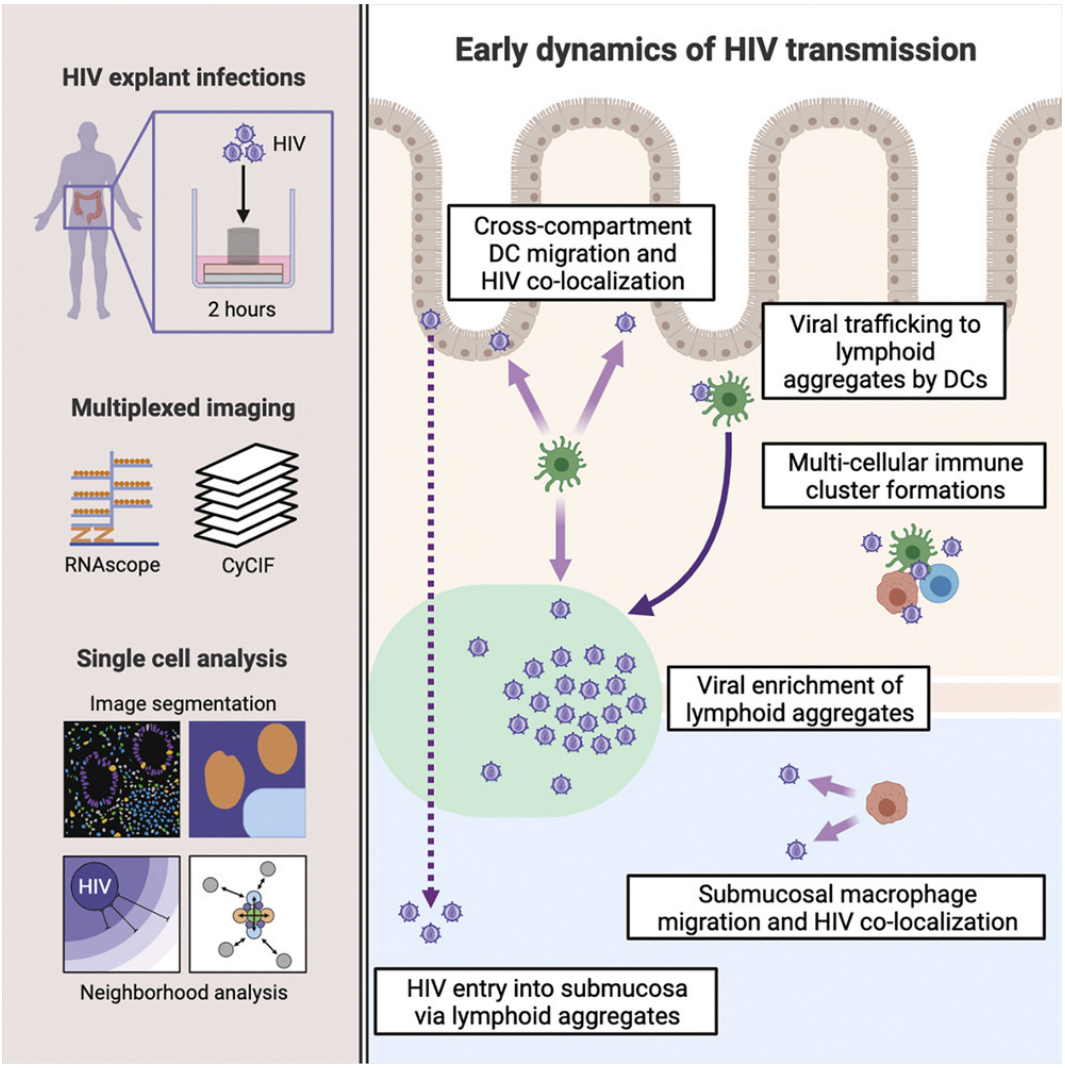
Preface

The primary motivation for the computational analysis techniques developed throughout the thesis was the HIV CyCIF data being acquired by the host lab. This chapter consists of a publication in Cell Reports, which describes the analysis of this data. The project was conceived and guided by first author H. Baharlou. As second author, I designed a significant portion of the image analysis pipeline used in analyzing the dataset. One of my biggest contributions was the implementation of the MATLAB codebase used to perform single-cell segmentation, cell classification, RNA spot counting and key data extraction (as described in Figure 4.1C-G and Supplementary Figure 4.1A-C). I also contributed to the development of the downstream analysis implemented in R, which facilitated the remaining quantitative analysis of the dataset. Under the direction of the first author, I helped guide the downstream analysis being performed, writing the codebase in R (Figure 4.2B-D,F, Figure 4.3A,C,E,F, Figure 4.2A-C,G-I, Figure 4.5, Figure 4.6A,C,E-G, Supplementary Figure 4.2B,D, Supplementary Figure 4.2, Supplementary Figure 4.3A-L, Supplementary Figure 4.4A,B,E). This analysis followed directly from the data that I had extracted with the MATLAB pipeline. Additionally, the tools developed in Chapters 2 (Figure 4.1 and Supplementary Figure 4.1) and 3 (Figure 4.6 and Supplementary Figure 4.4A,B,E) were implemented in this publication.

Summary

The initial immune response to HIV determines transmission. However, due to technical limitations we still do not have a comparative map of early mucosal transmission events. By combining RNAscope, cyclic immunofluorescence, and image analysis tools, we quantify HIV transmission signatures in intact human colorectal explants within 2 h of topical exposure. We map HIV enrichment to mucosal dendritic cells (DCs) and submucosal macrophages, but not CD4+ T cells, the primary targets of downstream infection. HIV+ DCs accumulate near and within lymphoid aggregates, which act as early sanctuaries of high viral titers while facilitating HIV passage to the submucosa. Finally, HIV entry induces recruitment and clustering of target cells, facilitating DC- and macrophage-mediated HIV transfer and enhanced infection of CD4+ T cells. These data demonstrate a rapid response to HIV structured to maximize the likelihood of mucosal infection and provide a framework for *in situ* studies of host-pathogen interactions and immune-mediated pathologies.

Graphical Abstract



4.1 Introduction

Thirty-seven million people are infected with HIV and, despite the introduction of pre-exposure prophylaxis, there were still 1.5 million new infections in 2020. Blocking transmission of HIV therefore remains a high global priority that requires an effective vaccine and, in the meantime, better prophylactic interventions. To achieve these goals, we need a better understanding of the initial events that govern transmission of HIV, particularly early viral interactions with antigen-presenting cells, such as dendritic cells (DCs) and macrophages, and their subsequent delivery of the virus to CD4+ T cells [1].

Simian immunodeficiency virus challenge studies have shown that productive mucosal infection at the site of transmission precedes the detection of plasma viremia, and that the viral reservoir is rapidly seeded at this site within days of exposure [2, 3]. Concordant data have also been described in human studies [4, 5]. Due to technological limitations, early transmission studies have been limited to time points after initial viral integration/replication [1, 6, 7] or the use of model systems, with human tissue studies mostly performed on isolated cells or by imaging of limited parameters [1, 8–12]. Thus, we still do not know the initial events that lead to mucosal HIV infection in the human tissues where transmission occurs.

The next stage in advancing our understanding of these events requires an *in situ* quantitative multi-parameter study to understand the relative involvement of multiple target cells within anatomically distinct tissue compartments. This "top down" approach is critical for establishing physiological relevance and guiding the rational selection of specific biological mechanisms for in-depth characterization studies. To our knowledge, no study has examined all these processes at once in the context of pathogen invasion of human tissue. Such studies have been hampered to date by a plethora of issues including parameter limitations of conventional microscopy, a lack of suitable image processing and analysis algorithms and, in the context of HIV, difficulties with *in situ* pathogen detection at early time points [13]. We have recently pioneered the use of RNAscope to visualize clinically relevant HIV virions interacting with anogenital target cells *in situ* within 2 h of topical exposure [8, 11]. We have also designed post image acquisition algorithms to remove autofluorescence [14] and quantify cell interaction changes between disease states [15]. In this study we have utilized these approaches as well as designed tools to segment full cell bodies more accurately, allowing us to quantify the signatures of HIV transmission across human colorectal tissue within 2 h of exposure. We have defined the spatial distribution of the three key colorectal HIV target cells (DCs, macrophages, and CD4+ T cells) across all colorectal tissue compartments (epithelium [EP], lamina propria [LP], lymphoid aggregates [LAs], and submucosa [SM]) and shown how they respond to HIV. We show conclusively that HIV is initially enriched within mucosal (EP, LP, LA) DCs and submucosal macrophages, but not CD4+ T cells and that the virus is preferentially enriched in LAs. We also provide strong circumstantial evidence that LP DCs traffic virus to LAs and that these structures themselves may provide a conduit for rapid HIV entry into the deeper submucosal layer, where it preferentially associates with macrophages. Finally, we show that HIV mucosal entry induces its target cells to form multi-cellular clusters within which HIV+ DCs and macrophages preferentially cluster with CD4+ T cells, leading to viral transfer and enhanced infection of CD4+ T cells, supported by *ex vivo* data.

4.2 Results

4.2.1 Analysis pipeline for mapping HIV-target cell interactions *in situ*

This study explores the interactions of HIV with its three key target cells — DCs, macrophages, and CD4+ T cells—in human colorectal tissue at the very earliest time points following HIV challenge using a combination of RNAscope, multiplexed fluorescence microscopy, and a custom image-processing and analysis pipeline (Figure 4.1A). Lab-adapted and transmitted/founder strains of HIV were applied to the apical surface of intact fresh human colorectal tissue for 2 h. RNAscope was performed to detect HIV RNA (virions), followed by cyclic immunofluorescence (CyCIF) microscopy, which was used to identify nuclei (DAPI), EP (E-Cadherin), CD4+ T cells (CD3+CD4+), CD4- T cells (CD3+CD4-), DCs (CD11c+), and macrophages (FXIIIa+) (Figure 4.1A-B). While having little to no capacity for HIV uptake themselves, CD4- T cells (primarily CD8+ T cells) served as a useful control for non-specific HIV-cell interactions *in situ*. We removed autofluorescence signal, a known feature of colorectal tissue imaging [16] post-acquisition using our autofluorescence identifier (AFid) algorithm [14] (Figure 4.1A) and Supplementary Figure 4.1A). We next performed segmentation of HIV virions, cells, and tissue compartments (EP, LP, LA, SM) (Figure 4.1C). Cells were classified and segmented using a custom approach that estimates full cell bodies and allows two cells to physically overlap. This enabled virions to be accurately assigned to amorphous cells such as macrophages and DCs (Supplementary Figure 4.1B-C). This pipeline allowed us to make compartment-specific measurements of target cell composition (Figure 4.1D), HIV status (HIV+ versus HIV-), virion load per cell (Figure 4.1E), target cell migration to/from HIV (Figure 4.1F), and the effect of HIV on interactions between target cells (Figure 4.1G).

To complement and validate these *in situ* data we used an orthogonal approach of flow cytometry analysis of HIV target cells following their dissociation from tissue and infection with HIV (Supplementary Figure 4.1D-F). Using our HIV p24 uptake assay, we compared HIV uptake *in situ* with that of *ex-vivo*-dissociated cells. In addition, HIV-induced cell:cell interactions observed *in situ* (Figure 4.1G) were further investigated by sorting and infecting co-cultures of these cells to determine whether their interaction lead to enhanced viral transfer and infection (Figure 4.1H). Together these approaches enabled us to create an *in situ* quantitative map of how HIV is distributed across colorectal target cells and tissue compartments, as well as the HIV-induced cell:cell interactions that occur in the mucosa within 2 h post-exposure and prior to systemic viral spread.

4.2.2 HIV-target cell composition and distribution within human colorectal tissue

Using our analysis pipeline, we defined the relative proportion and density of HIV target cells within EP, LP, LAs, and SM in fresh uninfected human colorectal tissue (Figure 4.2A-B). Although only 1% of EP cells were HIV target cells, DCs were the most abundant (58%), followed by CD4+ T cells (28%) then macrophages (14%). Of the LP cells, 11% were HIV target cells, with CD4+ T cells being most abundant (50%) followed by DCs (32%) and then macrophages (18%). In LAs, 35% of cells were HIV target cells consisting of CD4+ T cells (57%) and DCs (43%) with negligible presence of macrophages. Five percent of SM cells were HIV target cells, with macrophages being most abundant (56%)

followed by CD4+ T cells (25%) and DCs (19%). Cell density measurements closely followed these trends for each tissue compartment. Thus, LAs and LP had the highest density of HIV target cells, with DCs and CD4+ T cells dominating the mucosal layers and macrophages dominating the submucosal layer.

We next examined the spatial distribution of HIV target cells within each compartment. This was achieved by using the border between tissue compartments as an anchor and measuring changes in cell density from this reference point. LP macrophages were enriched <10 µm from the EP, while DCs, CD4+ T cells, and CD4- T cells preferentially localized >10 µm away from the EP (Wilcoxon, $p < 0.005$ for all, Figure 4.2C). In LP, DCs, and CD4+ T cells were enriched <200 µm from LAs (Wilcoxon, $p = 0.005$). This was also true for EP DCs (Wilcoxon, $p = 0.001$) but was not measurable for EP CD4+ T cells due to their low frequency in EP (Figure 4.2D).

We next focused our attention on characterizing LAs as they contained the highest density of HIV target cells, particularly DCs and CD4+ T cells (Figure 4.2A-B). LAs were present at a median density of 0.6 structures per mm^2 of tissue and varied in their area (median 0.09 mm^2; range 0.05-0.17 mm^2), diameter (median 283 µm; range 195-401 µm), and cell number (median 1,800 cells; range 860-3,015) (Figure 4.2E-F).

4.2.3 HIV viral particles are enriched in colorectal DCs and macrophages within 2 hours

We next assessed the interactions between HIV and its target cells using lab-adapted (BaL) or transmitted/founder (Z3678M) HIV strains. To ensure RNAscope probes were specific to HIV RNA we stained uninfected tissue and confirmed that no signal was detectable (Supplementary Figure 4.2A). As our pipeline incorporates automated virion detection, we compared mock and HIV-treated explants to calculate the false detection rate. One particle per 1,000 cells was detected in mock versus 30 particles per 1,000 cells in HIV-treated explants, indicating that only 3% of HIV+ cells in our treated samples were false positives (Supplementary Figure 4.2B). To determine whether HIV was enriched among its target cells we measured the percentage of total HIV particles in this population (HIV percentage), as well as the percentage of all cells that were target cells (target cell percentage). A chi-square test was then performed to test for whether enrichment among target cells was significant. The formula "log2(HIV percentage/target cell percentage)" measured the relative level of HIV enrichment between images. Although the percentage of total HIV particles localizing to target cells varied across images, the majority showed significant HIV enrichment in target cells compared with the remaining undefined "other cells" (chi-square test, $p < 0.05$, Figure 4.3A), with up to 4-fold enrichment observed in some images (Figure 4.3A, bottom). Visual inspection confirmed the specificity of HIV association with target cells over other cells for HIV_{BaL} (Figure 4.3B) and HIV_{Z3678M} (Supplementary Figure 4.2C). Using a similar approach, we tested for HIV enrichment in different target cell subsets and compared the degree of enrichment between populations (Figure 4.3C). As both HIV_{BaL} and HIV_{Z3678M} had similar trends in the degree of enrichment for each cell type (Supplementary Figure 4.2D), we combined these data to increase statistical power. HIV was enriched among DCs and macrophages (Wilcoxon, $p < 0.01$), with higher enrichment (though not significant) in DCs (Figure 4.3C). Although, HIV associated with CD4+ T cells, this occurred less than expected relative to their abundance (Wilcoxon, $p < 0.01$). In addition, CD4- T cells associated with HIV at the same frequency as their relative abundance, which is expected as they

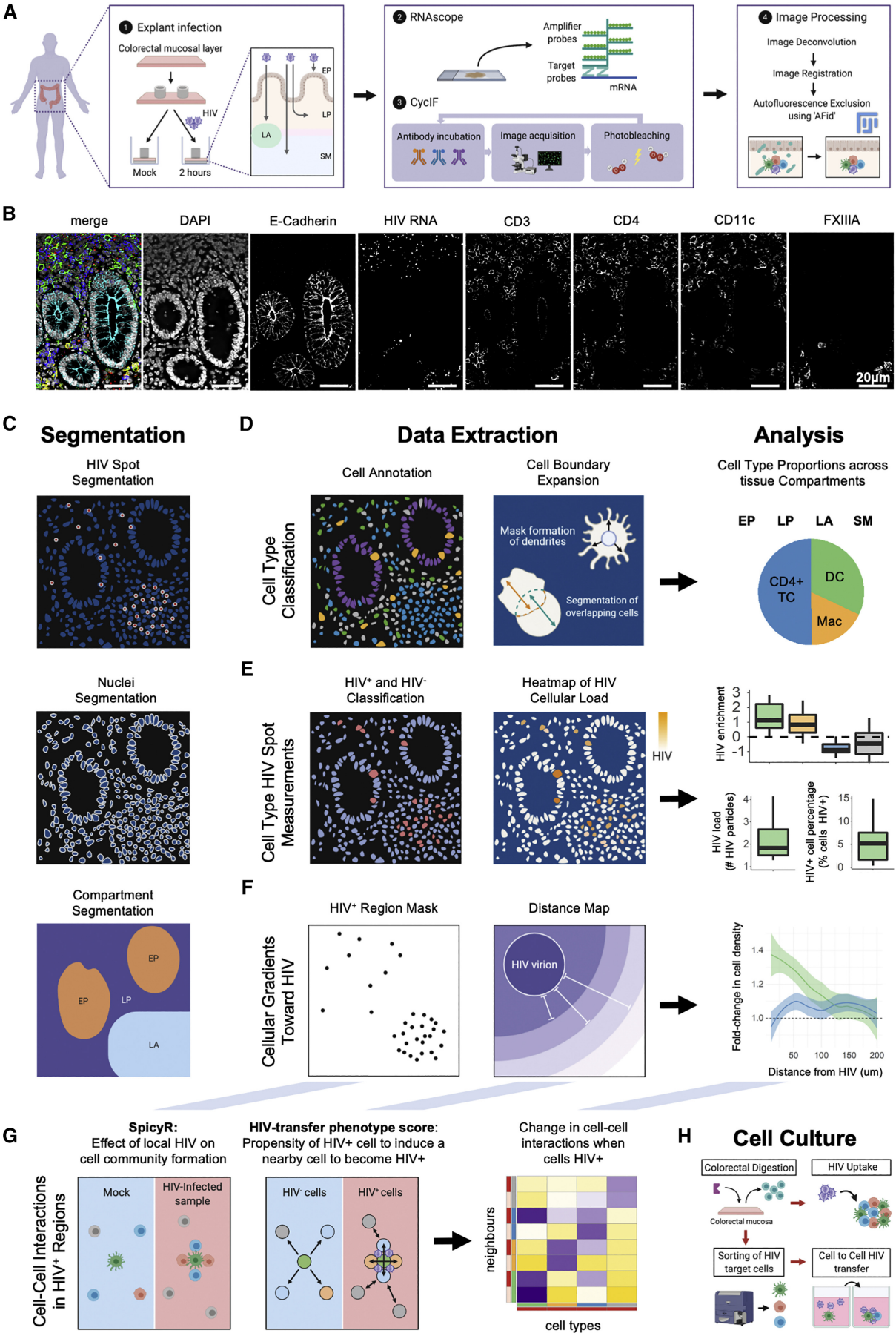


Figure 4.1: **Analysis pipeline for mapping HIV-target cell interactions *in situ*** (continued on next page)

Figure 4.1: Custom image analysis pipeline to measure HIV-cell interactions in fresh human colorectal explant tissue exposed to HIV_{Bal} or HIV_{Z3678M} for 2 h and probed with RNAscope and CyCIF. **(A)** Explant infection (1), staining with RNAscope (2), and CyCIF (3), and image processing protocol with autofluorescence removed digitally using "AFid" software (4). **(B)** Representative images from staining in (A). **(C)** Segmentation of virions, cell nuclei, and tissue compartments. **(D)** Segmented nuclei were annotated (left) and expanded to estimate full cell bodies using a custom approach (center) (details in Supplementary Figure 4.1C). Cell composition can then be determined across tissue compartments (right). **(E)** Annotation of cells as HIV+ ($\geq$1 virion) or HIV- (left) and measurement of total HIV load per cell (center). This allowed for comparisons of HIV association with different target cells (right). **(F)** Distance maps emanating from HIV particles (left and center) to measure changes in nearby cell density to infer potential cell migration to/from HIV. **(G)** SpicyR determines if the local presence of HIV alters cell communities (left), while the "HIV-transfer phenotype score" measures the propensity of an HIV+ cell to induce a nearby cell to become HIV+ (center). Frequency of interactions can be visualized using a heatmap (right). **(H)** Colorectal cell extraction and *ex vivo* HIV exposure to (1) compare 2 h HIV uptake with *in situ* results and (2) measure infection (intracellular p24 at 72 h) of CD4+ T cells co-cultured with cells that exhibited HIV-induced interactions with CD4+ T cells *in situ* in **(G)**.

do not express HIV-binding receptors. We next analyzed HIV load in individual cells in each population to determine the differential amount of virus associated with each type of target cell. Combining RNAscope, spot segmentation (Figure 4.1C), and cell segmentation with cell body estimation (Supplementary Figure 4.1B-C) enabled us to accurately measure single-particle differences between cells (Figure 4.3D). We pooled cells across all samples, stratified them by HIV particle count and measured the target cell composition within each group (Figure 4.3E). Among the 12,822 target cells analyzed, the majority (87%) contained only 1-3 HIV particles. Despite being the most prevalent cell type, CD4+ T cells were under-represented in all groups, particularly those with higher HIV particle numbers. Indeed, the only HIV target cells we could detect with >8 virions were DCs and macrophages, but never CD4+ T cells. This partly explains their relative HIV enrichment compared with CD4+ T cells.

Finally, we sought to verify whether data derived by infection of intact explants with *in situ* analysis would mirror results from infecting *ex vivo* cells extracted from tissue and analyzed by flow cytometry (Figure 4.3F). This is important, as the study of HIV transmission in human tissue has been largely confined to studies on *ex vivo* isolated cells [17–20]. Both approaches confirmed that the "percentage of cells containing HIV" and the "mean number of HIV particles per cell" were significantly greater for DCs and macrophages compared with CD4+ T cells (Wilcoxon, $p < 0.01$). However, measurements on *ex vivo* cells showed the CD4+ T cell "percentage" and "average" HIV measurements were significantly higher than CD4- T cells. In contrast no differences were observed *in situ*. This indicates that, at early time points *in situ*, HIV interactions with CD4+ T cells are not yet measurably different to that of a random interaction between HIV and a cell type with no HIV-binding capacity. Furthermore, *ex vivo* infection erroneously over-estimated macrophages as the dominant initial HIV-binding cell type, whereas *in situ* measurements showed DCs have a significantly higher mean HIV count per cell and a trend of higher HIV+ cell frequency. These results highlight the importance of physiologically relevant quantitative *in situ* studies.

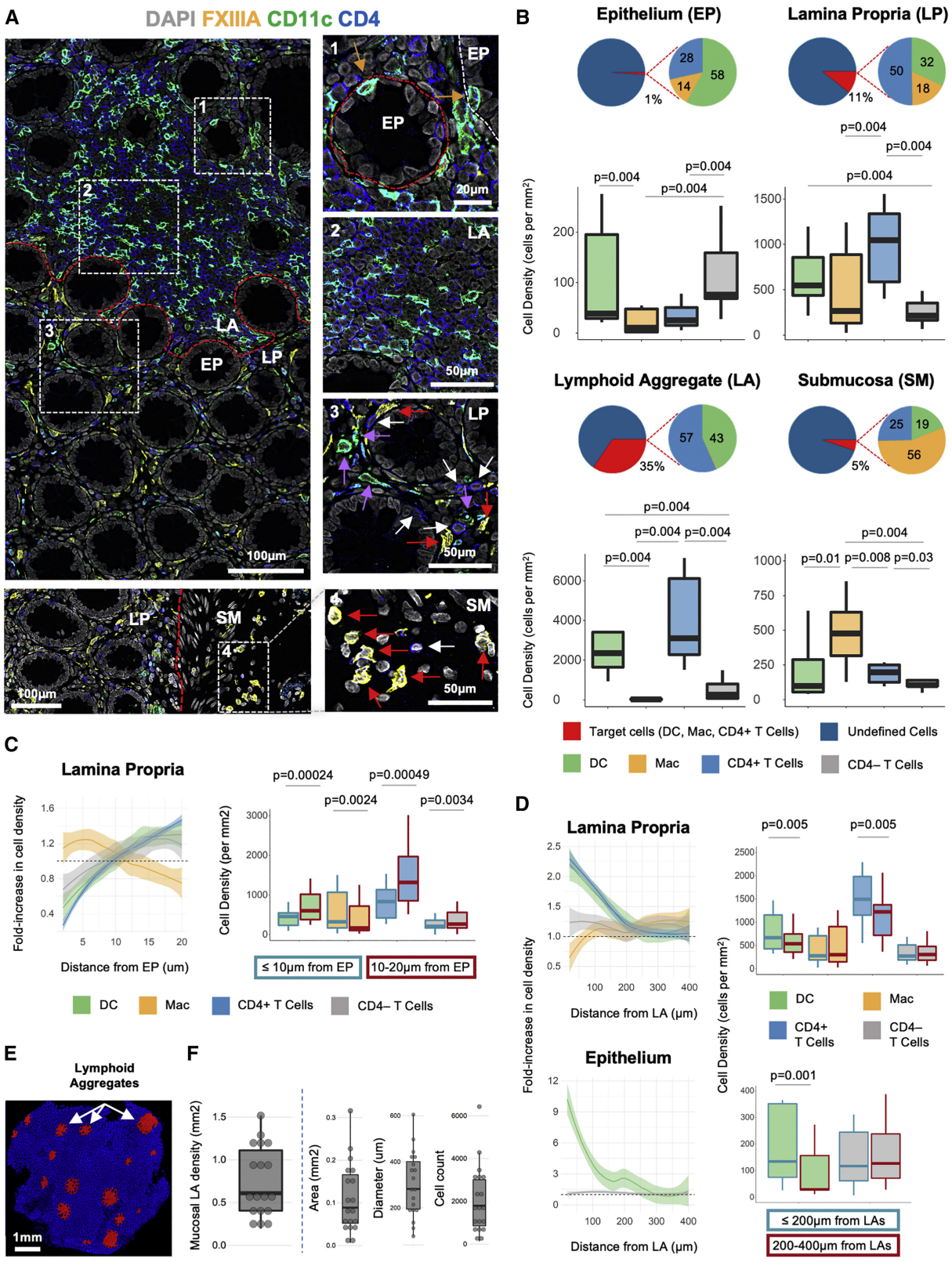


Figure 4.2: **Distribution of HIV target cells across human colorectal tissue compartments.** (continued on next page)

Figure 4.2: **(A)** Representative images of HIV target cells and their distribution. (1) DCs in the EP (brown arrows) near the LA. (2) LA DCs and CD4+ T cells. (3) LP DCs (purple arrows), macrophages (red arrows), and CD4+ T cells (white arrows). (4) Submucosal macrophages (red arrows). Broken red line, compartment borders. **(B)** HIV target cell percentage among all cells, composition, and density across tissue compartments. CD4- T cells also shown. **(C)** Change in LP target cell density with distance from EP. Left: fold change in target cell density (versus whole LP average) in 2 µm intervals from EP. Right: density of subsets in LP < 10 µm (blue border) or 10-20 µm (red border) from EP. **(D)** Change in EP or LP target cell density with distance from LAs. Left: fold change in density (versus whole EP or LP average) in 20 µm intervals from LAs. Right: density in EP or LP < 200 µm (blue border) or 200-400 µm (red border) from LAs. **(E)** Representative image of LA distribution in colorectum. Image is segmentation mask of LA (red) and non-LA (blue) cells. **(F)** Mucosal (EP + LP + LA) LA density (mm^2), average area (mm^2), diameter (µm), and cell count. n = 18 donors. Density = cells per mm^2 of DAPI. Statistics: Wilcoxon signed-rank test. n = 12 donors for **(A-E)**. LOESS curve of best fit for **(C and D)**.

Taken together, these results demonstrate that HIV can enter human colorectal explants as early as 2 h after inoculation resulting in preferential association with DCs and macrophages, compared with CD4+ T cells. Moreover, DCs and macrophages are capable of sampling larger quantities of virus at these early time points than other cells.

4.2.4 HIV localization patterns across colorectal tissue compartments and their associated target cells

We next turned our attention to tissue compartment-specific differences in the distribution of HIV particles and HIV-containing target cells. LAs contained the highest density of HIV particles reaching over 10,000 virions/mm^2 in several donors (Wilcoxon, $p < 3 \times 10^{-4}$, Figure 4.4A and Supplementary Figure 4.3A). HIV density was lowest in the EP and SM with the density in LP higher than EP (Wilcoxon, $p < 0.03$). To determine whether, upon EP penetration, HIV preferentially localizes to LP or LAs, we performed HIV enrichment analysis on compartments rather than cell types. We selected only LP and LA regions of images and compared the percentage of HIV in LAs with the expected percentage, defined as the percentage of the LP+ LA area comprised of LAs (Figure 4.4B). Accordingly, values above or below expected (dotted line) indicate preferential localization to LAs or LP, respectively. Analyzing the residuals (distance of points from dotted line) revealed preferential localization to LAs over LP (Wilcoxon, $p = 0.0007$, Figure 4.4C), which was observable upon visual inspection of images (Figure 4.4D).

Having observed HIV enrichment in DCs and macrophages (Figure 4.3C), we next determined the compartments in which this enrichment occurred. In EP, LP, and LAs HIV preferentially localized to DCs, whereas in SM the virus preferentially localized to macrophages (Figure 4.4E). Representative images are shown in Figure 4.3B (top panel) for LP and Figure 4.4F for all other compartments. In LAs 25% of donors showed >50% of HIV localized to DCs, which is far more than other compartments (Figure 4.4E). Correspondingly, LA CD4+ T cells harbored significantly less HIV than expected based on their frequency. Therefore, LA DCs may be more primed for antigen uptake, which would partly explain the high levels of HIV observed in these cells in this compartment.

Comparing the degree of HIV enrichment in cells between compartments, we observed

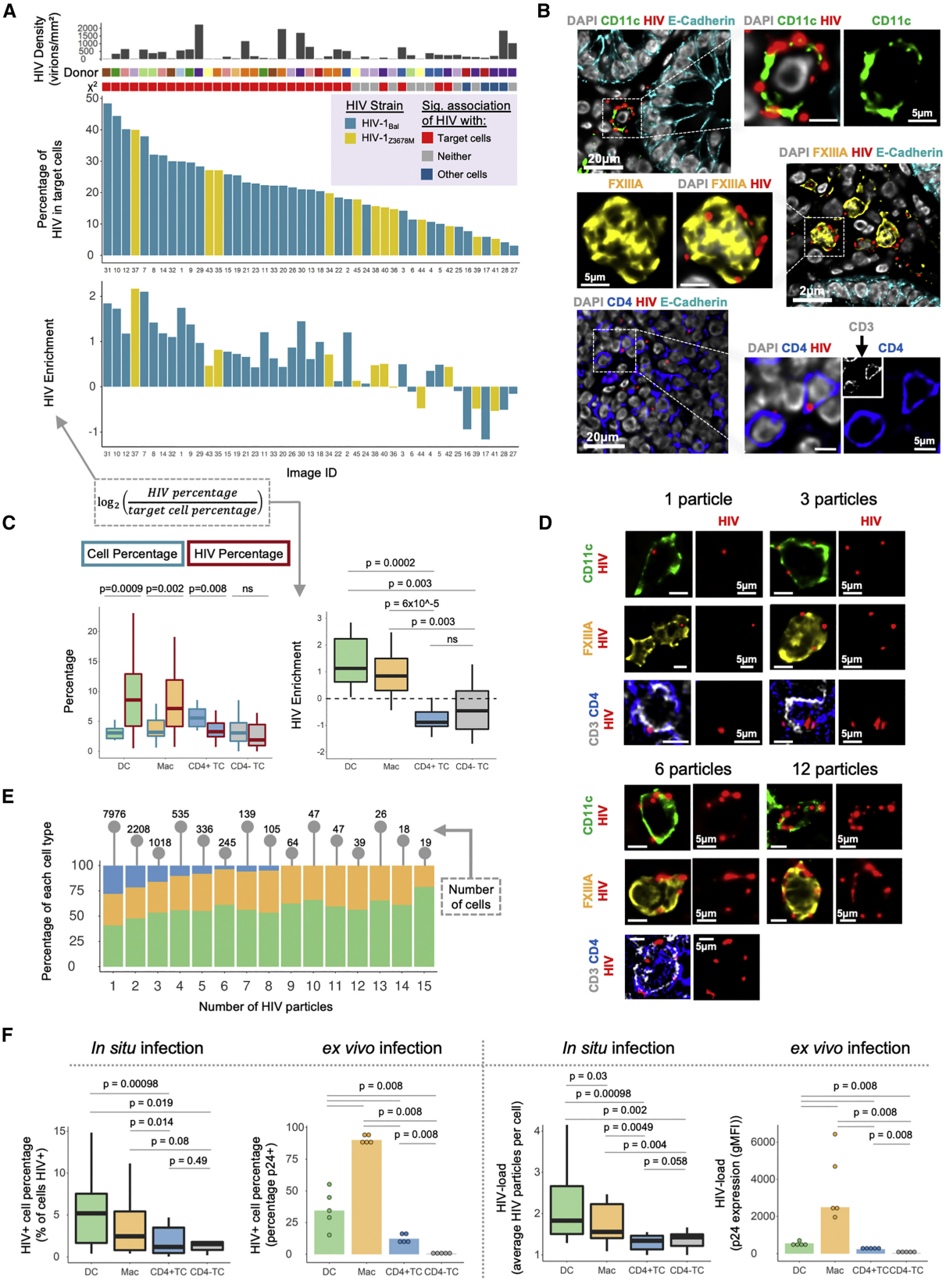


Figure 4.3: **Assessment of interactions of HIV with colorectal target cells.** (continued on next page)

Figure 4.3: **(A)** Target cell (DC, macrophage, CD4+ T cells) association with HIV across all images in this study. Middle: percentage of image HIV particles associated with target cells. Bottom: HIV enrichment in target cells (log2(HIV percentage/target cell percentage)). Top: whole image HIV particle density (per mm^2 of DAPI), donor number (color-coded) and chi-square test ($\chi 2$) of association indicating HIV enrichment in target cells (red), other cells (blue), or neither (gray); $p < 0.05$. HIV strain: blue bars, HIV_{Bal}; yellow bars, HIV_{Z3678M}. **(B)** Representative images of colorectal target cells interacting with HIV_{Bal} particles. **(C)** Left: cell type percentage (of all cells) versus percentage of all virions in cell type. Right: HIV enrichment as in **(A)**. Statistics: Wilcoxon signed-rank test. n = 15 explants (images average). **(D)** Representative images of target cells interacting with 1, 3, 6, or 12 HIV particles. **(E)** HIV+ target cells from all images were pooled and grouped by HIV load (1-15 particles). Composition of each cell type in each group is shown. Total cells in each group are annotated. **(F)** HIV uptake after 2 h HIV_{Bal} treatment of explants with analysis by microscopy (*in situ* infection) or *ex vivo* isolated cells analyzed by flow cytometry (*ex vivo* infection). Percentage of cells interacting with HIV ($\geq$1 virion or p24+) (left) or HIV load per cell (virion number or p24 gMFI) (right) is shown. *In situ* infection: n = 11 donors, Wilcoxon rank-sign test; *ex vivo* infection: n = 5 donors, Wilcoxon rank-sum test. n = 45 images from 15 explants (11 HIV_{Bal}, 4 HIV_{Z3678M}) from 12 donors for **(A-E)**.

that EP DCs had a 2-fold HIV enrichment compared with LP and LA DCs. In contrast, SM macrophages had a median 4-fold enrichment compared with their LP counterparts (Wilcoxon, $p < 0.05$, Supplementary Figure 4.3B). These compartment-specific differences in enrichment are explained by differences in cellular HIV load and the percentage of the population interacting with HIV. In particular, EP DCs had a higher HIV load per cell and a higher frequency of interactions with HIV, whereas SM macrophages had an increased frequency of HIV interactions, but a similar viral load to their LP counterparts (Wilcoxon, $p < 0.05$, Supplementary Figure 4.3C-D).

Beyond mapping HIV-target cell interactions to compartments we also explored how HIV+ cells were spatially distributed within compartments. Beginning with LP, we observed that, like uninfected tissue (Figure 4.2D), HIV+ DCs and CD4+ T cells were enriched near LAs (Supplementary Figure 4.3E). However, this enrichment was greater for HIV+ DCs and CD4+ T cells than their HIV- counterparts (paired two-sample t test, $p < 0.05$, Figure 4.4G), despite no difference in the frequency of viral particles between LA-proximal or -distal regions of the LP (Supplementary Figure 4.3F). Interestingly, despite our previous observation of enrichment of LP macrophages near EP and also EP DCs near LAs (Figure 4.2C-D), we did not observe any enrichment of HIV+ populations of these cells near these structures (data not shown).

In LAs themselves we measured HIV particle distribution using the formula for even-area concentric rings to divide these structures into roughly equal-area intervals (Supplementary Figure 4.3G). Using this approach, we observed a significant increase in HIV density toward the center of LAs (Supplementary Figure 4.3H). Stratifying LAs by size (Supplementary Figure 4.3I), HIV particles could even be detected toward the center of larger LAs (>500 µm in diameter), suggesting there may be a mechanism to focus virus centrally within these structures (Supplementary Figure 4.3J). Measuring HIV+ cells rather than individual particles, we observed that DCs were the only HIV-containing cell type to increase significantly in frequency toward the center of LAs (Figure 4.4H and Supplementary Figure 4.3K). This suggests that DC-mediated transport may in part

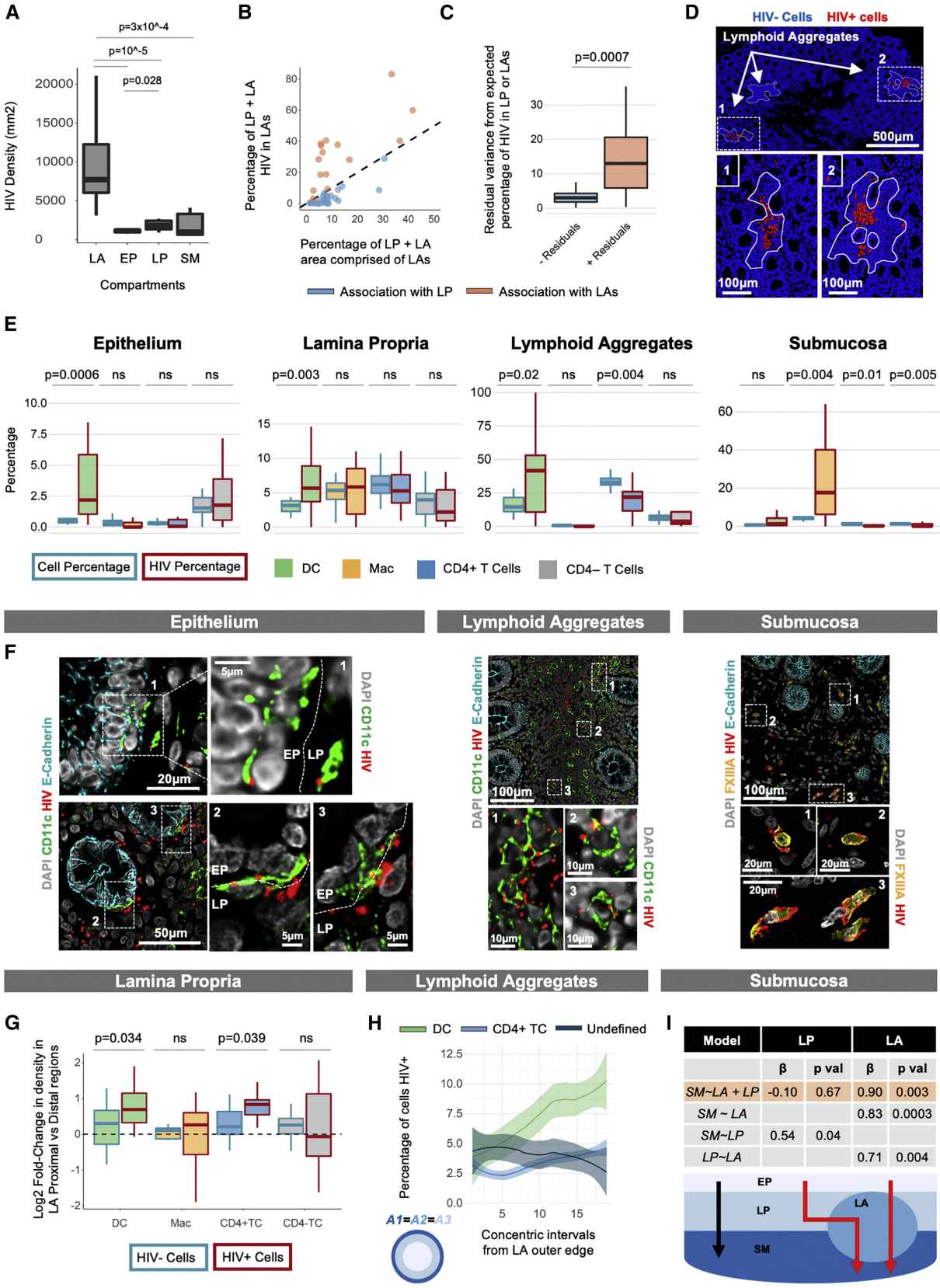


Figure 4.4: **Differential HIV uptake across colorectal tissue compartments.** (continued on next page)

Figure 4.4: **(A)** HIV density across tissue compartments. n = 10 images for each compartment (LA = 5, EP = 6, LP = 4, SM = 5 donors) (Methods). Statistics: Wilcoxon rank-sum test. **(B)** Percentage of virions in LAs versus percentage of image area (LP and LA only) comprised of LAs. Values above or below the line (y = x) indicate HIV enrichment in LAs or LP, respectively (Methods). n = 44 images (16 donors). **(C)** Positive and negative residuals (distance of points to line y = x) from **(B)**, which represent the magnitude of HIV association with LA and LP, respectively. Statistics: Wilcoxon rank-sum test. **(D)** Representative images showing preferential localization of HIV to LAs compared with the surrounding LP. Images are segmentation masks showing HIV+ cells (red), HIV- cells (blue), and outline of LAs. **(E)** Cell type percentage (of all cells) versus percentage of all virions across target cells in tissue compartments. n = 15 explants from 12 donors (images averaged). Statistics: Wilcoxon signed-rank test. **(F)** Representative images of DCs interacting with HIV in EP and LAs, and macrophage-HIV interactions in the SM. Broken line, EP-LP border. **(G)** Log2 fold change in HIV- or HIV+ cell density in LA-proximal (≤400 µm) versus -distal (400-800 µm) regions of LP. Paired regions only used if each contained at least 5 HIV+ and 5 HIV- cells. DC: 19 LAs (5 donors); Mac: 8 LAs (3 donors); CD4+ TC: 15 LAs (6 donors); CD4- TC: 12 LAs (5 donors) Statistics: paired t test (normality assessed by Q-Q plot). **(H)** Percentage of mucosal LA DCs, CD4+ T cells, or undefined cells (not a DC, macrophage, or T cell) that are HIV+ in intervals from the outer edge (x = 1) to center (x = 20) of HIV+ LAs (≥2 virions) (Methods). n = 112 LAs (14 donors). Statistics: Supplementary Figure 4.4K. LOESS curve of best fit. **(I)** Linear models of compartment HIV density to assess whether SM HIV density ("SM") is dependent on LP ("LP") or LA HIV density ("LA"). β weights and p values shown. Schematic of model results showing LA to SM HIV entry (red line) as the most likely pathway. n = 14 donors. Density = cells or virions per mm^2 of DAPI.

contribute to the central focusing of HIV within LAs.

Finally, as we were surprised that HIV interacted with macrophages in the deep SM layer as early as 2 h, we turned our attention to this compartment. We confirmed that HIV entry into the SM correlated with entry into the overlying mucosal layer (Pearson's r = 0.62, $p = 10^{-5}$, Supplementary Figure 4.3K). Fitting a linear model of HIV density in the SM as a function of LA and LP HIV density, we found that only LAs significantly predicted SM HIV entry ($\beta = 0.90$, $p = 0.003$), suggesting LAs as the dominant route of HIV entry into the SM (Figure 4.3I; Methods). Furthermore, we observed HIV particles throughout LAs, from the mucosal apical surface to the basal surface in the SM (Supplementary Figure 4.3M), suggesting that viral trafficking through the length of LAs is possible. Finally, we present evidence to confirm that SM entry was not due to leakage of virus from the cloning cylinder during the culture period. First, we observed a high correlation between mucosal and SM HIV densities (Supplementary Figure 4.3L), suggesting ordered entry of HIV into the SM from the mucosal layer. Second, HIV+ macrophages were in both superficial (near crypt bases) and deeper regions of the SM (Supplementary Figure 4.3N), whereas leakage would likely result in virus predominately in deeper regions, toward the bottom of the tissue. Finally, we collected the explant culture medium at the end of the culture period and confirmed that no HIV was present using a sensitive HIV detection assay (Supplementary Figure 4.3O).

Together these results reveal substantial differences in HIV distribution across colorectal tissue compartments, with LAs as key initial entry sites appearing to facilitate HIV access to the SM. Furthermore, HIV enrichment in DCs and macrophages occurred only in the mucosa and SM, respectively. This highlights that it is not only the cell type, but also its compartmental residence, that determines the degree of interaction with HIV.

4.2.5 Colorectal DCs form gradients toward HIV within and across tissue compartments

We next explored potential mechanisms of differential HIV-target cell interactions by analyzing the spatial organization of target cells in relation to HIV. In particular, we measured changes in target cell density from HIV particles, where steadily increasing or decreasing density gradients were used to infer potential cell migration in response to HIV (Figure 4.1F). In EP, LP, and LAs we observed that DCs formed an increasing density gradient toward HIV, whereas T cells and macrophages did not (Figure 4.5A). The effect was most pronounced in LP to the extent that DCs were depleted in regions >300 µm from HIV. Similarly, SM macrophages formed an increasing density gradient toward HIV.

As DCs exhibited migratory patterns toward HIV in all three mucosal compartments, we investigated whether they showed patterns consistent with crossing tissue compartments in response to HIV. To this end, we measured whether LP DCs redistributed "toward" and "into" EP or LAs when these compartments contained HIV. Starting with EP, we classified EP cells into HIV+ or HIV- populations and measured the density of DCs in the LP 10 µm from each EP population. Compared with HIV- EP, DCs were significantly more concentrated beneath (Figure 4.5B) and within HIV+ EP (Figure 4.5C). Using a similar approach for LAs, we found that, compared with HIV- LAs, LP DCs were significantly more enriched near HIV+ LAs (Figure 4.5D) and DCs were present at a higher density within HIV+ LAs (Figure 4.5E). In addition, DCs were more concentrated toward the outer edge of HIV+ LAs, further supporting the idea that DCs may migrate into these structures in response to HIV (Figure 4.5F). We did not observe any difference

in CD4+ T cell density near or within HIV+ versus HIV- LAs.

Taken together, these results suggest that HIV enrichment in mucosal DCs and SM macrophages may relate to their ability to migrate toward incoming HIV, with DCs going as far as to cross tissue compartments in response to incoming virus.

4.2.6 HIV induces the formation of target cell clusters within which DCs and macrophages traffic virus to CD4+ T cells

As cell:cell interactions are important for the spread of HIV [18, 21], we devised several spatial analysis tools to investigate the influence of HIV on interactions between target cells *in situ*.

We first used SpicyR [15], to compare cell:cell interactions between HIV+ and HIV- regions of treated explants to understand how HIV influences interactions, regardless of whether cells themselves are HIV+. This analysis showed that, in HIV+ regions, target cells cluster among each other in both LP and LAs. Interestingly, LP non-target CD4- T cells clustered among themselves in HIV+ regions but not with HIV target cells (Figure 4.6A). In addition, there was no difference between HIV- regions and the mock explants, confirming that clustering occurs specifically in response to HIV (Figure 4.6A). To verify that results are resistant to parameter variation, we ran SpicyR over a range of radius and distance cut offs (Supplementary Figure 4.4A). Qualitative inspection of images confirmed the presence of clusters of DCs, macrophages, and CD4+ T cells in HIV+ regions (Figure 4.6B). Interestingly, clusters tended to form away from the EP in the central area between the crypts of Lieberkühn. We hypothesized that this was due to target cell migration away from the EP interface in response to incoming HIV. To investigate this, we devised a method of temporal inference to model the progression of HIV entry into the mucosa (Methods). Here, images were divided into 100 μm^2 windows, each classified as "naive" (HIV density = 0), "early" (HIV density EP > LP), or "late" (HIV density EP < LP) in terms of HIV entry, with changes in target cell density measured close and far from the EP interface at each stage. Results showed that, like uninfected tissue, macrophages were enriched near the EP in naive regions; however, early in response to HIV they migrate deeper and persist there in the late phase (Supplementary Figure 4.4B). Interestingly we found that CD4+ T cells also migrate away from the EP, but only in the late phase, whereas CD4- T cells showed no change throughout the phases of entry. Curiously, DCs also showed no difference, although we can presume that this is due to their dual role of migration toward HIV+ EP (Figure 4.5B-C) as well as forming part of the central clusters in the LP (Figure 4.6A-B). These results suggest that a secondary effect of colorectal HIV entry is the rapid formation of a target cell-enriched community slightly distal to the site of entry, thus creating an ideal environment for cell-to-cell viral transfer.

We next investigated potential cell to cell HIV transfer. As HIV is present at the interface between cells during active viral transfer [22–26], cells engaged in transfer would both appear HIV+ in our dataset. Accordingly, we devised a "HIV-transfer phenotype score" to measure the change in association between cells when they both contain HIV. In brief, the frequency of HIV+ cell interactions with either HIV+ or HIV- cells was calculated using neighborhood analysis [27] and the transfer score defined as the difference in HIV+:HIV+ and HIV+:HIV- interaction frequencies. This was used as a proxy for viral transfer (Methods). Results showed that HIV+ DCs in LP and LAs, and macrophages in LP, had the highest transfer scores to CD4+ T cells (50% increase in images showing significant interactions) (Figure 4.6C, dotted boxes). Interestingly, we also observed

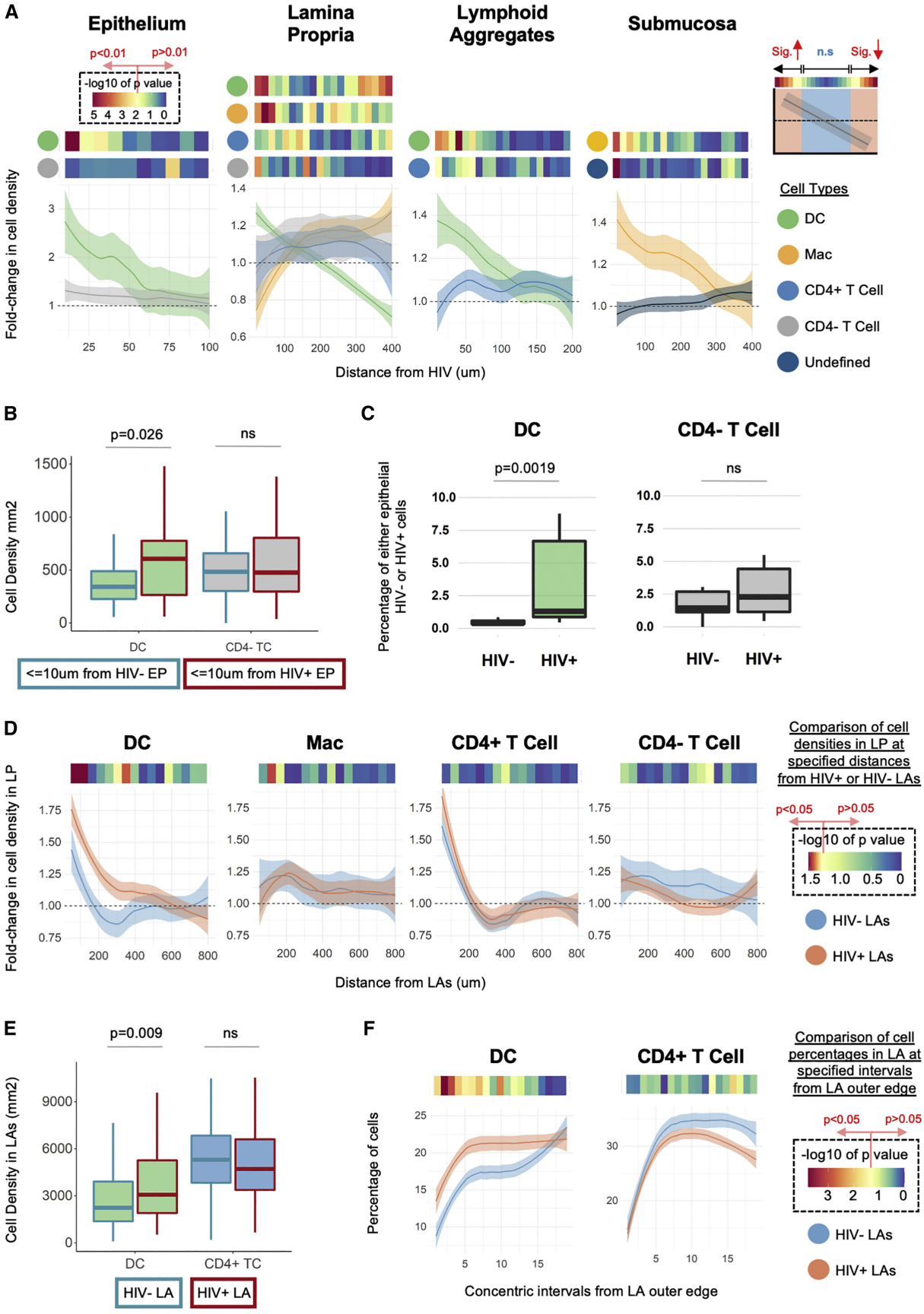


Figure 4.5: **Cellular gradients in response to HIV.** (continued on next page)

Figure 4.5: **(A)** Fold change in cell density (versus compartment average) in 10 µm (EP) or 20 µm (LP, LA, SM) intervals from HIV. n = 45 images (12 donors). Statistics: Wilcoxon signed-rank test comparing interval cell-type density with the compartment average. **(B)** LP DC and CD4- T cell density within 10 µm of HIV+ EP (≥1 virions) or HIV- EP (Methods). n = 40 images (12 donors). Statistics: Wilcoxon signed-rank test. **(C)** EP cells were classified as HIV+ or HIV- and the percentage of each population comprising DCs or CD4- T cells was measured. n = 33 images (11 donors). Statistics: Wilcoxon signed-rank test. **(D)** Fold change in cell density (versus LP average) in 50 µm intervals from HIV+ (≥2 virions) or HIV- (<2 virions) LAs. n = 38 images (12 donors). Statistics: Wilcoxon signed-rank test comparing cell density between intra-image HIV+ and HIV- LAs for each interval. **(E)** Density of DCs or CD4+ T cells in mucosal HIV+ versus HIV- LAs. Statistics: Wilcoxon rank-sum test comparing cell density between HIV- LAs (n = 84) and HIV+ LAs (n = 82) in 11 donors. **(F)** DC and CD4+ T cell frequency (percentage of all cells) in intervals of mucosal HIV+ versus HIV- LAs in (E), from their outer edge (x = 1) toward their center (x = 20). Statistics: Wilcoxon rank-sum test comparing cell frequencies in each interval between HIV+ and HIV- LAs. Density = cells per mm^2 of DAPI. LOESS curve of best fit for (A, D, and F). Heatmaps centered at $p = 0.01$ (A) or $p = 0.05$ (D and F).

considerable transfer scores between HIV+ DCs and macrophages. Importantly, transfer scores were close to 0 for CD4- T cells in LP, which are known to not transfer virus, as well as "unknown" cells in LAs. Potential transfer events were visually observable between DCs/macrophages and CD4+ T cells (Figure 4.6D and Supplementary Figure 4.4C) or between DCs and macrophages themselves (Supplementary Figure 4.4D) in both HIV_{BaL}- and HIV_{Z3678M}-treated explants. As transfer increases recipient cell HIV levels, we investigated whether CD4+ T cell interactions with DCs/macrophages was associated with increased T cell HIV load. Measuring cell body overlap as a proxy for an active interaction (Methods), we found that CD4+ T cell viral load was significantly positively associated with interactions with DCs, but not macrophages (Figure 4.6A). Importantly, DCs did not interact with CD4- T cells regardless of their (incidental) association with HIV (Supplementary Figure 4.4E), which confirms specificity and controls for the possibility of DC migration toward HIV (Figure 4.5A) driving the association. These results provide *in situ* quantitative evidence of cell-to-cell HIV transfer in intact tissue leading to increased CD4+ T cell HIV levels within just 2 h of exposure.

Having observed HIV-induced clustering among target cells and potential viral transfer *in situ* we determined whether this led to increased viral replication in CD4+ T cells *ex vivo*. We sorted all colorectal HIV target cells and co-cultured DCs or macrophages with autologous CD4+ T cells prior to infection with a transmitted founder HIV strain and subsequent assessment of CD4+ T cell infection 72 h later. We found that the presence of DCs or macrophages significantly enhanced infection of CD4+ T cells (Figure 4.6F). To determine whether enhancement was mediated by viral transfer we next infected DCs and macrophages prior to the addition of activated PBMC-derived CD4+ T cells. The results showed that both DCs and macrophages mediated HIV transfer to CD4+ T cells leading to increased viral replication but that this effect was significantly greater for DCs (Figure 4.6G).

All together, these results suggest that mucosal HIV entry induces its target cells to cluster together forming a community in which DCs and macrophages deliver virus to CD4+ T cells, facilitating infection of these cells within the mucosa itself.

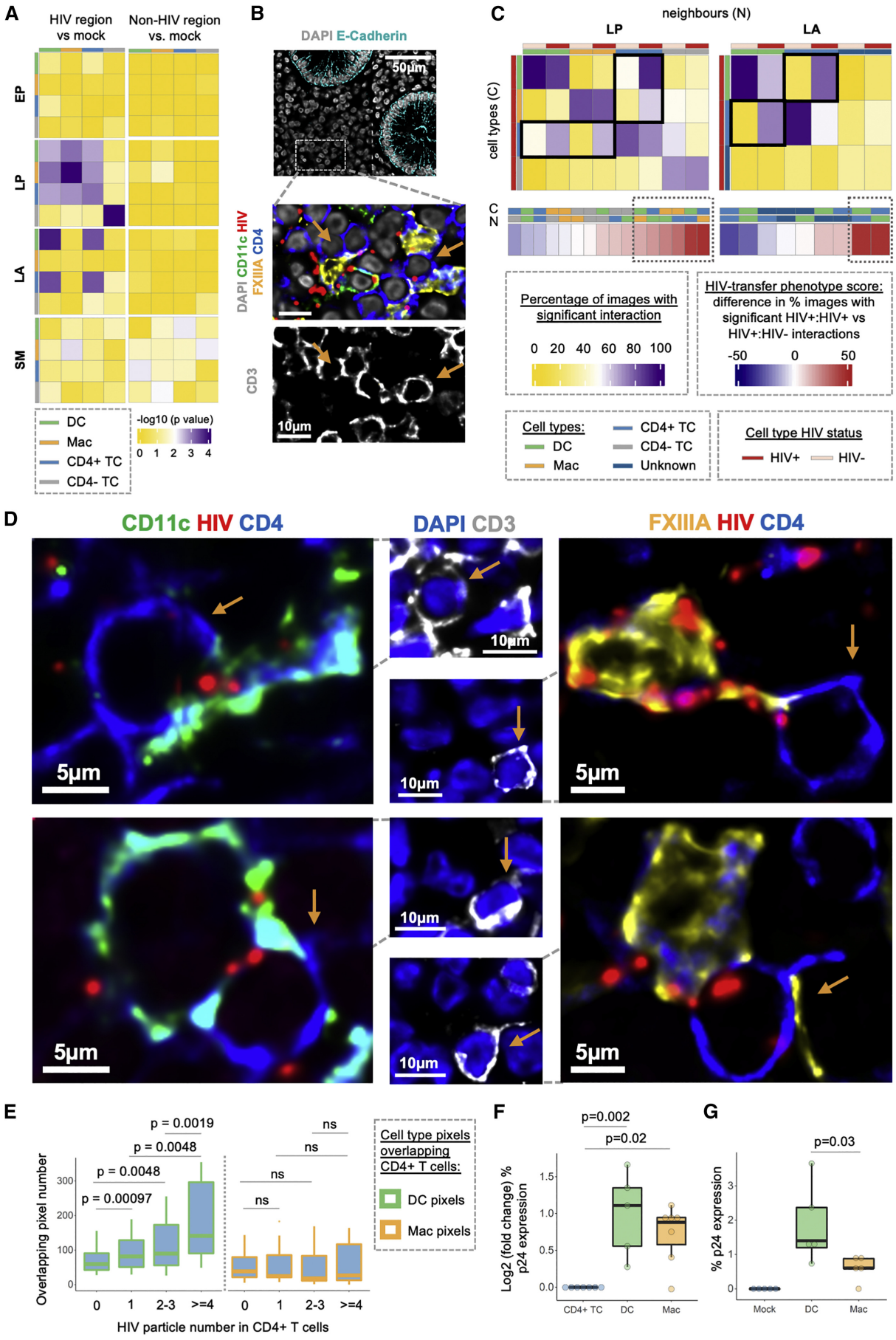


Figure 4.6: **Signatures of HIV-induced cell:cell interactions *in situ*.** (continued on next page)

Figure 4.6: **(A)** SpicyR analysis of differential cell:cell localization in HIV+ regions or HIV- regions versus mock-treated images in the EP, LP, LA, or SM (Methods). Purple indicates significant ($p < 0.01$) interactions. n = 45 images (12 donors). **(B)** Representative image of HIV target cells clustering (brown arrows) in an HIV+ region located away from the EP interface. **(C)** "HIV-transfer phenotype score" to estimate viral transfer between cells (Methods). Frequency of HIV+ target cell (rows) interactions with HIV- or HIV+ neighbors (columns) in LP or LAs. Colors: percentage of images with significant ($p < 0.005$) interactions (yellow-purple) or the HIV-transfer phenotype score (blue-red gradient), which is the difference in frequency between HIV+:HIV+ and HIV+:HIV- interactions. Key interactions encased by boxes. n = 45 images (12 donors). **(D)** Representative images of DCs and macrophages interacting with CD4+ T cells (brown arrow) where HIV_{Bal} is present at the interface between cells. **(E)** Number of DC (CD11c) or macrophage (FXIIIa) pixels overlapping with the body of CD4+ T cells that harbor varying levels of HIV particles. n = 11 images (5 donors). Statistics: Wilcoxon rank-sum test. **(F)** HIV co-culture assay. Primary colorectal CD4+ T cell p24 expression 72 h after HIV_{Z3678M} treatment (2 h, MOI = 1) of either CD4+ T cells alone (n = 7), or with added DCs (n = 5) or macrophages (n = 7) (10:1 ratio). Statistics: Wilcoxon rank-sum test comparing p24 fold change in co-cultures versus lone CD4+ T cell infections. **(G)** HIV transfer assay. PBMC-derived CD4+ T cell p24 expression 72 h after HIV_{Z3678M} or mock (PBS) treatment (2 h, MOI = 1) of primary colorectal DCs or macrophages with subsequent addition of CD4+ T cells. Statistics: Wilcoxon rank-sum test to compare.

4.3 Discussion

In this study we developed a pipeline for multiplexed *in situ* quantification of initial host-pathogen transmission events and applied it to study human colorectal HIV transmission within 2 h of exposure. This was made possible by a combination of RNAScope, CyCIF, and image analysis algorithms to enable accurate quantification. In particular, RNAscope overcame issues of low signal-to-noise inherent in antibody- and fluorophore-tagged HIV detection approaches [13, 28–30] and, when combined with CyCIF, enabled quantitative comparison of HIV localization to DCs, macrophages, CD4+ T cells, and CD4- T cells (as a control) in four colorectal tissue compartments. Our AFid tool reduces false identification of HIV+ cells due to autofluorescence [14]. Our segmentation approach allows cell overlap to outline full cell bodies. This enables accurate assignment of HIV particles to amorphous cells, such as DCs and macrophages (Supplementary Figure 4.1), while ensuring key cell:cell interactions are not missed as HIV can be transmitted between cells via membranous extensions [31, 32]. Finally, we employed spatial techniques to investigate cell:cell spread of HIV including our recently developed SpicyR algorithm [15], and a HIV-transfer phenotype score derived from neighborhood analysis [27]. Together these approaches enabled us to dissect early transmission events *in situ* and provides a framework for *in situ* quantification of cellular microenvironment changes that could be applied to a range of disease and physiological settings.

The relative involvement of different HIV target cells in initial viral uptake is fundamental to understanding the determinants of transmission. However, this has not been previously described in intact primary human colorectal tissue. We reveal that HIV is preferentially enriched in DCs and macrophages rather than CD4+ T cells, with DCs exhibiting the highest per-cell viral sampling capacity (Figure 4.3). Importantly, this was not replicable upon infection of *ex vivo* isolated rectal cells where macrophages were the

dominant initial target cell, in agreement with previous work [33]. We postulate that this difference is because the tissue microenvironment is a critical factor in HIV target cell migration and viral interactions, as illustrated in this study and others [9, 34–36]. This reinforces the importance of *in situ* studies to accurately define initial host-pathogen interactions.

A key strength of this study is the analysis of tissue compartments (Figures 4.4 and 5), which revealed that HIV enrichment in DCs and macrophages was specific to the mucosa and submucosa, respectively, and that HIV preferentially localizes to LAs. Mucosal DCs and submucosal macrophages steadily increased in density toward HIV particles, suggestive of migration to sites of HIV entry. LP DCs also appeared capable of crossing compartments to sample HIV, which has been observed for epithelium [34], but which we report here for LAs. As such, migration may be a key mechanism of early viral enrichment in these populations. This may be the function of one or more subsets of either DCs or macrophages, as subset-specific differences in both migratory capacity [37] and HIV binding [8, 11] have been observed. HIV binding itself appeared to influence cell location with LP HIV+ DCs and CD4+ T cells tending to locate near LAs. This could represent specific cell subsets in LA-adjacent regions, which are known to vary in murine studies of Peyer's patches [38] but are unstudied in humans. Another possibility, at least for DCs, is that HIV binding induces enhanced LA-directed migration and entry. This is supported by the increased DC density in HIV+ versus HIV- LAs and the preponderance of HIV+ DCs in these structures. Compartment-based analysis also revealed that LAs are key HIV-containing compartments within 2 h of exposure to the virus. This may be due to delivery by HIV+ DCs or other cells in the adjacent LP, or via follicle-associated epithelium and resident M cells, which is a key site of entry for other enteric pathogens [39]. As LAs are enriched in rectal tissue [40], contain abundant HIV target cells and are known sites of HIV persistence [41], rapid access to these structures may facilitate sustained infection and early reservoir formation within the mucosa. Indeed, naive CD4+ T cells are enriched in LAs [42] and are amenable to both viral transfer and activation by HIV+ DCs, which can promote latency [43] and HIV-susceptible Th17 programs [7, 44, 45]. Entry into LAs but not LP was also associated with increased SM HIV levels, suggesting LAs as a possible conduit for HIV access to the underlying SM. This could be through direct passage as LAs traverse the muscularis mucosa barrier [42] and HIV easily penetrates deep into LAs (Supplementary Figure 4.3H-J). Alternatively, as HIV disseminates via lymphatics [2], it may use the extensive LA lymphatic network [42] to gain access to the SM, through which mucosal lymphatics drain [46].

CD4+ T cells are the major initial targets of HIV integration and productive infection [6, 7, 47]. However, the early events leading to infection remain poorly understood. DCs and macrophages facilitate HIV transfer and enhanced CD4+ T cell infection *in vitro*; however, *in situ* characterization using human tissue is lacking and mostly qualitative [1]. Using our SpicyR algorithm and temporal inference we show that, after HIV penetrates the EP, HIV target cells appear to move away from the EP and form clusters consisting of DCs, macrophages, and CD4+ T cells (Figure 4.6). This may occur temporally as our data suggest that macrophages migrate away from the EP initially, followed by CD4+ T cells. Importantly, this was specific to HIV target cells and creates an ideal setting for viral transfer between cell types. Indeed, using a scoring system for potential HIV transfer events, we showed that, upon HIV binding, DCs and macrophages preferentially cluster with CD4+ T cells, with virus present at the cellular interface. Interestingly, CD4+ T cell HIV load was tightly correlated with their physical overlap with DCs, suggesting that the

early presence of high viral loads in CD4+ T cells may be dependent on their interactions with DCs. The lack of this association for macrophages may be due to lower levels of viral transfer, or a lack of physical overlap during interactions with T cells, unlike DCs which can entirely envelope CD4+ T cells during transfer [48]. Finally, we assessed infection of rectal tissue-derived co-cultures and confirmed viral transfer and enhanced infection of CD4+ T cells by both DCs and macrophages, in agreement with other studies [8, 11, 49]. This provides the strongest evidence to date that viral transfer to CD4+ T cells occurs as early as 2 h within the mucosa.

In summary, this study contributes quantitative *in situ* data on the initial events of HIV transmission in intact human mucosal explant tissue. Although some of the findings of this study are circumstantial, we believe they give rise to important hypotheses regarding HIV transmission. Particularly, the role of niche-specific cell subsets, the drivers of cell recruitment and cell:cell interactions at sites of HIV entry, and the possibility of LAs as key sites of early viral amplification, persistence, and extra-mucosal dissemination. We anticipate that recent advancements in high-parameter imaging modalities [50, 51], animal models and human organoid systems [52] will allow unprecedented insight regarding the early determinants of HIV transmission. The approach and results presented here provide a foundation for such future studies, which could inform prophylactic interventions or the design of a mucosal vaccine.

4.3.1 Limitations of the study

Cells were defined as HIV+ if they spatially overlapped with HIV particles, meaning that some will be false positive due to random encounter. We worked around this limitation with comparative rather than absolute measurements such as HIV enrichment (HIV+ cell frequency/cell frequency). Precise identification of virion-cell interactions would require a sensitive reporter construct that produces a signal following the virion-membrane binding event. However, this would be challenging as the virus rapidly contacts multiple cell types in tissue and so downstream virion-cell interactions would produce false negatives.

Increasing/decreasing cell density gradients were used to infer cell migration to and from areas of tissue. Formal proof would require real-time imaging of virions and cells in fresh human explant tissue, which is technically challenging if not impossible with current technologies.

The goal of our study was to map out the early host-cell interactions that characterize colorectal HIV transmission. However, our system is unable to determine the relative contribution of these early interactions to downstream mucosal infection. Future studies may address this by combining pro-longed explant culture, rapid targeted inhibition of cell subsets, and a reporter for downstream productive infection [7]. However, studying the role of newly migrated blood-derived immune cells in mucosal infection would likely require small animal models [53].

4.4 Methods

4.4.1 Resource availability

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Andrew Harman (andrew.harman@sydney.edu.au).

Materials

This study did not generate new unique reagents.

4.4.2 Experimental model and subject details

Human subjects

Healthy human colorectal tissue was obtained within 15 min of resection from patients undergoing surgical intervention for diverticulitis or colorectal cancer. Only healthy tissue distal to the site of disease process were used for this study. Details on patient age and sex are provided in Supplementary Table 1. This study was approved by the Western Sydney Local Area Health District (WSLHD) Human Research Ethics Committee (HREC); reference number HREC/2013/8/4.4(3777) AU RED HREC/13/WMEAD/232. Written consent was obtained from all donors.

Cell lines

Both human embryonic kidney-derived 293T (HEK293T Cells) and HeLa-derived TZM-bls were cultured in Dulbecco's Modified Eagle Medium (Lonza) with 10% Fetal Calf Serum (FCS) (Lonza) (DMEM10) at 37°C/5% CO2 and passaged using TrypLE express (Gibco) dissociation at a 1:10 dilution three times a week and 1:12 dilution twice a week, respectively. SUPT1.CCR5-CL.30 cells were maintained in RPMI (Lonza) with 10% FCS (RF10) at 37°C/5% CO2 and passaged at a 1:10 dilution twice per week.

This page has been left blank intentionally.

Table 4.1: Key resources table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Sheep Polyclonal anti-FXIIIA	Affinity Biologicals	Cat: SAF13A-AP
Rabbit monoclonal anti-CD11c (clone EP1347Y)	Abcam	Cat# ab52632; RRID: AB_2129793
Mouse monoclonal anti-CD3 (clone F7.2.38)	Abcam	Cat# ab181724; RRID: AB_302587
Rabbit anti-CD4 (clone EPR6855)	Abcam	Cat# ab181724; RRID: AB_2864377
Rabbit anti E-Cadherin-AF647 (clone 24E10)	Cell Signal	Cat# 9835; RRID: AB_10828228
Donkey anti-sheep-AF488	Invitrogen	Cat: A-11015
Donkey anti-rabbit-AF647	Invitrogen	Cat: A-31573
Donkey anti-mouse-DyLight755	Invitrogen	Cat: SA5-10171
HLA-DR BUV395 (L243)	Biolegend	N/A
HLA-DR PerCP (AC122)	Miltenyi Biotec	Cat# 130-108-056; RRID: AB_2661330
CD19 BV750 (HI819)	Biolegend	N/A
CD19 APC Vio770 (L719)	Miltenyi Biotec	N/A
CD4 BV785 (OKT4)	Biolegend	Cat# 317442; RRID: AB_2563242
CD4 BV650 (OKT4)	Biolegend	Cat# 317436; RRID: AB_2563050

Table 4.1 continued from previous page

CD3 BUV496 (UCHT1)	BD	Cat# 612940; RRID: AB_2870222
CD3 APC Vio770 (REA613)	Miltenyi Biotec	Cat# 130-113-698; RRID: AB_2726239
CD14 BUV737 (M5E2)	BD	Cat# 612763; RRID: AB_2870094
CD14 BV421 (M5E2)	Biolegend	BioLegend Cat# 301830; RRID: AB_10959324
CD11c BB515 (B-ly6)	BD	Cat# 564490; RRID: AB_2744273
CD11c PE CF594 (B-ly6)	BD	Cat# 562393; RRID: AB_11153662
p24-PE (KC57)	Beckman Coulter	Cat# 6604667; RRID: AB_1575989
p24-APC (28b7)	Medimabs	Cat: MM-0289-APC
Bacterial and virus strains		
HIV- BaL	HEK293T transfection with pWT/HIV_{BaL}	N/A
HIV- Z3678M	HEK293T transfection with pHIV$HIV_{Z3678MTF}$	N/A
Biological samples		
Human colorectal tissue	Department of Colorectal Surgery Westmead Hospital	N/A
Critical commercial assays		
RNAscope 2.5HD Reagent Kit-RED	ACD Bio	Cat: 322360
sulfo-Cyanine7 antibody labelling kit	Lumiprobe	Cat: 5321-10rxn
Dead Cell Removal Kit	Miltenyi Biotec	Cat: 130-090-101
EasySep Human CD4+ T Cell Enrichment Kit	StemCell Technologies, Vancouver, Canada	Cat: 19052

Table 4.1 continued from previous page

Human CD45 Microbead Enrichment Kit	Miltenyi Biotec	Cat# 130-045-801; RRID: AB_2783001
Human CD14 Microbead Enrichment Kit	Miltenyi Biotec	Cat# 130-050-201; RRID: AB_2665482
Human CD19 Microbead Enrichment Kit	Miltenyi Biotec	Cat# 130-050-301; RRID: AB_2848166
Deposited data		
All data to reproduce figures in this paper	This study	https://doi.org/10.5281/zenodo.6992156
Experimental models: Cell lines		
HEK293T Cells	ATCC	N/A
HeLa-derived TZM-bls	NIH AIDS Reagent Program	John Kappes and Xiaoyun Wu
SUPT1.CCR5-CL.30 cells	Human Non-Hodgkin's T lymphocyte Lymphoma	James Hoxie, University of PA
Oligonucleotides		
custom probes targeting HIV-1_{Bal} (85zz pairs spanning base pairs 1144-8431)	ACD Bio	REF: 486631
custom probes targeting HIV_{Z3678M} (85zz pairs spanning base pairs 1149 - 8505)	ACD Bio	REF: 811791
Recombinant DNA		
pWT/HIV_{BaL}	NIH AIDS Research and Reference Reagent Program	Dr. Bryan R. Cullen
$pHIVHIV_{Z3678MTF}$	Gift from Eric Hunter	Genbank: KR820393
Software and algorithms		
Image processing: Huygens Professional 18.10	Scientific Volume Imaging	N/A
Image Processing: Fiji (Madison Version)	ImageJ	N/A
Image processing: Autofluorescence Identifier 'AFid' software	(Baharlou et al., 2020)	https://ellispatrick.github.io/AFid
Data Extraction: MATLAB 2017b	Mathworks	N/A

Table 4.1 continued from previous page

Analysis code to reproduce figures in this paper	This study: https://doi.org/10.5281/zenodo.6992156	https://github.com/heevaBaharlou/HIVImageAnalysis
SpicyR package	Bioconductor	10.18129/B9.bioc.spicyR
FlowJo (Treestar)	FlowJo (Treestar).	N/A
R programming language	GNU	N/A
Other		
Schematics were created with Biorender	Biorender	Agreement number: HX23I22GBL

4.4.3 Method details

HIV-1 virus production

Lab-adapted (HIV_{BaL}) or transmitted founder (HIV_{Z3678M}) strains were produced by transfection using a previously described protocol [8]. 1.6×10^7 HEK293T cells were seeded in a T150 flask (Becton Dickinson, Franklin Lakes, New Jersey, USA) and transfected with 80ug of $pWT/_{BaL}$ or $pHIVZ_{Z3678MTF}$ plasmid DNA. The following reaction mixture (all from Sigma-Aldrich) was prepared separately and added in addition to the plasmid DNA: 10µL 0.15M Na2HPO4 (pH 7.1), 128uL 2M CaCl2, 1mL Hepes-buffered saline (280mM NaCl, 50mM HEPES, pH 7.1), 1mL (1mM Tris, 0.1mM EDTA, pH 8.0), all diluted in 15mL DMEM10 (Lonza). Culture media was replaced the next day with DMEM10 and cells cultured for a further 2 days, after which media was collected, centrifuged (1600g, 20min) and the resultant supernatant concentrated (3000g, 20min) to 1mL using 300kDa filters (Vivaspin 20, Sartorius, Göttingen, Germany). High titer stocks for HIV_{BaL} were achieved by infection of the SUPT1 T cell line. Stocks were depleted of macrovesicles by adding 18mL viral supernatant to 2mL CD45 magnetic beads (Miltenyi Biotech) for 2h prior to filtering through an LS column (Miltenyi Biotech). The resultant CD45- HIV_{BaL} supernatant as well as HIV_{Z3678M} were under-layed with 1mL of 20% sucrose and ultracentrifuged (100,000g, 4°C, 1.5h, Beckman Optima XL-100K, 70Ti rotor) to further concentrate viral stocks. This method yielded viral titers between 108-109, as measured by 50% tissue culture infective dose (TCID50)/mL on TZM-bl cells by LTR β-galactosidase reporter gene expression following one round of infection. Briefly, serial dilutions of viral stocks were performed on plated TZM-bls (37°C, 3 days), followed by media removal, addition of 50ul X-gal solution and incubation for 1h at 37°C. Wells were then diluted with 50ul of 4% PFA and incubated for a further 20min at room temperature (RT), followed by solution removal and EliSPOT imaging. The Spearman Kärber algorithm was used for TCID50 measurements. Virus aliquots were stored at 80°C. Endotoxins (Limulus amebocyte lysate assay; Sigma), TNF-α, IFN-α, and IFN-β (Enzyme-linked immunosorbent assay (ELISA) were all below the limit of detection.

Tissue digestion

To perform HIV-uptake (Figure 4.3F) and co-culture/transfer assays (Figure 4.6F-G), we extracted HIV target cells from human colorectal tissue. Underlying fat and mesentery were removed using a scalpel and forceps and remaining tissue cut into $5mm^2$ pieces. Surface epithelium and mucus was stripped by two incubations in RPMI with 10% FCS, 0.3% DTT (Sigma) and 2mM EDTA (Sigma) (15min, 37°C). Tissue was washed in DPBS and underwent two incubations in 20mL RPMI with 0.3% Collagenase IV (Worthington) and 0.5% DNase (Sigma) (30min, 37°C) to liberate cells which were then passed through a 100µm cell strainer and washed twice in DPBS. Cells were resuspended in 35mL RPMI, under-layed with 15mL Ficoll-Paque (GE Healthcare) and centrifuged (400g, 20min, no brakes). Buffy coats were collected and washed twice in DPBS. Red Cell Lysis buffer (All Sigma: 150 mM ammonium chloride (v/v), 10mM potassium bicarbonate (v/v), 0.1 mM EDTA (v/v) in ddH20) was used to remove remaining red blood cells as per the manufacturer's instructions.

HIV uptake assay

Following our tissue digestion protocol (see above), liberated cells underwent positive selection for CD45+ cells (Miltenyi Biotec). 5×10^5 cells in 150ul of DC Culture Media were then treated with HIV_{Bal} (MOI = 5, 2h, 37°C) or PBS (mock). Cells were then washed three times in DPBS, 200ul of DPBS, stained with 0.05ul FVS700 for 30min at 4°C, washed in FACS wash (1% FCS (v/v), 2 mM EDTA, 0.1% sodium azide (w/v) in PBS) and 10ul Brilliant Stain Buffer (BD) added. Cells were then incubated with an antibody panel for 30min at 4°C with the final volume made to 50ul. The antibody panel included Biolegend: 1ul HLA-DR BUV395 (L243), 1ul CD19 BV750 (HI819), 2.5ul CD4 BV786 (OKT4); BD: 5ul CD3 BUV496 (UCHT1), 2.5ul CD14 BUV737 (M5E2), 1.5ul CD11c BB515 (B-ly6). Cells were washed in FACS wash, permeabilised with 100ul Cytofix/Cytoperm (BD) for 20min at RT and washed in Perm Wash (1% FCS (v/v), 1% BSA (w/v), 0.1% saponin (w/v), 0.1% sodium azide (w/v) in PBS). Cells were resuspended in 50ul Perm Wash and underwent intracellular staining with antibodies Beckman Coulter: p24-PE (KC57) and Medimabs: p24-APC (28b7) for 30min at RT. Cells were again washed in Perm Wash and HIV expression assessed by dual p24 expression using an LSRFortessa (BD). P24 expression was measured by gMFI and dual p24+ positive cells were defined as HIV+ cells as per [8, 11]. Mock treated cells were used to set the gates to minimize background staining interference.

Target cell selection and sorting

Following tissue digestion (see above), liberated cells were positively selected for CD45+ cells as per the manufacturer's instructions (EasySep Human CD45+ Cell Enrichment Kit, StemCell Technologies) using a QuadroMACS separator with LS columns. 2.5×10^6 cells were resuspended in 200ul of DPBS, stained with 0.05ul FVS700 for 30min at 4°C, washed in FACS wash and 10ul Brilliant Stain Buffer added. Our antibody sort panel was added for 30min at 4°C with the final volume made to 50ul. The sort panel included Miltenyi: 2.5ul CD3 APC Vio770 (REA613), 2.5ul CD19 APC Vio770 (L719), 1ul HLA-DR PerCP (AC122); Biolegend: 2ul CD4 BV650 (OKT4); BD: 1.5ul CD11c PE CF594 (B-ly6), 2.5ul CD14 BV421 (M5E2). Cells were then washed twice in FACS wash and once in pre-sort buffer (BD). 1mL of pre-sort buffer was used to resuspend cells which were then filtered using a 100µm cell strainer just prior to sorting on either the BDInflux (BD) or BDAriaIII (BD) cell sorters. CD4+ T cells were defined as live CD3+CD4+, Dendritic Cells as live HLA-DR+CD3-CD19-CD14-CD11c+ and Macrophages as live HLA-DR+CD3-CD19-CD14+. Sorted cells were placed in FACS tubes with 500ul of DC Culture Media and kept at 4°C until co-culture/transfer assay setup as described in the sections below.

HIV co-culture assay

Sorted DCs, macrophages and CD4+ T cells were plated as follows: CD4+ T cells alone, DCs with CD4+ T cells (1:10 ratio), macrophages with CD4+ T Cells (1:10 ratio). Cultures were topped to 150ul with DC Culture Media and treated with HIV_{Z3678M} (MOI = 1, 2h, 37°C). An additional mock treated CD4+ T cell culture was maintained as a control. Cells were then washed three times in DC Culture Media, resuspended in 200ul of DC Culture Media with 0.02% Normocin (InvivoGen) and cultured for 3 days at 37°C. HIV infection of CD4+ T cells was determined by p24 expression using flow cytometry

as described in the section on 'HIV assessment by Flow Cytometry'.

HIV transfer assay

Sorted DCs and macrophages as well as activated PBMC-derived CD4+ T cells (see section below) were plated in individual wells in 150ul DC Culture media, then exposed to HIV_{Z3678M} (MOI = 1, 2h, 37°C). An additional mock treated activated CD4+ T cell culture was maintained as a control. Cells were then washed three times in DC Culture Media and resuspended in 200ul of DC Culture Media with 0.02% Normocin. Activated CD4+ T cells were then added to DC and Macrophage cultures at a 2:1 ratio and cultured for 3 days at 37°C. HIV infection of CD4+ T cells was determined by p24 expression using flow cytometry as described in the section on 'HIV assessment by Flow Cytometry'.

HIV assessment by Flow Cytometry

Cells cultured with HIV were washed in DPBS, resuspended in 200ul, stained with FVS700 for 20min at 4°C and washed with FACS wash. Cells were stained with Miltenyi: CD3 APC-Vio770 (REA613) for 30min at 4°C, washed twice in FACS wash, permeabilized with 100ul Cytofix/Cytoperm for 20min at RT and washed in Perm Wash. Cells were resuspended in 50ul Perm Wash and underwent intracellular staining with antibodies Beckman Coulter: 1ul p24-PE (KC57) and Medimabs: 1ul p24-APC (28b7) for 30min at RT. Cells were again washed in Perm Wash and HIV expression assessed by dual p24 expression using an LSRFortessa.

PBMC-derived CD4+ T cell isolation

Transfer assays (Figure 4.6G) were performed using allogenic activated CD4+ T cells selected from peripheral blood mononuclear cells (PBMCs). PBMCs were derived from leukoreduction system chambers (LRSC) (Australian Red Cross Blood Service), on the same day as platelet donation. LRSCs were diluted 1:5, distributed across Falcon tubes with 35mL in each tube, then under-layed with 15mL Ficoll-Paque and centrifuged (400g, 20min, no brakes). Buffy coats were collected and washed x2 in DPBS. Red Cell Lysis buffer was used to remove remaining red blood cells as per the manufacturer's instructions. CD4+ T cells were selected using a CD4 selection kit (StemCell Technologies) and activated by culturing 1×10^6 cells/mL for 3 days at 37°C in RPMI supplemented with 10% FCS, 5 mg/mL PHA (Sigma) and 150 IU/mL IL-2 (Peprotech). Cells were transferred to cryovials containing FCS with 10% DMSO, placed in a CoolCell (Corning) and stored at 80°C.

HIV explant infection

Human colorectal tissue was obtained within 15 min of surgical resection. Underlying fat and mesentery were removed using a scalpel and forceps and tissue spread out in a Petri dish with the mucosal surface face-up. Gel-foam sponges (Pfizer) were cut into $1cm^2$ pieces (one for each explant), placed in the well of a 24-well plate and soaked in culture media consisting of 10µM HEPES (Gibco), non-essential amino acids (Gibco), 1 mM sodium pyruvate (Gibco), 50µM 2-Mercaptoethanol (Gibco), 10ug/mL Gentamycin (Gibco), 10% FCS (Lonza), all diluted in RPMI-1640 (Lonza). Here after this is referred to as 'DC Culture Media'. Sponges were left to soak whilst tissue was further processed. As

previously described [42, 54] a dissection light microscope was used to select appropriate placement of cloning cylinders for the topical application of HIV. Appropriate areas were defined as containing visible lymphoid aggregates and being free of any signs of trauma. Once an appropriate area was located, an 8mm cloning cylinders (Sigma-Aldrich) was lightly coated on one side with histoacryl surgical glue (B Braun) using a fine paint brush and carefully placed over the region. 500ul of PBS was then added to the cloning cylinder to prevent the tissue surface from drying out and to hasten setting of the surgical glue. This step is critical as excess glue can slowly spread over time from the cylinder edge and cover the tissue surface thus preventing viral entry. Once all cylinders were placed, tissue within cylinders were checked again to ensure they were free of glue. 1 or 2 selected regions did not have cylinders placed but were resected with a scalpel as a $1cm^2$ area and fixed in 4% PFA (diluted in PBS) (electron microscopy sciences) for 18h. These samples were used to analyze target cell distribution in fresh uninfected colorectal tissue (Figure 4.2). The Petri dish was then filled with PBS so as to just cover its surface, which reduced friction and the likelihood of displacement of the cylinders during cutting and lifting of explants. Soaked gel-foam sponges were then distributed across the 24-well plate. The perimeter surrounding each cloning cylinder was then cut, after which forceps were used to lift explants and place them on the sponges. Wells were then filled with DC Culture Media to the level of the tissue. A final quality check for cloning cylinder sealing was performed by ensuring that the previously applied PBS was still present and at the same level across explants. Solutions were removed from all cylinders and either 100ul PBS (mock), or a $TCID_{50}$ of 70,000 (diluted in 100ul PBS) of lab adapted (HIV_{BaL}) or transmitted founder (HIV_{Z3678M}) strains applied to the inner chamber of cloning cylinders for 2h at 37°C. The time point of 2h was chosen as it allowed sufficient viral penetration into the tissue and was short enough that the explants did not start to degrade. Indeed, initial optimizations with 30min and 6h timepoints showed insufficient viral penetration past the epithelial surface and early signs of tissue degradation respectively (data not shown). It should be noted that degradation as early as 6h was likely due to the large size ($1cm^2$) of our explants. A $TCID_{50}$ of 70,000 was selected as it is comparable to peak levels of infectious virus found in semen during the acute stage [55]. Supernatant from inner chamber, as well as the DC Culture Media below were collected and stored at 80°C for assessment of HIV leakage from cylinders during the culture period (Supplementary Figure 4.3O). Explants were then washed x3 with 500ul PBS to remove excess HIV from the mucosal surface prior to cylinder removal and tissue fixation in 4% PFA (diluted in PBS) for 18h.

Tissue embedding and sectioning

After fixation, explants were removed and trimmed with a scalpel along the impression left by the cloning cylinder during the culture period. This ensured that only areas exposed to virus remained. Explants were kept in 70% ethanol prior to embedding in paraffin. Tissues were then embedded flat so that the mucosal surface faced away from the sectioning surface of the paraffin block. This ensured that the thin mucosal surface would not accidentally be trimmed away whilst sectioning. 4µm transverse sections were taken through the entire tissue, from the submucosal side to the end of mucosa, and placed on Superfrost Ultra Plus Adhesion Slides (Thermo Scientific) Sections were stored in the dark at RT until staining.

RNAScope

Detection of HIV RNA was performed using RNAscope as previously described [8, 11, 13] using the RNAscope 2.5HD Reagent Kit-RED with custom probes targeting HIV-1_{Bal} or $HIV_{Z3678M.}$ In the protocol that follows, reagents included in the 'RNAscope 2.5HD Reagent Kit-RED' are indicated. Unless otherwise written, wash steps were for 2min on a rotator set to low and incubations were carried out in a hybridization oven (HybEZ Hybridization System (220VAC), ACD Bio). 4µm paraffin sections were baked at 60°C for 1h and dewaxed by sequentially submerging slides in xylene (2 × 2min) and 100% ethanol (2 × 2min). Slides were air-dried and antigen retrieval performed for 20min at 95°C using a pH9 buffer (RNAscope kit) and a decloaking chamber (Biocare). Sections were washed in TBS (Amresco, Cat: 0788), then Milli-Q H2O, followed by dipping slides 3-5 times in 100% ethanol, leaving them to air dry and then encircling sections with a hydrophobic pen (RNAscope kit). Sections were then incubated with protease pre-treatment 3 (diluted 1:5 in PBS and kept ice-cold) (RNAscope kit) for 30min at 40°C. Sections were washed ×2 in Milli-Q H2O and incubated with probes targeting HIV-1Bal or HIVZ3678M for 2h at 40°C. Sections were washed ×2 in RNAscope wash buffer (RNAscope kit). Signal from probes was then amplified using Amps 1-6 (RNAscope kit) which were added in sequence with ×2 washes in RNAscope wash buffer between Amps. Amp incubation times were as follows Amp 1 = 30min at 40°C; Amp 2 = 15min at 40°C; Amp 3 = 30min at 40°C; Amp 4 = 15min at 40°C; Amp 5 = 30min at RT; Amp 6 = 15min at RT. HIV RNA Signal was developed using Fast Red substrate made by mixing Red-B and Red-A (RNAscope kit) at a 1:75 ratio for 5min at RT, followed by washing in Milli-Q H_2O then TBS.

Cyclic immunofluorescence staining and image acquisition

Unless otherwise indicated all washes were 2 × 2min in TBS on rotator set to low and incubations were in a humidified chamber protected from light. Following RNAscope, sections were blocked for 30min at RT with blocking buffer (10% donkey serum (Sigma), 1% BSA (Sigma), 0.1% Saponin (Sigma), all diluted in TBS) and washed. Sections were incubated with sheep anti-FXIIIa and rabbit anti-CD11c antibodies (diluted in block buffer) overnight at 4°C, washed and donkey anti-sheep-AF488 (Invitrogen, Cat: A-11015) and donkey anti-rabbit-AF647 (Invitrogen, Cat: A-31573) antibodies added for 30min at RT. Sections were washed and further blocked for 30min at RT with block buffer with 10% rabbit serum (DAKO) to block excess binding sites from the donkey anti-rabbit antibodies. 0.5% PFA (diluted in PBS) was added for 15min at RT to fix blocking rabbit IgGs in place. Rabbit anti-CD4-Cy7 (Abcam, Cat: ab181724, conjugation with sulfo-Cyanine7 antibody labeling kit (lumiprobe, Cat: 5321-10rxn)) was then added overnight at RT and sections washed 3 × 5min in TBS. Sections were stained with 1ug/mL DAPI (diluted in TBS) (Thermo Scientific, Cat: 62248) for 3min at RT, washed then rinsed in Milli-Q water. Sections were mounted with SlowFade Diamond Antifade Mountant (Invitrogen, Cat: S36963) and cover-slipped (Menzel-Glaser 22 × 60 mm Coverslip, Thermo Scientific). Images were acquired as per the section below on 'Image Acquisition'. After imaging, slides were submerged in TBS until coverslips dissociated. Sections were then washed and treated with bleach solution (5% H2O2 (Sigma) and 20mM NaOH (Sigma) diluted diluted in PBS) for 1h with light (15 watt, 2700k light bulb, 5cm above sample). Sections were checked under the microscope to ensure signal removal from all channels, and subsequently washed and incubated with blocking buffer with 10% rabbit serum for 15min at RT, followed by washing in TBS. Rabbit E-Cadherin-AF647 (Cell Signal, Cat: 9835) and

mouse anti-CD3 (Abcam, Cat: ab17143) were added overnight at RT. Sections were washed 3 × 5min in TBS and donkey anti-mouse-DyLight755 antibody (Invitrogen, Cat: SA5-10171) then added for 30min at RT. After washing 3 × 5min in TBS and rinsing in Milli-Q water slides were again mounted with SlowFade Diamond Antifade Mountant and imaged as described below.

Image acquisition

Images were acquired with a VS120 Slide Scanner equipped with an ORCA-FLASH 4.0 VS: Scientific CMOS camera (Olympus) and VS-ASW 2.9 software used for image acquisition and file conversion from vsi to tiff format. The entire tissue area was imaged using an ×20 objective (UPLSAPO 20X/NA 0.75, WD 0.6/CG Thickness 0.17) and select areas for representative images were acquired using an ×40 objective (UPLSAPO 40X/NA 0.95, WD 0.18/CG Thickness 0.11-0.23). For x40images, Z-stacks were acquired 3.5µm above and below the plane of focus with 0.5µm step sizes. Channels used include: DAPI (Ex 387/11-25nm; Em: 440/40-25nm), FITC (Ex: 485/20-25nm; Em: 525/30-25nm), TRITC (Ex: 560/25-25 nm; Em: 607/36-25nm), Cy5 (Ex: 650/13-25 nm; Em: 700/75-75nm) and Cy7 (Ex: 710/75nm, Em: 810/90nm). All channels were checked, and antibodies titrated beforehand, to ensure against signal-spill over between channels.

4.4.4 Quantification and statistical analysis

Image deconvolution and registration

Huygens Professional 18.10 (Scientific Volume Imaging, The Netherlands, http://svi.nl) CMLE algorithm, with SNR: 20 and 40 iterations were used for deconvolution of both single plane images acquired at x20, and also ×40 Z-stacks. Images were aligned using the ImageJ plugin multiStackReg vs1.45 with the DAPI channel serving as a reference for alignment.

Autofluorescence removal

Colorectal tissue is prone to autofluorescence from many sources such as red blood cells, blood vessels, apoptotic cells, intrinsically autofluorescent cells etc. Our early analyzes showed this substantially interfered with cell phenotyping and we could not remove autofluorescence using commercial quenching kits without significantly reducing our staining intensity. As such we developed 'Autofluorescence Identifier' (AFid) which analyzes pixels from two input fluorescent channels and outlines autofluorescent objects [14]. The code (https://ellispatrick.github.io/AFid) is implemented in MATLAB, R and Fiji. The Fiji version was used for this analysis. Pairs of channels compared were FXIIIA (on FITC) vs HIV RNA (on Texas Red) and HIV RNA vs CD11c. Autofluorescence masks were 'OR' combined and the resultant mask was used to exclude autofluorescent pixels (values set to 0) from images during data extraction as described below. The input parameters for AFid were as follows:

- Threshold: Niblack
- Min Area = 20 pixels.
- Max Area = 100000 pixels.

- Sigma = 2 pixels.
- Correlation cut off = 0.6.
- Number of clusters = 1.
- Max Value to automate k = 0.
- Glow Removal = Yes.
- Expansion Sensitivity = 20 pixels.
- Above values were for image resolutions of 3 pixels per µm.

HIV spot segmentation

HIV RNA particles were segmented using a custom MATLAB script based on a previously described spot counting algorithm [56]. First, a manual threshold of the HIV RNA channel was set to approximate areas of HIV stain. The IdentifySpots2D function by Battich et al. was then used to identify the number of spots. The detection threshold was set to 0.01 and deblending steps was set to 2. Identified spots were excluded if they did not overlap with the manually generated threshold mask in the first step.

Cell segmentation and Classification

Single cell segmentation was performed using a customized implementation of CellProfiler [57] in MATLAB (nucleiSegment.m, segRun.m function). Nuclei segmentation was performed by applying a local otsu filter to threshold the DAPI image and applying object-based watershed to identify boundaries. Objects with diameter between 3.3 and 16.7µm were kept. Masks of the CD3, CD11c and FXIIIa images were obtained by Gaussian blurring each image (sigma = 1.5) and performing a manual threshold to capture the full membrane. T he nuclei are then classified based on the percentage overlap of each nuclei object with each membrane mask. A cell is classified as a T cell if the overlap with CD3 is >20%, classified as a DC if the overlap with CD11c is >20%, and classified as a macrophage if the overlap with FXIIIa is >40%. Finally, three separate labeled cell masks are obtained for each cell type by expanding the nuclei to fill the membrane one pixel at a time (expandNucleus.m). For example, a nucleus identified as a macrophage is expanded into the FXIIIa mask space. A schematic with further details is provided in (Supplementary Figure 4.1C).

Tissue compartment segmentation

A manual threshold of the E-Cadherin stain was determined and nuclei (as segmented in the above section) belonging to this compartment were extracted using the BinaryReconstruct function in the 'Morphology' package in Fiji. The E-Cadherin and nuclei masks were then combined. The submucosa and lymphoid aggregates were manually outlined in Fiji and masks generated. The submucosa was defined as starting from the base of the crypts of Lieberkühn, whilst the border of lymphoid aggregates was determined by the increased density of CD11c and CD4 expressing DCs and T cells respectively. A mask of the whole tissue was then generated with a manually determined threshold. Subtracting the epithelium, lymphoid aggregates and submucosa from the whole tissue mask provided

a mask of the lamina propria. All masks were combined into a single image stack and assigned unique pixel values (eg, all epithelial pixels = 1, lamina propria pixels = 2 etc) so that they could be thresholded to extract data from each compartment.

Data extraction in MATLAB

Data extraction was performed in MATLAB (toRunNoNeighbrs.m, to RunNeighbrs.m). The mean marker expression and number of HIV particles was identified. From the compartment masks, a cell was considered to be part of that compartment if the overlap with the compartment mask was >25%. Distances from the compartments and HIV were obtained by creating a distance map from these objects and measuring the minimum value of the distance map within each cell. For cells within the LAs, the distance from the LA border was also measured in a similar approach. Neighbors were generated using the approach described in [27]. Finally, using the cell masks, the overlap between cells were also identified. All data were exported into a single csv file with rows as individual cells and columns as cell features such CD4 expression, HIV particle number, distance from LAs etc.

Analysis in R

All image analysis for this study was performed in R using the csv spreadsheet of cells and their features, generated as described in the previous section. Procedures for statistical analysis generating the results in this paper are described in the methods below and links to the source code are provided in Table 4.1. Donor and image numbers used for each analysis are indicated in figure legends.

HIV enrichment and association testing in cells

As the virus-cell interactions were from incoming viral particles and not those synthesized by the cell itself, some of the interactions observed were false positives due to random cell encounter with virus. We reasoned that the degree of background interactions for a given cell type would be proportional to the cell population's frequency in tissue. As such, measuring the percentage of total image HIV particles in each cell type, would largely reflect the relative abundance of each cell type. To get around this, we performed tests of association and enrichment, which are able to account for cell type abundance.

A Chi Square test of association was used to measure HIV enrichment in the grouped target cell population (DCs, macrophages, CD4+ T cells) in each image used for this study (Figure 4.3A). Here the expected number of HIV particles in target cells was determined by the proportion of all cells that comprised target cells. This showed that in most cases, HIV preferentially associates with the target cell population under study, rather than the undefined remaining cells in each image.

HIV enrichment was also calculated for specific populations. HIV enrichment was defined as the log2 transform of the percentage of virions associated with a cell type divided by the cell type's percentage of all cells (Figure 4.3C).

HIV enrichment testing in tissue compartments

The calculation of HIV enrichment was extended to measure preferential virus localization between LP and LA compartments, rather than between cell populations (Figure 4.4B-C). The 'expected' percentage of HIV particles in each compartment was based on the

proportional area of these compartments. For example, a 5:1 ratio of LP to LA area means that, by random chance, we expect 80% of virions in the LP and 20% in LAs. The null hypothesis of there being no difference in HIV localization between compartments (i.e. it just follows the area ratio) is represented visually as the line y = x in Figure 4.5B. Consequently, the residual variance (Euclidean distance of datapoints from the line y = x) represents the magnitude of HIV enrichment in LAs (for points above the line) or the LP (for points below the line) in each image. Comparing residuals as in Figure 4.5D is therefore akin to comparing the propensity of HIV to localize to LAs or the LP, accounting for differences in area between these compartments. Important to note, only virions in the LP and LA were used for this analysis (i.e. EP and SM virions excluded). Accordingly, the reciprocal values on the x and y axes of Figure 4.5B are measurements for the LP. For example, the rightmost red data point is an image with 40% of the area comprised of LAs, and 60% of HIV in LAs. In this same image, 60% of the image area is comprised of LP which contains 40% of HIV.

Measuring HIV density across tissue compartments

HIV density was calculated for each tissue compartment separately. For a given compartment, HIV density was calculated using the top 10 images with the highest HIV density for that compartment. The HIV density of all compartments, calculated in this way, was then compared (Figure 4.4A). This approach was chosen to mitigate variation due to inter-image differences in HIV concentration, penetration depth and compartment proportions. For example, for a given donor the density of HIV in LAs often varies substantially by section due to difference in depth of viral penetration at 2h. This can lead to extremely high LA HIV densities in one section, and almost no virions in LAs in other sections from the same donor. In this case, using all images would mask the high HIV density observed in LAs in the earlier section. To get around this issue, we compared the highest observable HIV densities in each compartment, rather than simply measuring all compartments in all images.

Cellular spatial distribution within compartments

Measurement of cellular spatial distribution within compartments was performed using distance maps emanating from compartment borders (see 'Data extraction in MATLAB'). Cell density was measured either in non-cumulative intervals (Figure 4.2D, left) or in just two regions defined as "-proximal" and "-distal" to the compartment border (Figure 4.2D right). The latter was mostly used for statistical comparisons as comparing only two regions allowed for more observations (cells) per region and better statistical power. For EP, LP and SM, the non-cumulative intervals were linear (e.g., 0–20µm, 20–40µm ... etc) (Figure 4.2D, left). To measure changes throughout LAs we used a different approach. To achieve comparable areas for each interval we assumed LAs to be spherical and hence circular in 2D. We then calculated the radial edges of each interval from the outer edge using the formula: 1 – sqrt (1 – k/n) where k = 1,2 ... n-1 and n = max interval number. For a perfect circle this would derive intervals of equal area (Figure 4.4H and Supplementary Figure 4.3G-K).

Cellular gradients in response to HIV

Distance maps emanating from HIV particles (see 'Data extraction in MATLAB') were used to analyze changes in cell density in the vicinity of HIV (Figure 4.5). The formation of increasing/decreasing cell density gradients in non-cumulative intervals from HIV particles was inferred as potential cell migration to/from HIV within a given compartment (Figure 4.5A).

Potential migration of LP target cells into the EP (Figure 4.5B-C) was assessed by measuring the cell density both near (≤10µm from the EP-LP interface) and within HIV+ EP vs HIV- EP. A single relatively small interval of 10µm, rather than multiple intervals, was chosen due to the small space between epithelial crypts. If a comparative increase in the density of a target cell was observed both leading up to HIV+ EP and within the HIV+ EP itself, then we inferred this was likely the target cell migrating from the LP into the EP to sample HIV particles. We performed several filtering steps to remove potential interfering variables. In particular, HIV+ and HIV- EP cells could be located next to one another, and so HIV- EP were specifically selected as >50µm away from HIV+ EP. Additionally, many regions have more HIV in the LP than the EP, which could draw target cells away from HIV+ EP and toward HIV+ LP instead. To mitigate this effect images were binned into 100 × 100µm quadrats and only quadrats with more HIV particles in EP than LP were used for analysis. Finally, only images with at least one cell in each category were analyzed (e.g. at least one DC beneath each HIV+ and HIV- EP).

Potential migration of LP target cells into LAs (Figure 4.5D-E) was assessed by measuring the cell density in non-cumulative 50µm intervals from LAs and within LAs themselves. These measurements were split by whether LAs were HIV+ or HIV- and so only donors with both HIV+ and HIV- LAs were used in this analysis. We also compared the spatial distribution of target cells throughout intervals of HIV+ vs HIV- LAs (see 'Data extraction in MATLAB' for methodology). This was to determine whether target cells clustered more toward the outer intervals of HIV+ LAs, which is where cells entering LAs from the LP would likely be located.

Analysis of pathways for submucosal HIV entry

HIV entry into the submucosa could only occur via either the LP or LAs, as these are the mucosal compartments directly overlying the submucosa. To analyze which pathway was most likely, we created linear models (lm() function in R) of submucosal HIV density as a function of LP HIV density, LA HIV density, or both (Figure 4.3I). Datapoints were individual explants, i.e., a compartment across all images from an explant was measured, producing a single value for that explant. The combined model showed that LA HIV density was the only significant predictor of SM HIV density. Although LP HIV density alone showed some predictive capacity that was borderline significant ($\beta = 0.54$, $p = 0.04$), this effect was diminished in the combined model. This was likely due to collinearity between LA and LP HIV density ($\beta = 0.71$, $p = 0.004$), causing the LP to transmit effects from the LA HIV density when attempting to predict SM HIV density. This assumes the direction of the association is that LP HIV density affects LA HIV density, and not the other way around. To explain why this assumption is likely true, consider two possible scenarios for HIV entry into the SM (1) LP → LA → SM and (2) LA → LP → SM. In each case, the middle variable is called the 'mediator' as it transmits the effects of the first variable to the last one. The criteria for a variable to be classed as a mediator are

well established [58] and are explained in simple by MacKinnon et al., [59].

1. There must be a significant relationship between the independent variable and the dependent variable,

2. There must be a significant relationship between the independent variable and the mediating variable, and

3. The mediator must be a significant predictor of the outcome variable in an equation including both the mediator and the independent variable.

Referencing the equations in Figure 4.3I, we can see that only the LA variable satisfies these properties and can be classed as a mediator, indicating that the LP $\rightarrow$ LA $\rightarrow$ SM pathway is more likely.

SpicyR analysis

The R package 'SpicyR' [15] was used to analyze differential cell-cell localization between HIV and mock-treated samples in the EP, LP, LA and SM (Figure 4.6A). As HIV particles were only present in specific regions of the images from HIV-treated explants, we selected regions near HIV particles to compare to the mock sample. This is because the numerous HIV- regions would dilute the HIV-induced effects on cell-cell interactions when comparing to mock. The HIV-proximal region was defined as all cells within 30µm of HIV for the EP, LP and LA compartments, and 100µm for the SM due to the lower cell density in this compartment. As a control, HIV-distal regions (>30µm or 100µm) were compared to the mock sample. p values were centered at $p = 0.01$ which was the assigned cut-off for significant changes for this analysis. Importantly, the SpicyR test function measures interactions of a cell type within a user specified radius around the cell. As the chosen radius is a possible source of variation in results, we tested multiple radii. We did this alongside testing multiple cut-offs to define the HIV-proximal region and found our results were robust to variation in these parameters (Supplementary Figure 4.4A).

Temporal inference analysis

Having observed that LP target cells clustered away from the EP-interface we wanted to know whether this could be due to target cell migration away from the EP-interface in response to incoming HIV particles (Supplementary Figure 4.4B). As we only performed a single 2h time point for our explants we devised a method of temporal inference to investigate this phenomenon. We observed different proportions of HIV in the EP and LP in different regions of our images. This was likely due to each region being at a different stage of HIV entry, with regions where most HIV was in EP representing early stages of HIV entry, and regions where most HIV was in LP representing later stages (as HIV would have already passed the EP barrier to enter the LP). As such we reasoned that we could use the relative proportion of HIV in EP vs LP as a proxy for 'early' vs 'later' stages of HIV entry. In particular, we divided images into 100 $\times$ 100µm windows with each window classified as HIV- windows (no HIV particles), 'early' (EP > LP HIV particle count) or 'late' (EP < LP HIV particle count) in terms of HIV entry. We then calculated the log2 fold change in LP target cell density in EP proximal ($\geq$10µm) vs distal regions (10–50µm) from either HIV- EP or HIV+ EP. A negative fold change indicates target cell enrichment in the distal region. If the magnitude of the negative fold change was higher

in the distal region from HIV+ EP vs HIV- EP, then it means that HIV exposure to the tissue caused further relocation of the target cell away from the EP.

It's important to note that all windows contained a mixture of HIV+ and HIV- EP which is why the distance of LP target cells from each type of EP could be calculated for all windows. The HIV- EP was >50µm from HIV+ EP to minimize interference (as also described in 'Cellular gradients in response to HIV'). The reason for measuring LP target cell distance from HIV- vs HIV+ EP is that it controlled for local fluctuations in LP target cell density, allowing us to better observe how a local region changes in early vs later stages of HIV entry. This is opposed to measuring cell distance from the EP in general and directly comparing changes in target cell density in early vs late windows. These windows would be in different regions and therefore could have vastly different steady-state target cell densities (therefore affecting the fold change measurements close and far from the EP).

HIV-transfer phenotype score

As HIV is present at the interface between cells during active viral transfer [22–26], neighboring cells engaged in transfer would both appear as HIV+ in our data. We sought to quantify the frequency of this phenotype between cell pairs, as a proxy for cellular engagement in viral transfer. This was achieved by employing a modified version of Neighborhood Analysis [27] and constructing a 'HIV-transfer phenotype score' from the results to estimate the degree to which a HIV+ cell increases the likelihood of nearby cells being HIV+, whilst controlling for possible confounders (Figure 4.6C). In particular, we first quantified the number of HIV+ and HIV- cell neighbors for HIV+ cells. The cell labels (locations) of HIV+ cells, rather than all cells, were randomized 999 times to generate a null distribution of HIV+ cell neighborhoods, against which the actual HIV+ cell neighborhood was compared to determine significant interactions ($p < 0.005$). Randomization was restricted to HIV+ cells to control for the background effect of HIV particles on the localization of cells in tissue (i.e. cells are already at an increased density in HIV+ regions). This is a key difference to standard neighborhood analysis which randomizes on all cells. T his results in association scores for HIV+:HIV+ interactions (HIV+ cell with HIV+ neighbor) and 'HIV+:HIV-' interactions (HIV+ cell with HIV- neighbor), which was expressed as the percentage of all images showing significant interactions for the cell pair. Note that in this analysis, cells of the same type often appear to significantly interact with each other, which has been noted in previous spatial studies analyzing cell:cell interactions in tissue [60–62].

Using these data we constructed a 'HIV-transfer phenotype score' which was defined as the difference in the frequency of 'HIV+:HIV+' interactions and 'HIV+:HIV-interactions. HIV+:HIV- interactions represent the moment before potential transfer and it's frequency is determined by all conditions leading up to a transfer event, including the steady-state likelihood of two cells interacting and any HIV-induced effects on the movement of either cell just prior to a potential transfer. HIV+:HIV+ interactions include this, plus the direct influence of cellular HIV-binding on the formation of interactions. As such, the difference in these measurements reflects the propensity of a HIV+ cell to transfer virus to another specific cell type in its immediate neighborhood. Important to note, HIV+:HIV+ interactions could be in the form of two cells that each have HIV particles, but are not engaged in the process of transfer. As HIV is not influencing the interaction, the frequency of such events would be the same as the steady-state interaction frequency of the cell pair. This frequency is included in the HIV+:HIV- interaction frequency and

is therefore accounted for (subtracted) in the 'HIV-transfer phenotype score'.

Association between T cell HIV-load and DC/macrophage interaction

This analysis was performed to determine whether the HIV-load (number of HIV particles) in a CD4+ T cell is associated with increased interactions with either DCs or macrophages (Figure 4.6E). In particular, CD4+ T cells were split into 4 groups based on the number of HIV particles they contained (0, 1, 2–3 or $\geq$4 virions). For T cells in each group, their degree of interaction with DCs or macrophages was measured as the number of pixels in the T cell body that belonged to either of these cells (DC: CD4+ pixels; Macrophage: FXIIIa+ pixels). To account for inter-image variation, we performed all measurements intra-image which meant that a single image was required to have CD4+ T cells with a spectrum of HIV levels to allow for a fair comparison. Accordingly, only images with at least 3 CD4+ T cells in each category (0, 1, 2–3, $\geq$4 virions) were used for the analysis.

Acknowledgements

This work was funded by the National Health and Medical Research Council Ideas grant (GNT1181482). The authors would like to acknowledge the Westmead Cell Imaging and Flow Cytometry Core Facilities, supported by the Westmead Institute, Westmead Research Hub, Cancer Institute New South Wales.

Author contributions

Conceptualization, H.B. and A.N.H.; methodology, H.B., A.N.H., N.C., E.E.V., K.M.B., and J.W.R.; experimental investigation, H.B., E.E.V., and D.Y.; computational analysis, H.B., N.C., K.H., and E.P.; novel reagents and samples, J.T., G.C., N.P., G.C., F.R., A.D.R., and M.P.G,; visualization, H.B., K.H., and E.E.V.; writing, H.B. and A.N.H.; funding acquisition, A.N.H. and A.L.C. intellectual input and critical revisions, H.B., A.N.H., A.L.C., E.E.V., K.J.S., K.M.B., N.-P.N., S.N.B., J.D.E., and M.A.H. All authors read and approved of the manuscript.

Declaration of interests

The authors declare no competing interests.

Data and code availability

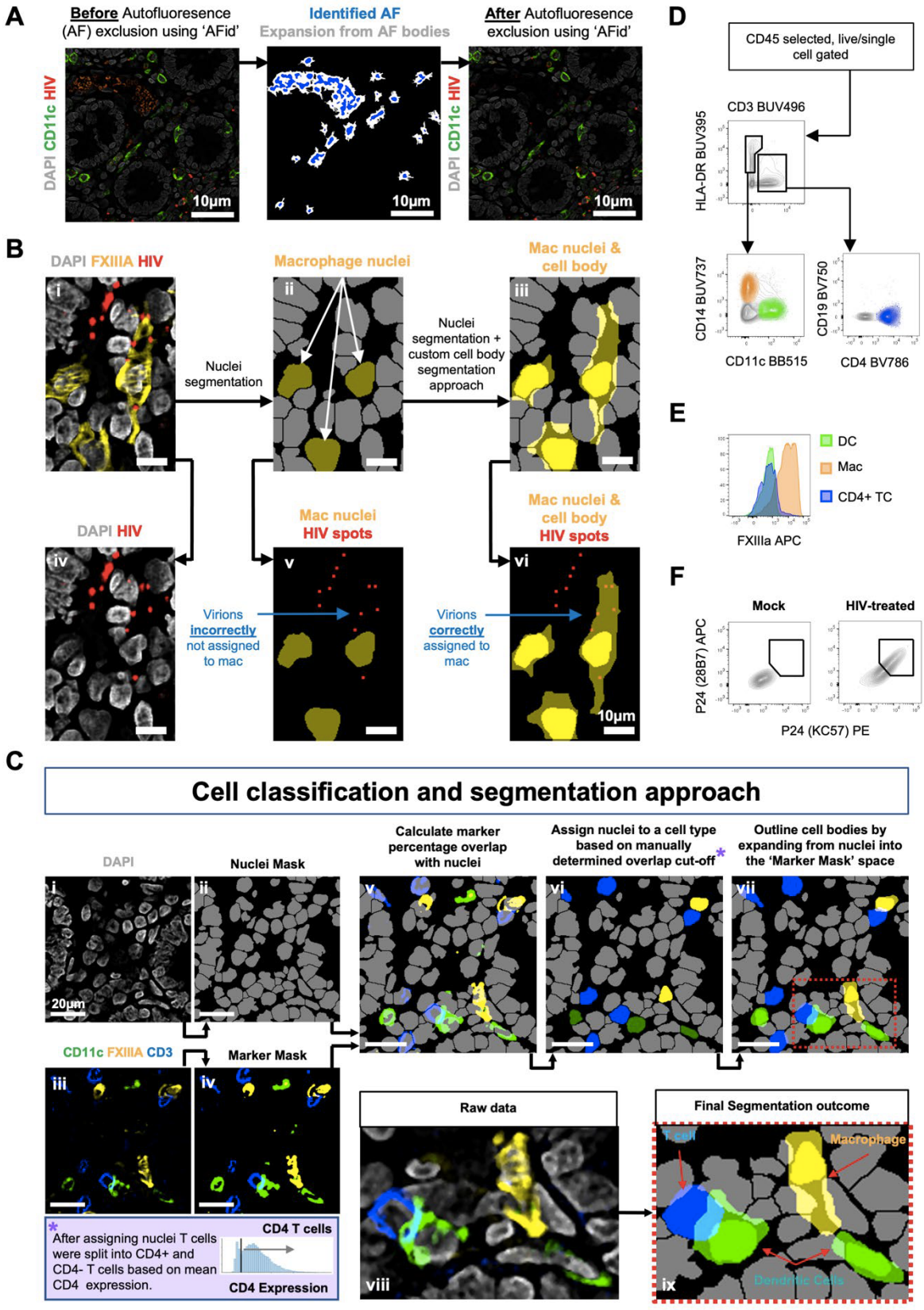
- Raw image data and the processed spreadsheet used for analysis have been deposited at Zenodo at https://doi.org/10.5281/zenodo.6992156 and are publicly available as of the date of publication. Flow cytometry data reported in this paper will be shared by the lead contact upon request.
- All original code and instructions for image processing/analysis is available at GitHub (https://github.com/heevaBaharlou/HIVImageAnalysis) and deposited at Zenodo at https://doi.org/10.5281/zenodo.6992156.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

4.5 Supplementary Material

Supplementary Table 4.1: **Donor age and sex information for colorectal tissues used in this study** (related Methods Human subjects). *New batch images were used for the majority of analyses in the paper as they were stained for all cell type defining markers. The old batch images were stained with fewer markers and therefore only used for analysis of HIV localisation across compartments (e.g. HIV virion density in compartments (Figure 4.4A).

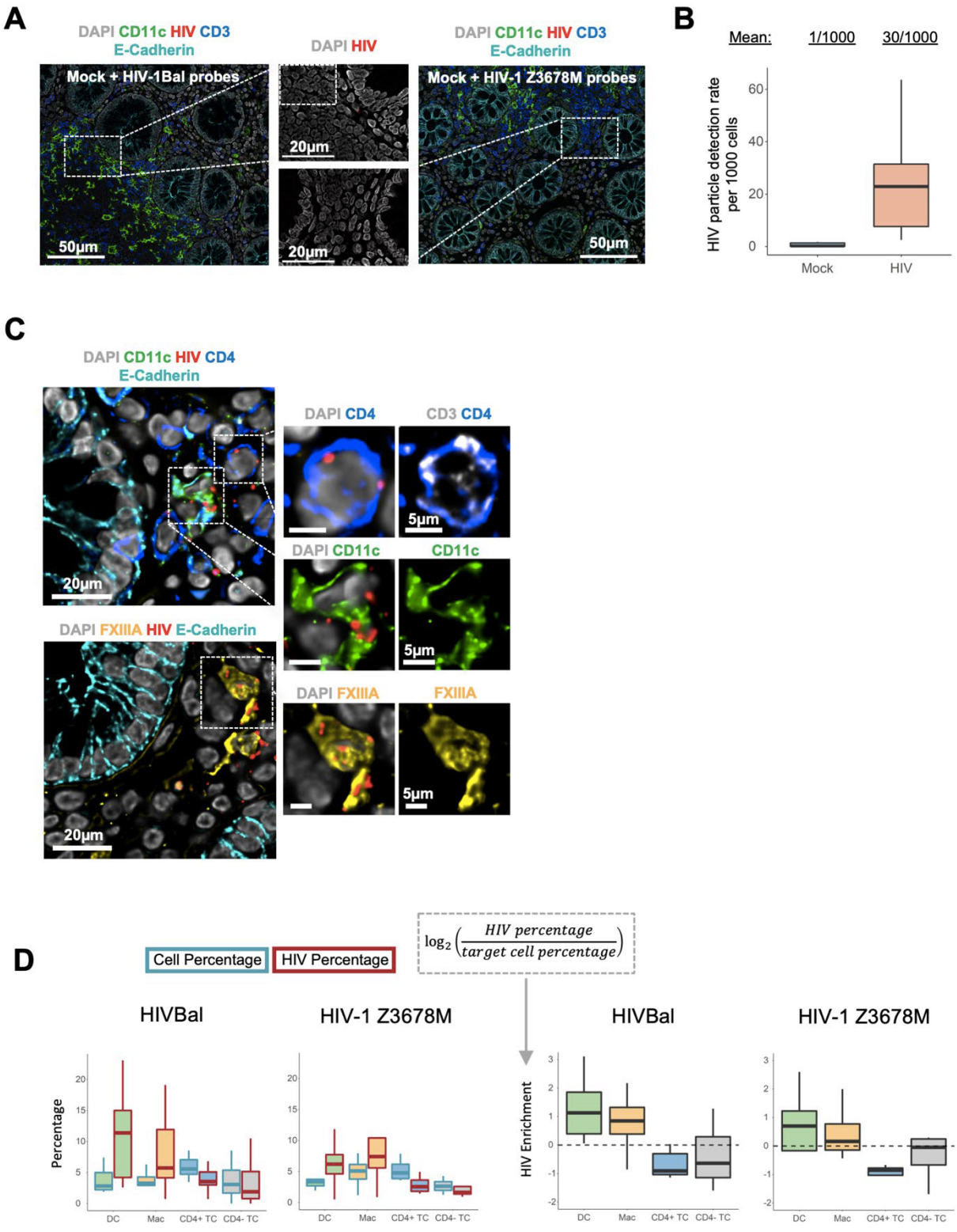
Donor	Batch*	Age	Sex
1	new	54	M
2	new	46	F
3	new	58	F
4	new	91	F
5	new	61	M
6	new	80	F
7	new	89	F
8	new	74	M
9	new	67	M
10	new	84	M
11	new	53	F
12	new	84	M
13	old	50	M
14	old	60	F
15	old	74	F
16	old	51	M
17	old	60	M
18	old	36	F
19	old	46	F
20	old	51	M

This page has been left blank intentionally.



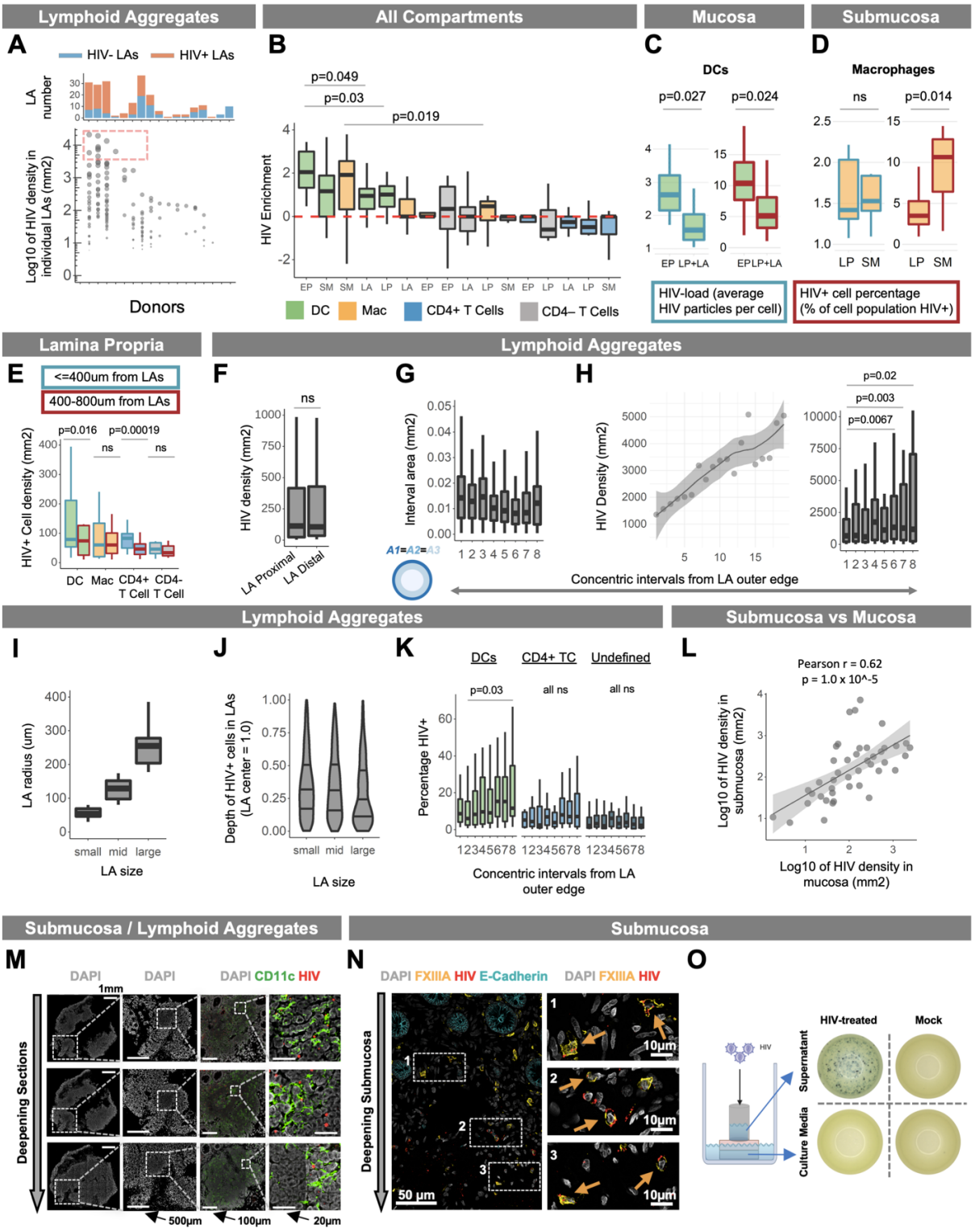
Supplementary Figure 4.1: **Cell Segmentation method (related to Figure 4.1).** (continued on next page)

Supplementary Figure 4.1: **(A)** Images before (left) and after (right) autofluorescence removal using AFid. The middle image shows the mask of the identified autofluorescence with core autofluorescent bodies in blue and the expansion mask shown in white. The expansion mask flows out from the autofluorescent bodies to capture all autofluorescence and is designed with a halting condition whereby the detection of stromal background fluorescence or antibody derived signal halts the expansion. Note that the algorithm also detects low signal autofluorescence and so some of the outlined autofluorescent bodies in the middle panel are not easily visible by eye. Importantly none of these signals overlap with the CD11c or HIV signals in the image. **(B)** These images illustrate the necessity of the custom segmentation approach outlined in part C for accurately assigning HIV virions to the correct cell type. A macrophage sampling multiple HIV particles is shown **(i)**, as is the inadequacy of nuclei segmentation alone **(ii)** for accurately assigning HIV particles to the macrophage **(v)**. This is rectified via the cell boundary estimation incorporated into our segmentation method **(iii)** which allows for the accurate assignment of HIV particles to the macrophage. **(C)** Cell classification and segmentation approach. First, a mask of the nuclei in each image was created by performing thresholding on DAPI followed by a morphological watershed **(i-ii)**. Masks of CD11c, FXIIIa and CD3 were then created via manual thresholding **(iii-iv)**. DCs, macrophages and T cells were classified by measuring the percentage overlap of marker masks from iv with the segmented nuclei from **ii (v)**. The threshold for classifying a cell type was determined visually and was typically around 20% overlap for membrane markers CD11c and CD3 and 40% for the intracellular marker FXIIIa **(vi)**. Once classified we then split the nuclei mask into three separate masks, each containing the nuclei belonging to one of the 3 cell types. In each mask, we then outlined the cell body, using distance maps emanating from nuclei centers, and restricted to the area of the masks of cell-type defining markers from part iv. This was followed by a morphological watershed to separate touching cells. The result was 3 masks which estimate the full cell body for each cell type **(vii)**. Notably, a pixel from a single coordinate across the masks can belong to multiple cell types as the expansion in part vii was performed on separate masks. This is reflective of the actual situation *in situ* whereby cells interact in 3D and so will exhibit overlap with one another when taking 2D image slices. T cells were classified as CD4+ and CD4- by a manually determined cut off for CD4 expression (bottom left box). The raw data and the final segmentation outcome are shown side by side **(viii-ix)**. **(D)** Gating strategy for HIV target cell identification by Flow Cytometry [11]. **(E)** Representative histograms of FXIIIa expression on DCs, macrophages and CD4+ T cells (n = 3), showing FXIIIa expression is specific to macrophages as defined in part **D**. **(F)** Cell types defined in part **D** were either treated with HIV_{Bal} , HIV_{Z3678M} or untreated (mock) and stained with two different clones of antibodies (clones 28B7 and KC57) targeting p24. Dual p24+ positive cells were defined as HIV+ cells as per [8, 11].



Supplementary Figure 4.2: **Assessment of interactions of HIV with colorectal target cells (related to Figure 4.3)** (continued on next page)

Supplementary Figure 4.2: **(A)** Images from mock (PBS treated) explants stained using probes against either HIVBal or HIVZ3678M. **(B)** HIV virions detected per 1000 cells in mock or HIV-treated samples stained using probes against either HIVBal or HIVZ3678M. **(C)** Representative images of colorectal target cells interacting with HIVZ3678M particles. **(D)** Comparison of HIV uptake (HIV Percentage) relative to opportunity (Cell Percentage) across target cells and for two strains of HIV which were HIVBal and HIVZ3678M. Left: For each cell type their percentage among all cells (blue border) or the percentage of all HIV particles in those cells (red border) is shown. Right: HIV particle percentage normalized to the cell percentage, and log2 transformed. Data points represent individual donors, where the results from multiple images were averaged for each donor. HIV_{Bal}: n = 11, HIV_{Z3678M}: n = 4.

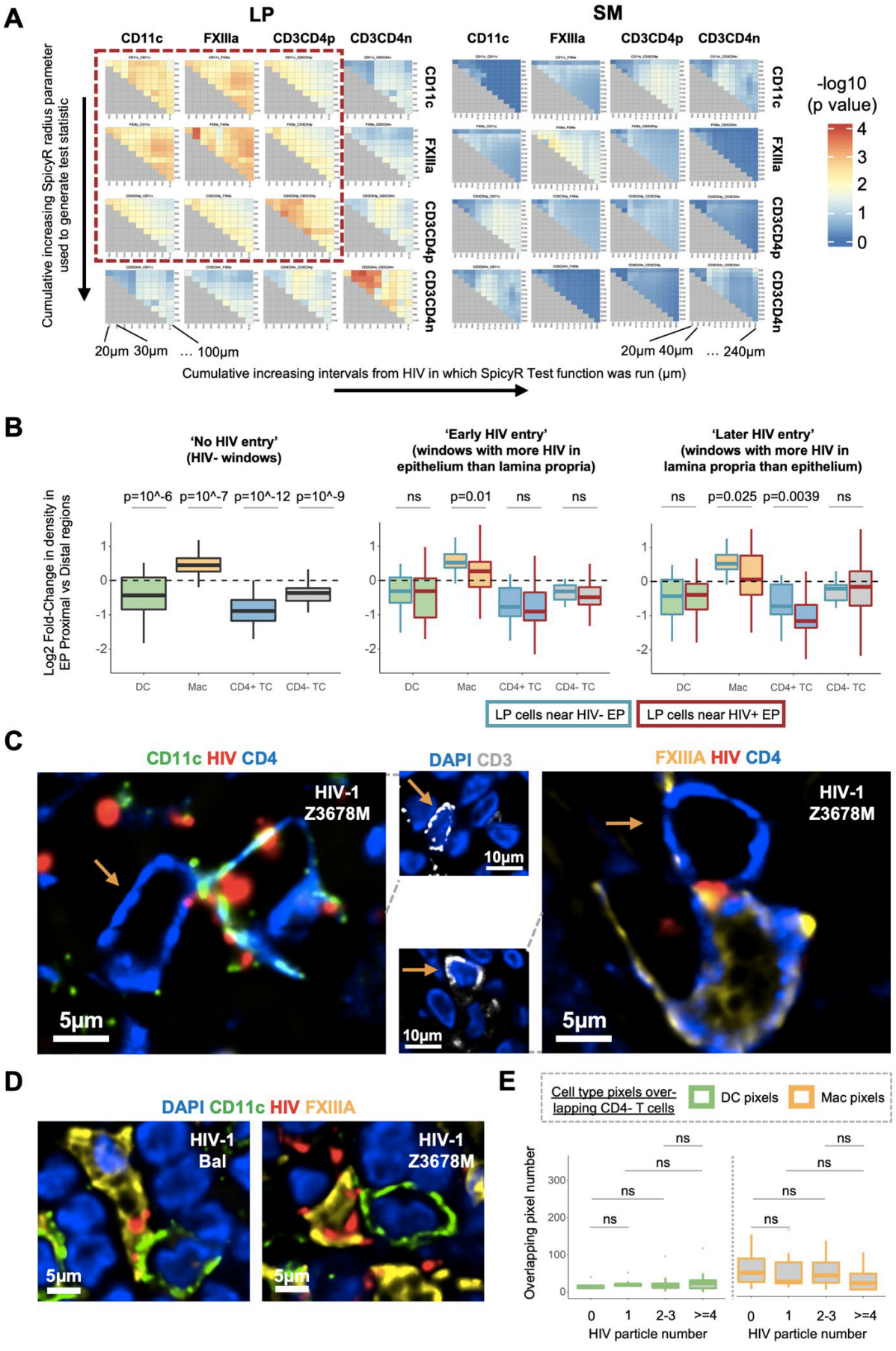


Supplementary Figure 4.3: **Differential HIV uptake across colorectal tissue compartments (related to Figure 4.4)** (continued on next page)

Supplementary Figure 4.3: **(A)** Density of HIV virions per mm^2 in 215 LA images (y-axis) across 17 donors (x-axis). Each data point represents an individual LA from a given donor with dot size correlating to LA HIV density. The annotation above indicates the total number of LAs counted and is colored by whether they are HIV+ (red) or HIV- (blue). Dotted box indicates a sample of LAs that were highly enriched with HIV (>4000 virions per mm^2). **(B)** HIV enrichment in different cell types across tissue compartments. The HIV particle percentage was normalized to the cell percentage (data from Figure 4.4E), and log2 transformed. A Wilcoxon signed-rank test was performed to compare this HIV enrichment score across data points (specific cell types in specific tissue compartments). Data represent 15 explants from 12 donors, where the results from multiple images were averaged for each explant. **(C-D)** HIV-load (average virions per cell) or HIV+ cell percentage (percentage of cell population that is HIV+) was compared between 'EP vs sub-EP' compartments (LP + LA) for DCs and 'LP vs SM' for macrophages. Data represent individual donors where a donor was only included if at least 5 HIV+ cells were detected in both compartments for HIV-load measurements or 10 cells detected in both compartments for HIV+ cell percentage measurements. This was to allow for a fair comparison between compartments. A Wilcoxon rank-sum test was performed to compare measurements across compartments. LP vs SM: n=10 donors; 'EP vs sub-EP': 12 donors for 'HIV-load' measurements and 11 donors for HIV+ cell percentage measurements. **(E)** Density of HIV+ ($\geq$1 particle) target cells and CD4- T cells in LA-proximal ($\leq$400µm, blue border) vs - distal (400-800µm, red border) regions of LP. 400µm was chosen (instead of 200µm as in Figure 4.2D) in order to capture a larger pool of HIV+ cells, thus improving the reliability of comparisons. Data represents the 'proximal' and 'distal' LP from individual LAs. HIV+ cells for each cell type were compared between regions (Wilcoxon signed-rank test) only if each region contained at least 5 HIV+ cells, so as to allow for a fair comparison. DC: n=19 LA (5 donors), Mac: n=12 LA (3 donors), CD4+ TC: 15 LA (6 donors): CD4- TC:12 LA (5 donors). **(F)** Density of HIV particles in LA-proximal ($\leq$400µm) vs -distal (400-800µm) regions of LP. A Wilcoxon signed rank test was between LA-proximal and -distal regions of LP. N=143 LAs from 11 donors. **(G)** Area (mm^2) of LA intervals. In order to achieve comparable areas for each interval we assumed LAs to be spherical and hence circular in 2D. We then calculated the radial edges of each interval from the outer edge using the formula: 1 - sqrt (1 - k/n) where k = {1,2 ... n-1} and n = max interval number. In this example n = 8. For a perfect circle this would derive intervals of equal area. N=115 LAs from 15 donors. **(H)** Left: Density of HIV+ cells in intervals from the outer edge of HIV+ LAs (x = 1) toward their center (x = 20). Results are shown as a LOESS curve of best fit to highlight cell density trends from outside LAs toward their center. Right: statistical comparisons (Wilcoxon signed rank test) of discrete intervals where LAs were split into 8 intervals instead of 20 in order to increase the number of cells measured per interval, thus reducing error and increasing the reliability of comparisons. N=115 LAs from 15 donors. **(I-J)** HIV+ cell depth in LAs (continued on next page)

Supplementary Figure 4.3: **(J)** was measured in LAs categorized as small, medium or large **(I)**. HIV+ cell depths were scaled such that '0 = LA outer edge' and '1 = LA center'. LAs were assigned to each group by performing a quantile split, creating 3 even groups based on LA radius. Only LAs with at least 10 HIV+ cells were measured. Note that since area increases exponentially with increasing radius, 75% of the LA area exists half-way (y=0.5) toward the LA center. As such in this graph it appears as though HIV+ cell frequency is higher toward the LA edge, when in fact the opposite is true as shown in Figure S3H. n=58 LAs from 11 donors. **(K)** Percentage of LA DCs, CD4+ T cells or undefined cells (not a DC, Mac or T cell) that are HIV+ in intervals from the outer edge of HIV+ LAs (x = 1) toward their center (x = 8). For each cell type, only LAs with at least 1 HIV+ cell in each interval were selected for analysis. A Wilcoxon signed rank test was performed between indicated intervals. DC: 87 LAs (13 donors); CD4+ TC: 72 LAs (12 donors); Undefined cells: 92 LAs (13 donors). **(L)** HIV virion density (per mm^2 of DAPI) in SM vs mucosa (EP + LP + LA) in individual images. Smoothed linear regression is shown with Pearson's r and its associated p value. Images were only included if the SM and mucosa both contained HIV virions. N=43 images (16 donors). **(M)** Representative images showing HIV virions throughout various depths of a single LA from superficial sections in the mucosa (top row) to deeper sections where the LA has penetrated the submucosa (bottom row). **(N)** Representative images showing HIV+ Macs exist throughout various depths of the SM, including regions close to the LP (box 1) and deeper regions (box 3). **(O)** Representative images of TZM-bl assay results using either viral inoculum within cloning cylinders or the culture media in which the explants sit. Solutions were collected at the end of the viral culture period and then cultured with TZM-bls for 72h. Infected cells were visualized upon addition of X-gal substrate with each blue dot corresponding to an infected cell. n=6 explants. All density measurements were performed per mm^2 of DAPI for indicated regions.

This page has been left blank intentionally.



Supplementary Figure 4.4: **Signatures of HIV-induced cell-cell interactions *in situ* (related to Figure 4.6)** (continued on next page)

Supplementary Figure 4.4: **(A)** SpicyR analysis as in Figure 4.6A with variation in radius (in which to measure cell-cell interactions) and HIV region distance cut-off (distance from HIV in which interactions were analysed). This analysis was run for LP and SM. The dotted red box encases HIV target cells in the lamina propria which can be seen to form significant clusters with most parameter settings. **(B)** LP target cell proximity to the EP in response to inferred stages of HIV entry. Images were divided into 100x100µm windows with each window classified as HIV- windows, 'early' (EP > LP HIV particle count) or 'late' (EP < LP HIV particle count) in terms of HIV entry. The log2 fold-change in LP target cell density in EP proximal (≤10µm from EP) vs distal regions (10-50µm from EP) was calculated for HIV- windows (left). For 'early' and 'late' stage windows the log2 fold-change was calculated from HIV- (blue border) and HIV+ EP (red border) separately. Only HIV- EP greater than 50µm from HIV+ EP were used. A Wilcoxon signed-rank test was used to compare 'EP proximal vs EP distal fold-change in LP cells between HIV+ vs HIV- EP. Data from 40 images across 12 donors were used for this analysis. **(C)** Representative images of DCs and macrophages interacting with CD4+ T cells where HIV_{Z3678M} is present at the interface between the cells. For clarity, CD3 staining is shown in a separate image with a brown arrow pointing to CD3+CD4+ T cells. **(D)** Representative images of DCs and macrophages interacting with one another where HIV_{Z3678M} is present at the interface between these cells. **(E)** Number of DC (CD11c) or macrophage (FXIIIa) positive pixels overlapping with the body of CD4- T cells (y axis) that harbor varying levels of HIV particles (x axis). Only images with at least 3 CD4- T cells in each category (0, 1, 2-3, ≥4 HIV particles) were selected for the analysis which consisted of 9 images across 5 donors. This comprised X images from Y donors. A Wilcoxon signed-rank test was performed to compare levels of DC or Mac pixel overlap between CD4- T cells with varying levels of HIV. A Wilcoxon rank-sum test was performed to compare differences in the magnitude of membrane overlap between CD4- T cells of varying HIV load. All density measurements were performed per mm^2 of DAPI for indicated regions.

Bibliography

[1] Erica E Vine, Jake W Rhodes, Freja A Warner van Dijk, Scott N Byrne, Kirstie M Bertram, Anthony L Cunningham, and Andrew N. Harman. Hiv transmitting mononuclear phagocytes and integrating the old and new. *Mucosal immunology*, 15(4):542–550, 2022. doi: 10.1038/s41385-022-00492-0. URL https://doi.org/10.1038/s41385-022-00492-0.

[2] Claire Deleage, Taina T. Immonen, Christine M. Fennessey, Arnold Reynaldi, Carolyn Reid, Laura Newman, Leslie Lipkey, Timothy E. Schlub, Celine Camus, Sean P. O'Brien, Jeremy Smedley, Jessica M. Conway, Gregory Q. Del Prete, Miles P. Davenport, Jeffrey D. Lifson, Jacob D. Estes, and Brandon F. Keele. Defining early siv replication and dissemination dynamics following vaginal transmission. *Science advances*, 5(5):eaav7116–NA, 2019. doi: 10.1126/sciadv.aav7116. URL https://doi.org/10.1126/sciadv.aav7116.

[3] James B. Whitney, Alison L. Hill, Srisowmya Sanisetty, Pablo Penaloza-MacMaster, Jinyan Liu, Mayuri Shetty, Lily Parenteau, Crystal Cabral, Jennifer L. Shields, Stephen Blackmore, Jeffrey Y. Smith, Amanda L. Brinkman, Lauren Peter, Sheeba I. Mathew, Kaitlin M. Smith, Erica N. Borducchi, Daniel I. S. Rosenbloom, Mark G. Lewis, Jillian Hattersley, Bei Li, Joseph Hesselgesser, Romas Geleziunas, Merlin L. Robb, Jerome H. Kim, Nelson L. Michael, and Dan H. Barouch. Rapid seeding of the viral reservoir prior to siv viraemia in rhesus monkeys. *Nature*, 512(7512):74–77, 2014. doi: 10.1038/nature13594. URL https://doi.org/10.1038/nature13594.

[4] Tae Wook Chun, Delphine Engel, M. Michelle Berrey, Theresa Shea, Lawrence Corey, and Anthony S. Fauci. Early establishment of a pool of latently infected, resting cd4+ t cells during primary hiv-1 infection. *Proceedings of the National Academy of Sciences of the United States of America*, 95(15):8869–8873, 1998. doi: 10.1073/pnas.95.15.8869. URL https://doi.org/10.1073/pnas.95.15.8869.

[5] Donn J Colby, Lydie Trautmann, Suteeraporn Pinyakorn, Louise Leyre, Amélie Pagliuzza, Eugene Kroon, Morgane Rolland, Hiroshi Takata, Supranee Buranapraditkun, Jintana Intasan, Nitiya Chomchey, Roshell Muir, Elias K. Haddad, Sodsai Tovanabutra, Sasiwimol Ubolyam, Diane L. Bolton, Brandie A. Fullmer, Robert J. Gorelick, Lawrence Fox, Trevor A Crowell, Rapee Trichavaroj, Robert J. O'Connell, Nicolas Chomont, Jerome H. Kim, Nelson L. Michael, Merlin L. Robb, Nittaya Phanuphak, and Jintanat Ananworanich. Rapid hiv rna rebound after antiretroviral treatment interruption in persons durably suppressed in fiebig i acute hiv infection. *Nature medicine*, 24(7):923–926, 2018. doi: 10.1038/s41591-018-0026-6. URL https://doi.org/10.1038/s41591-018-0026-6.

[6] Danijela Maric, Wesley A Grimm, Natalie Greco, Michael D. McRaven, Angela J. Fought, Ronald S. Veazey, and Thomas J. Hope. Th17 t cells and immature dendritic cells are the preferential initial targets after rectal challenge with a simian immunodeficiency virus-based replication-defective dual-reporter vector. *Journal of virology*, 95(19):e0070721–NA, 2021. doi: 10.1128/jvi.00707-21. URL https://doi.org/10.1128/jvi.00707-21.

[7] Daniel J. Stieh, Edgar Matias, Huanbin Xu, Angela J. Fought, James Blanchard, Preston A. Marx, Ronald S. Veazey, and Thomas J. Hope. Th17 cells are pref-

erentially infected very early after vaginal transmission of siv in macaques. *Cell host & microbe*, 19(4):529–540, 2016. doi: 10.1016/j.chom.2016.03.005. URL https://doi.org/10.1016/j.chom.2016.03.005.

[8] Kirstie M. Bertram, Rachel A. Botting, Heeva Baharlou, Jake W. Rhodes, Hafsa Rana, J. Dinny Graham, Ellis Patrick, James Fletcher, Toby M. Plasto, Naomi R. Truong, Caroline Royle, Chloe M Doyle, Orion Tong, Najla Nasr, Laith Barnouti, Mark P. Kohout, Andrew J. Brooks, Michael Wines, Peter A. Haertsch, Jake Lim, Martijn P. Gosselink, Grahame Ctercteko, Jacob D. Estes, Melissa J Churchill, Paul U. Cameron, Eric Hunter, Muzlifah Haniffa, Anthony L. Cunningham, and Andrew N. Harman. Identification of hiv transmitting cd11c+ human epidermal dendritic cells. *Nature communications*, 10(1):2759–2759, 2019. doi: 10.1038/s41467-019-10697-w. URL https://doi.org/10.1038/s41467-019-10697-w.

[9] Yonatan Ganor, Z Zhou, Daniela Tudor, Alain Schmitt, Marie-Cécile Vacher-Lavenu, Gibault L, N Thiounn, J Tomasini, Jean-Philippe Wolf, and Morgane Bomsel. Within 1 h, hiv-1 uses viral synapses to enter efficiently the inner, but not outer, foreskin mucosa and engages langerhans-t cell conjugates. *Mucosal immunology*, 3(5):506–522, 2010. doi: 10.1038/mi.2010.32. URL https://doi.org/10.1038/mi.2010.32.

[10] Florian Hladik, Polachai Sakchalathorn, Lamar Ballweber, Gretchen M. Lentz, Michael F. Fialkow, David A. Eschenbach, and M. Juliana McElrath. Initial events in establishing vaginal entry and infection by human immunodeficiency virus type-1. *Immunity*, 26(2):257–270, 2007. doi: 10.1016/j.immuni.2007.01.007. URL https://doi.org/10.1016/j.immuni.2007.01.007.

[11] Jake W. Rhodes, Rachel A. Botting, Kirstie M. Bertram, Erica E. Vine, Hafsa Rana, Heeva Baharlou, Peter Vegh, Thomas R. O'Neil, Anneliese S. Ashhurst, James Fletcher, Grant P Parnell, J. Dinny Graham, Najla Nasr, Jake J. K. Lim, Laith Barnouti, Peter A. Haertsch, Martijn P. Gosselink, Angelina Di Re, Faizur Reza, Grahame Ctercteko, Gregory Jenkins, Andrew J. Brooks, Ellis Patrick, Scott N. Byrne, Eric Hunter, Muzlifah Haniffa, Anthony L. Cunningham, and Andrew N. Harman. Human anogenital monocyte-derived dendritic cells and langerin+cdc2 are major hiv target cells. *Nature communications*, 12(1):2147–2147, 2021. doi: 10.1038/s41467-021-22375-x. URL https://doi.org/10.1038/s41467-021-22375-x.

[12] Radiana Trifonova, Brooke Bollman, Natasha S. Barteneva, and Judy Lieberman. Myeloid cells in intact human cervical explants capture hiv and can transmit it to cd4 t cells. *Frontiers in immunology*, 9(NA):2719–2719, 2018. doi: 10.3389/fimmu.2018.02719. URL https://doi.org/10.3389/fimmu.2018.02719.

[13] Claire Deleage, Stephen W. Wietgrefe, Gregory Q. Del Prete, David R. Morcock, Xing Pei Hao, Michael Piatak, Julian W. Bess, Jodi Anderson, Katherine Perkey, Cavan S. Reilly, Joseph M. McCune, Ashley T. Haase, Jeffrey D. Lifson, Timothy W. Schacker, and Jacob D. Estes. Defining hiv and siv reservoirs in lymphoid tissues. *Pathogens & immunity*, 1(1):68–106, 2016.

[14] Heeva Baharlou, Nicolas P Canete, Kirstie M. Bertram, Kerrie J Sandgren, Anthony L. Cunningham, Andrew N. Harman, and Ellis Patrick. Afid: a tool for automated identification and exclusion of autofluorescent objects from microscopy images. *Bioinformatics (Oxford, England)*, 37(4):559–567, 2020. doi: 10.

1093/bioinformatics/btaa780. URL https://doi.org/10.1093/bioinformatics/btaa780.

[15] Nicolas P Canete, Sourish S Iyengar, John T Ormerod, Heeva Baharlou, Andrew N Harman, and Ellis Patrick. spicyr: spatial analysis of in situ cytometry data in r. *Bioinformatics (Oxford, England)*, 38(11):3099–3105, 2022. doi: 10.1093/bioinformatics/btac268. URL https://doi.org/10.1093/bioinformatics/btac268.

[16] Jonas Wizenty, Muhammad Imtiaz Ashraf, Nadine Rohwer, Martin Stockmann, Sascha Weiss, Matthias Biebl, Johann Pratschke, Felix Aigner, and Tilo Wuensch. Autofluorescence: A potential pitfall in immunofluorescence-based inflammation grading. *Journal of immunological methods*, 456(NA):28–37, 2018. doi: 10.1016/j.jim.2018.02.007. URL https://doi.org/10.1016/j.jim.2018.02.007.

[17] Zahra Ahmed, Tatsuyoshi Kawamura, Shinji Shimada, and Vincent Piguet. The role of human dendritic cells in hiv-1 infection. *The Journal of investigative dermatology*, 135(5):1225–1233, 2014. doi: 10.1038/jid.2014.490. URL https://doi.org/10.1038/jid.2014.490.

[18] Kirstie M. Bertram, Orion Tong, Caroline Royle, Stuart Turville, Najla Nasr, Anthony L. Cunningham, and Andrew N. Harman. Manipulation of mononuclear phagocytes by hiv: Implications for early transmission events. *Frontiers in immunology*, 10(NA):2263–2263, 2019. doi: 10.3389/fimmu.2019.02263. URL https://doi.org/10.3389/fimmu.2019.02263.

[19] Rachel A. Botting, Hafsa Rana, Kirstie M. Bertram, Jake W. Rhodes, Heeva Baharlou, Najla Nasr, Anthony L. Cunningham, and Andrew N. Harman. Langerhans cells and sexual transmission of hiv and hsv. *Reviews in medical virology*, 27(2):e1923–NA, 2017. doi: 10.1002/rmv.1923. URL https://doi.org/10.1002/rmv.1923.

[20] Jake W. Rhodes, Orion Tong, Andrew N. Harman, and Stuart Turville. Human dendritic cell subsets, ontogeny, and impact on hiv infection. *Frontiers in immunology*, 10(NA):1088–1088, 2019. doi: 10.3389/fimmu.2019.01088. URL https://doi.org/10.3389/fimmu.2019.01088.

[21] Lucie Bracq, Maorong Xie, Serge Benichou, and Jérôme Bouchet. Mechanisms for cell-to-cell transmission of hiv-1. *Frontiers in immunology*, 9(NA):260–260, 2018. doi: 10.3389/fimmu.2018.00260. URL https://doi.org/10.3389/fimmu.2018.00260.

[22] Eduardo Garcia, Marjorie Pion, Annegret Pelchen-Matthews, Lucy M. Collinson, Jean-François Arrighi, Guillaume Blot, Florence Leuba, Jean-Michel Escola, Nicolas Demaurex, Mark Marsh, and Vincent Piguet. Hiv-1 trafficking to the dendritic cell-t-cell infectious synapse uses a pathway of tetraspanin sorting to the immunological synapse. *Traffic (Copenhagen, Denmark)*, 6(6):488–501, 2005. doi: 10.1111/j.1600-0854.2005.00293.x. URL https://doi.org/10.1111/j.1600-0854.2005.00293.x.

[23] David McDonald, Li Wu, Stacy M. Bohks, Vineet N. KewalRamani, Derya Unutmaz, and Thomas J. Hope. Recruitment of hiv and its receptors to dendritic cell-t cell junctions. *Science (New York, N.Y.)*, 300(5623):1295–1297, 2003. doi: 10.1126/science.1084238. URL https://doi.org/10.1126/science.1084238.

[24] Jian-Hua Wang, Alicia M. Janas, Wendy J. Olson, and Li Wu. Functionally distinct transmission of human immunodeficiency virus type 1 mediated by immature and mature dendritic cells. *Journal of virology*, 81(17):8933–8943, 2007. doi: 10.1128/jvi.00878-07. URL https://doi.org/10.1128/jvi.00878-07.

[25] Jian-Hua Wang, Clive Wells, and Li Wu. Macropinocytosis and cytoskeleton contribute to dendritic cell-mediated hiv-1 transmission to cd4+ t cells. *Virology*, 381(1):143–154, 2008. doi: 10.1016/j.virol.2008.08.028. URL https://doi.org/10.1016/j.virol.2008.08.028.

[26] Hyun Jae Yu, Morgan A. Reuter, and David McDonald. Hiv traffics through a specialized, surface-accessible intracellular compartment during trans-infection of t cells by mature dendritic cells. *PLoS pathogens*, 4(8):e1000134–NA, 2008. doi: 10.1371/journal.ppat.1000134. URL https://doi.org/10.1371/journal.ppat.1000134.

[27] Denis Schapiro, Hartland W. Jackson, Swetha Raghuraman, Jana R. Fischer, Vito R. T Zanotelli, Daniel Schulz, Charlotte Giesen, Raul Catena, Zsuzsanna Varga, and Bernd Bodenmiller. histocat: analysis of cell phenotypes and interactions in multiplex image cytometry data. *Nature methods*, 14(9):873–876, 2017. doi: 10.1038/nmeth.4391. URL https://doi.org/10.1038/nmeth.4391.

[28] Edward M. Campbell, Omar Perez, Marta Melar, and Thomas J. Hope. Labeling hiv-1 virions with two fluorescent proteins allows identification of virions that have productively entered the target cell. *Virology*, 360(2):286–293, 2006. doi: 10.1016/j.virol.2006.10.025. URL https://doi.org/10.1016/j.virol.2006.10.025.

[29] Eliseo A. Eugenin and Joan W. Berman. Improved methods to detect low levels of hiv using antibody-based technologies. *Methods in molecular biology (Clifton, N.J.)*, 1354(NA):265–279, 2016. doi: 10.1007/978-1-4939-3046-3_18. URL https://doi.org/10.1007/978-1-4939-3046-3_18.

[30] Lisa Prevedel, Nancy Ruel, Paul Castellano, Carla Smith, Shaily Malik, Courtney Villeux, Morgane Bomsel, Susan Morgello, and Eliseo A. Eugenin. Identification, localization, and quantification of hiv reservoirs using microscopy. *Current protocols in cell biology*, 82(1):e64–NA, 2018. doi: 10.1002/cpcb.64. URL https://doi.org/10.1002/cpcb.64.

[31] Eliseo A. Eugenin, Peter J. Gaskill, and Joan W. Berman. Tunneling nanotubes (tnt) are induced by hiv-infection of macrophages: a potential mechanism for intercellular hiv trafficking. *Cellular immunology*, 254(2):142–148, 2008. doi: 10.1016/j.cellimm.2008.08.005. URL https://doi.org/10.1016/j.cellimm.2008.08.005.

[32] Damjan S. Nikolic, Martin Lehmann, Richard L. Felts, Eduardo Garcia, Fabien Blanchet, Sriram Subramaniam, and Vincent Piguet. Hiv-1 activates cdc42 and induces membrane extensions in immature dendritic cells to facilitate cell-to-cell virus propagation. *Blood*, 118(18):4841–4852, 2011. doi: 10.1182/blood-2010-09-305417. URL https://doi.org/10.1182/blood-2010-09-305417.

[33] Kevin B. Gurney, Julie Elliott, Hoorig Nassanian, Carol Song, Elizabeth J. Soilleux, Ian McGowan, Peter A. Anton, and Benhur Lee. Binding and transfer of human immunodeficiency virus by dc-sign+ cells in human rectal mucosa. *Journal of virology*,

79(9):5762–5773, 2005. doi: 10.1128/jvi.79.9.5762-5773.2005. URL https://doi.org/10.1128/jvi.79.9.5762-5773.2005.

[34] Mariangela Cavarelli, Chiara Foglieni, Maria Rescigno, and Gabriella Scarlatti. R5 hiv-1 envelope attracts dendritic cells to cross the human intestinal epithelium and sample luminal virions via engagement of the ccr5. *EMBO molecular medicine*, 5(5): 776–794, 2013. doi: 10.1002/emmm.201202232. URL https://doi.org/10.1002/emmm.201202232.

[35] Andrea Imle, Peter Kumberger, Nikolas D Schnellbächer, Jana Fehr, Paola Carrillo-Bustamante, Janez Ales, Philip Schmidt, Christian Ritter, William J. Godinez, Barbara Müller, Karl Rohr, Fred A. Hamprecht, Ulrich S. Schwarz, Frederick Graw, and Oliver T. Fackler. Experimental and computational analyses reveal that environmental restrictions shape hiv-1 spread in 3d cultures. *Nature communications*, 10(1):2144–2144, 2019. doi: 10.1038/s41467-019-09879-3. URL https://doi.org/10.1038/s41467-019-09879-3.

[36] Z Zhou, Lin Xu, Alexis Sennepin, C Federici, Yonatan Ganor, Daniela Tudor, Diane Damotte, N. Barry Delongchamps, Marc Zerbib, and Morgane Bomsel. The hiv-1 viral synapse signals human foreskin keratinocytes to secrete thymic stromal lymphopoietin facilitating hiv-1 foreskin entry. *Mucosal immunology*, 11(1):158–171, 2017. doi: 10.1038/mi.2017.23. URL https://doi.org/10.1038/mi.2017.23.

[37] Diana Domanska, Umair Majid, Victoria T Karlsen, Marianne A Merok, Ann-Christin Røberg Beitnes, Sheraz Yaqub, Espen S Bækkevold, and Frode L. Jahnsen. Single-cell transcriptomic analysis of human colonic macrophages reveals niche-specific subsets. *The Journal of experimental medicine*, 219(3):NA–NA, 2022. doi: 10.1084/jem.20211846. URL https://doi.org/10.1084/jem.20211846.

[38] Johnny Bonnardel, Clément Da Silva, Sandrine Henri, Samira Tamoutounour, Lionel Chasson, Frederic Montañana-Sanchis, Jean-Pierre Gorvel, and Hugues Lelouard. Innate and adaptive immune functions of peyer's patch monocyte-derived cells. *Cell reports*, 11(5):770–784, 2015. doi: 10.1016/j.celrep.2015.03.067. URL https://doi.org/10.1016/j.celrep.2015.03.067.

[39] Nobuhide Kobayashi, Daisuke Takahashi, Shunsuke Takano, Shunsuke Kimura, and Koji Hase. The roles of peyer's patches and microfold cells in the gut immune system: Relevance to autoimmune diseases. *Frontiers in immunology*, 10(NA):2345–2345, 2019. doi: 10.3389/fimmu.2019.02345. URL https://doi.org/10.3389/fimmu.2019.02345.

[40] JM Langman and R Rowland. The number and distribution of lymphoid follicles in the human large intestine. *Journal of anatomy*, 149:189, 1986.

[41] Tae Wook Chun, David C. Nickle, Jesse S. Justement, Jennifer Hartt Meyers, Gregg Roby, Claire W. Hallahan, Shyam Kottilil, Susan Moir, JoAnn M. Mican, James I. Mullins, Douglas J. Ward, Joseph A. Kovacs, Peter J. Mannon, and Anthony S. Fauci. Persistence of hiv in gut-associated lymphoid tissue despite long-term antiretroviral therapy. *The Journal of infectious diseases*, 197(5):714–720, 2008. doi: 10.1086/527324. URL https://doi.org/10.1086/527324.

[42] Thomas M. Fenton, Peter Bjørn Jørgensen, Kristoffer Niss, Samuel J. S Rubin, Urs M. Mörbe, Lene Riis, Clément Da Silva, Adam W Plumb, Julien Vandamme, Henrik Loft Jakobsen, Søren Brunak, Aida Habtezion, Ole Haagen Nielsen, Bengt Johansson-Lindbom, and William W. Agace. Immune profiling of human gut-associated lymphoid tissue identifies a role for isolated lymphoid follicles in priming of region-specific immunity. *Immunity*, 52(3):557–570, 2020. doi: 10.1016/j.immuni.2020.02.001. URL https://doi.org/10.1016/j.immuni.2020.02.001.

[43] Paul U. Cameron, Suha Saleh, Georgina Sallmann, Ajantha Solomon, Fiona Wightman, Vanessa A. Evans, Geneviève Boucher, Elias K. Haddad, Rafick-Pierre Sekaly, Andrew N. Harman, Jenny L. Anderson, Kate L. Jones, Johnson Mak, Anthony L. Cunningham, Anthony Jaworowski, and Sharon R. Lewin. Establishment of hiv-1 latency in resting cd4+ t cells depends on chemokine-induced changes in the actin cytoskeleton. *Proceedings of the National Academy of Sciences of the United States of America*, 107(39):16934–16939, 2010. doi: 10.1073/pnas.1002894107. URL https://doi.org/10.1073/pnas.1002894107.

[44] Wan Hon Koh, Paul Lopez, Oluwaseun Ajibola, Roshan Parvarchian, Umar Mohammad, Ryan Hnatiuk, Jason Kindrachuk, and Thomas T. Murooka. Hiv-captured dcs regulate t cell migration and cell-cell contact dynamics to enhance viral spread. *iScience*, 23(8):101427–NA, 2020. doi: 10.1016/j.isci.2020.101427. URL https://doi.org/10.1016/j.isci.2020.101427.

[45] Constance Renault, Nicolas Veyrenche, Franck Mennechet, Anne-Sophie Bedin, Jean-Pierre Routy, Philippe Van de Perre, Jacques Reynes, and Edouard Tuaillon. Th17 cd4+ t-cell as a preferential target for hiv reservoirs. *Frontiers in immunology*, 13 (NA):822576–NA, 2022. doi: 10.3389/fimmu.2022.822576. URL https://doi.org/10.3389/fimmu.2022.822576.

[46] JOSEPH L Unthank and H GLENN Bohlen. Lymphatic pathways and role of valves in lymph propulsion from small intestine. *American Journal of Physiology-Gastrointestinal and Liver Physiology*, 254(3):G389–G398, 1988.

[47] Phalguni Gupta, Kelly B. Collins, Deena Ratner, Simon C. Watkins, Gregory J. Naus, Daniel V. Landers, and Bruce K. Patterson. Memory cd4+ t cells are the earliest detectable human immunodeficiency virus type 1 (hiv-1)-infected cells in the female genital mucosal tissue during hiv-1 transmission in an organ culture system. *Journal of virology*, 76(19):9868–9876, 2002. doi: 10.1128/jvi.76.19.9868-9876.2002. URL https://doi.org/10.1128/jvi.76.19.9868-9876.2002.

[48] Richard L. Felts, Kedar Narayan, Jacob D. Estes, Dan Shi, Charles M. Trubey, Jing Fu, Lisa M. Hartnell, Gordon Ruthel, Douglas K. Schneider, Kunio Nagashima, Julian W. Bess, Sina Bavari, Bradley C. Lowekamp, Donald Bliss, Jeffrey D. Lifson, and Sriram Subramaniam. 3d visualization of hiv transfer at the virological synapse between dendritic cells and t cells. *Proceedings of the National Academy of Sciences of the United States of America*, 107(30):13336–13341, 2010. doi: 10.1073/pnas.1003040107. URL https://doi.org/10.1073/pnas.1003040107.

[49] Najla Nasr, Joey Lai, Rachel A. Botting, Sarah K. Mercier, Andrew N. Harman, Min Kim, Stuart Turville, Teresa Domagala, Paul R Gorry, Norman Olbourne, and Anthony L. Cunningham. Inhibition of two temporal phases of hiv-1 transfer from

primary langerhans cells to t cells: The role of langerin. *Journal of immunology (Baltimore, Md. : 1950)*, 193(5):2554–2564, 2014. doi: 10.4049/jimmunol.1400630. URL https://doi.org/10.4049/jimmunol.1400630.

[50] Heeva Baharlou, Nicolas P Canete, Anthony L. Cunningham, Andrew N. Harman, and Ellis Patrick. Mass cytometry imaging for the study of human diseases-applications and data analysis strategies. *Frontiers in immunology*, 10(NA):2657–2657, 2019. doi: 10.3389/fimmu.2019.02657. URL https://doi.org/10.3389/fimmu.2019.02657.

[51] Sabrina M. Lewis, Marie Liesse Asselin-Labat, Quan Nguyen, Jean Berthelet, Xiao Tan, Verena C. Wimmer, Delphine Merino, Kelly L. Rogers, and Shalin H. Naik. Spatial omics and multiplexed imaging to explore cancer biology. *Nature methods*, 18(9):997–1012, 2021. doi: 10.1038/s41592-021-01203-6. URL https://doi.org/10.1038/s41592-021-01203-6.

[52] Jihoon Kim, Bon-Kyoung Koo, and Juergen A. Knoblich. Human organoids: model systems for human biology and medicine. *Nature reviews. Molecular cell biology*, 21(10):571–584, 2020. doi: 10.1038/s41580-020-0259-3. URL https://doi.org/10.1038/s41580-020-0259-3.

[53] Amy Gillgrass, Jocelyn M. Wessels, Jack X. Yang, and Charu Kaushic. Advances in humanized mouse models to improve understanding of hiv-1 pathogenesis and immune responses. *Frontiers in immunology*, 11(NA):617516–617516, 2021. doi: 10.3389/fimmu.2020.617516. URL https://doi.org/10.3389/fimmu.2020.617516.

[54] Chloe M Doyle, Erica E. Vine, Kirstie M. Bertram, Heeva Baharlou, Jake W. Rhodes, Suat Dervish, Martijn P. Gosselink, Angelina Di Re, Geoffrey P. Collins, Faizur Reza, James Wei Tatt Toh, Nimalan Pathma-Nathan, Golo Ahlenstiel, Grahame Ctercteko, Anthony L. Cunningham, Andrew N. Harman, and Scott N. Byrne. Optimal isolation protocols for examining and interrogating mononuclear phagocytes from human intestinal tissue. *Frontiers in immunology*, 12(NA):727952–NA, 2021. doi: 10.3389/fimmu.2021.727952. URL https://doi.org/10.3389/fimmu.2021.727952.

[55] Christopher D. Pilcher, George Joaki, Irving F. Hoffman, Francis Martinson, Clement Mapanje, Paul W. Stewart, Kimberly A. Powers, Shannon Galvin, David Chilongozi, Syze Gama, Matthew Price, Susan A. Fiscus, and Myron S. Cohen. Amplified transmission of hiv-1: comparison of hiv-1 concentrations in semen and blood during acute and chronic infection. *AIDS (London, England)*, 21(13):1723–1730, 2007. doi: 10.1097/qad.0b013e3281532c82. URL https://doi.org/10.1097/qad.0b013e3281532c82.

[56] Nico Battich, Thomas Stoeger, and Lucas Pelkmans. Image-based transcriptomics in thousands of single human cells at single-molecule resolution. *Nature methods*, 10(11):1127–1133, 2013. doi: 10.1038/nmeth.2657. URL https://doi.org/10.1038/nmeth.2657.

[57] Anne E. Carpenter, Thouis R. Jones, Michael R. Lamprecht, Colin Clarke, In Han Kang, Ola Friman, David A. Guertin, Joo Han Chang, Robert A. Lindquist, Jason Moffat, Polina Golland, and David M. Sabatini. Cellprofiler: image analysis software for identifying and quantifying cell phenotypes. *Genome biology*, 7(10):

1–11, 2006. doi: 10.1186/gb-2006-7-10-r100. URL https://doi.org/10.1186/gb-2006-7-10-r100.

[58] Reuben M. Baron and David A. Kenny. The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of personality and social psychology*, 51(6):1173–1182, 1986. doi: 10.1037/0022-3514.51.6.1173. URL https://doi.org/10.1037/0022-3514.51.6.1173.

[59] David P. MacKinnon, Jennifer L. Krull, and Chondra M. Lockwood. Equivalence of the mediation, confounding and suppression effect. *Prevention science : the official journal of the Society for Prevention Research*, 1(4):173–181, 2000. doi: 10.1023/a:1026595011371. URL https://doi.org/10.1023/a:1026595011371.

[60] Nicolas Damond, Stefanie Engler, Vito R. T Zanotelli, Denis Schapiro, Clive Wasserfall, Irina Kusmartseva, Harry S. Nick, Fabrizio Thorel, Pedro Luis Herrera, Mark A. Atkinson, and Bernd Bodenmiller. A map of human type 1 diabetes progression by imaging mass cytometry. *Cell metabolism*, 29(3):755–768, 2019. doi: 10.1016/j.cmet.2018.11.014. URL https://doi.org/10.1016/j.cmet.2018.11.014.

[61] Hartland W. Jackson, Jana R. Fischer, Vito R. T Zanotelli, H. Raza Ali, Robert Mechera, Savas D. Soysal, Holger Moch, Simone Muenst, Zsuzsanna Varga, Walter P. Weber, and Bernd Bodenmiller. The single-cell pathology landscape of breast cancer. *Nature*, 578(7796):615–620, 2020. doi: 10.1038/s41586-019-1876-x. URL https://doi.org/10.1038/s41586-019-1876-x.

[62] Leeat Keren, Marc Bosse, Diana M. Marquez, Roshan Angoshtari, Samir Jain, Sushama Varma, Soo-Ryum Yang, Allison W. Kurian, David Van Valen, Robert B. West, Sean C. Bendall, and Michael Angelo. A structured tumor-immune microenvironment in triple negative breast cancer revealed by multiplexed ion beam imaging. *Cell*, 174(6):1373–1387, 2018. doi: 10.1016/j.cell.2018.08.039. URL https://doi.org/10.1016/j.cell.2018.08.039.

Chapter 5

Discussion

Understanding the early events that facilitate the transmission of HIV is important for the development of strategies for disease prevention. However, due to technical limitations, these events are still poorly understood. Through access to human tissue and real-world HIV strains, as well as experimental innovations such as cyclic immunofluorescence (CyCIF) and RNA scope, the host lab is in a unique position to visualize early HIV-cell interactions. In this thesis, HIV-cell interactions were able to be visualized as early as 30 minutes in human colorectal tissue, a key site for HIV transmission. However, the complexity of the data necessitated the need for image analysis strategies in order quantitatively analyze the unique cellular interactions that underpin early HIV transmission.

Hence, this thesis had two important aims. The main biological aim of this thesis was to develop a computational pipeline for analyzing high parameter CyCIF images of early HIV transmission in human colorectal tissue. The secondary aim of this project was to develop new computational tools to facilitate the analysis of high parameter microscopy images.

The main biological aim of this thesis was addressed through the development of a complete image analysis pipeline for the analysis of CyCIF images of early HIV transmission in colorectal tissue (Chapter 4). This pipeline was able to identify the unique cellular interactions that underpin early HIV transmission, and can serve as a framework for the study of human disease in the context of imaging host-cell interactions. The secondary aim was achieved through the development of 'Autofluorescence Identifier' (AFid, Chapter 2), a simple tool for removing autofluorescence from fluorescent microscopy images, and the development of 'SPatial analysis of In situ Cytometry data in R' (spicyR, Chapter 3), a tool which facilitates inference on changes in spatial co-localizations. While motivated by the analysis of the HIV CyCIF data, both tools were published and made readily available given their applicability to other high parameter image datasets. With the development of mass cytometry imaging (MCI) and serial staining immunofluorescence techniques being more widely used in the study of human disease, it has become increasingly important to develop analysis strategies given the complexity of the data. Thus, the AFid and spicyR serve to fill gaps in the analysis currently present in high parameter imaging.

Overall, this thesis presents three publications addressing these aims. While extensive discussions have already been provided, this section will discuss these publications within the context of the overall thesis.

5.1 AFid

The first challenge that was present in the analysis of the HIV CyCIF data was the presence of autofluorescence, with these signals being difficult to distinguish from real immunofluorescent staining. While strategies for addressing autofluorescence have been previously described, chemical bleaching methods were incompatible with our staining protocols [1–4], and the tools required for spectral unmixing [5] or background subtraction [6–8] were not readily available. To address this, we developed AFid, a post-acquisition algorithm that identifies and removes autofluorescence from fluorescent images.

The AFid algorithm was constructed around the assumption that the textural features of autofluorescent signals are conserved across different image channels, occupying a unique position in feature space that is then identified through clustering (Supplementary Figure 2.3A). Through experimental perturbations and *in silico* simulation we illustrated that AFid can distinguish autofluorescence from real signal given two fluorescent image channels.

In the analysis of the HIV CyCIF dataset, there were significant differences in results before and after the application of AFid. Importantly, it was found that the number of dendritic cells that were identified to be HIV positive was greatly overestimated due to autofluorescent signal (Figure 2.3). Ultimately, AFid allowed for the accurate phenotyping of cells to facilitate a more robust analysis of fluorescent image data. This is increasingly important particularly as high parameter fluorescent microscopy techniques continue to improve. Hence, AFid was extensively used in the further analysis of the HIV CyCIF dataset.

Furthermore, AFid has also been applied to a range of datasets since it became publicly available. The algorithm has been applied to fluorescent images of pancreatic tissue exploring the immunology of type 1 diabetes [9] and fluorescent images of olfactory epithelial tissue exploring COVID-19-induced anosmia [10]. These publications illustrate that AFid can be utilized in a diverse range of tissue types to assist in the visualization of different diseases.

Overall, AFid addresses a key problem in fluorescent image analysis. A key advantage over other methods is that AFid does not require additional experimental steps, being applied post-acquisition. It can be applied to existing datasets and does not require the use of specialist equipment or proprietary software for use. AFid has been implemented as an easy-to-use tool in ImageJ, MATLAB and R (`https://ellispatrick.github.io/AFid`), allowing for its use across different analysis environments.

5.2 spicyR

A key challenge in the analysis of high parameter images is understanding the important cell-cell interactions within an image. Various neighbourhood approaches have been developed for identifying spatial association or avoidance of cells [11–13]. However, there did not exist any robust methods for identifying differential cell-type co-localization across clinical or experimental groups. To this end, we developed the R package spicyR, a tool which facilitates inference on changes in spatial co-localization between cell types. Briefly, spicyR was designed to answer the question "How does cell-type co-localization change with different experimental or clinical conditions?".

The first innovation of spicyR is the development of the co-localization score u (Equation 3.3), providing an interpretable and comparable measure of spatial co-localization between

two cell types within a single image. Next, spicyR is able to model information across multiple images per subject and account for large differences in cell numbers across images, implementing a random effects model and a weighting scheme based on cell counts.

Overall, spicyR provides an interpretable and easy-to-use tool for spatial analysis, adding to the repertoire of analysis techniques that can be applied to high parameter imaging data. However, it should be acknowledged that while spicyR may be a key step in image analysis, it should not be interpreted in isolation. The results of spicyR should be understood within the context of other simple analysis techniques, such as cell-type composition and cell marker analysis.

To highlight its utility, spicyR was able to be applied to an existing type 1 diabetes IMC dataset [14], revealing both novel findings as well as findings in line with the original publication (Figure 3.4). Since the release of the R package on Bioconductor, spicyR has also been applied to cutaneous squamous cell carcinoma images [15], chronic rhinosinusitis images [16], and was utilized as a key step in the spatial analysis of the HIV CyCIF dataset, identifying key cell-cell interactions (Figure 4.6). Thus, spicyR is demonstratively applicable for investigating images from diverse disease environments.

5.3 Analysis of HIV CyCIF Dataset

While this thesis is primarily focused on the development of computational methods, we first briefly describe the novel aspects of the experimental design presented in the publication which motivated the analysis approaches developed. Through collaborations with colorectal surgeons within the Westmead Hospital Precinct, the host lab was able to collect healthy human colorectal tissue and apply HIV within less than an hour. Furthermore through an international collaboration with Professor Eric Hunter (Emory University, USA), the host lab gained access to transmitted founder strains of HIV, representing a strain derived from a real-world HIV transmission event. This places the host lab in the unique position of modelling HIV transmission directly in the tissue of interest. Next, the optimization of RNAscope, an *in situ* hybridization technique, allowed for HIV to be visualized at very early timepoints with single virion precision, allowing their interaction with key HIV target cells to be identified. Finally, the use of CyCIF has allowed multiple markers to be visualized within colorectal tissue. This allowed for the visualization of key HIV target cells, namely dendritic cells, macrophages and CD4 T cells, HIV via RNAscope, and structural markers for identifying key structural compartments within colorectal tissue. While the experimental innovations pioneered by the host lab provide a unique opportunity for investigating early HIV transmission, the complexity of the data presents an important challenge regarding the robust quantitative analysis of the imaging data. To address this, a novel image analysis pipeline was developed, utilizing ImageJ, MATLAB and R. This section will hence discuss the computational innovations presented in this publication.

One of the key computational innovations present was the development of a novel MATLAB image processing pipeline, with key functions including performing single-cell segmentation and cell-type classification, HIV spot counting, and important image data quantification for further downstream analysis. Firstly, a custom cell segmentation and classification strategy was developed (Supplementary Figure 4.1B). This unique approach resulted in the segmentation of the full cell body of key HIV target cells and, unlike other segmentation pipelines, allowed for two cells to physically overlap. This allowed for the complex amorphous shape of cells such as macrophages and DCs to be identified and in

turn allowing cell-cell and cell-HIV interactions to be identified precisely. The combined segmentation and classification approach led to a quantification of the cellular composition of colorectal tissue, and together with RNA spot counting allowed the number of HIV virus particles to be measured. Furthermore, to facilitate spatial analysis, a modified version of the neighbourhood analysis by Schapiro et al. [11] was implemented, allowing for its use with the new cell segmentation masks being generated. From this neighbourhood analysis, a HIV-transfer phenotype score was derived (Figure 4.6C), allowing the viral transfer between cells to be estimated. Finally, the distance of cells and HIV with respect to the tissue compartments segmented in ImageJ (Figure 4.4).

Another key computational innovation was the downstream analysis techniques that were implemented in R, facilitating the quantitative analysis of the image data extracted with ImageJ and MATLAB. Specifically, this step of analysis allowed for the quantification and visualization of: 1) the distribution of HIV target cells (Figure 4.2) and HIV (Figure 4.4) in human colorectal tissue, HIV-cell interactions (Figures 4.3 and 4.5), and cell-cell interactions (Figure 4.6). Such analysis resulted in the quantification of cell- and tissue compartment-specific enrichment of HIV.

Through both experimental and computational innovations, cellular events leading to early HIV transmission were able to be visualized and interrogated. In brief, these results revealed early HIV sampling by dendritic cells and macrophages, as well as early presentation of HIV from these cells to CD4 T cells, forming clusters that may facilitate early viral transfer and enhance T cell infection. Furthermore, analysis also revealed trafficking of HIV by dendritic cells to lymphoid aggregates, facilitating rapid entry of HIV into deeper structures of colorectal tissue. Thus, these results present the first quantitative evidence of the early events of HIV transmission in human colorectal tissue *in situ*.

We expect that the tools developed in this paper can be applicable in the study of other pathogens in different *in situ* models and clinical scenarios. As such we have made the code used for analysis readily available (Chapter 4).

5.4 Concluding Remarks and Future Directions

Through collaborations with surgeons, as well as novel techniques in the experimental design, the host lab is in a unique position to be able to visualize and describe the events which facilitate early HIV transmission. This has resulted in the generation of complex high parameter images, necessitating the need for the development of image analysis techniques. Such an approach was important in order to quantitatively analyze the complex relationship between HIV and the mucosal immune system during early HIV transmission within colorectal tissue. These innovations have ultimately inspired the work presented within this thesis.

Ultimately, through the development of this novel analysis pipeline, key cell-cell, HIV-cell and HIV-tissue relationships important for early HIV transmission were able to be identified for the first time *in situ*. Furthermore, there is an opportunity for the analysis techniques presented here to be applied to other datasets, particularly those looking at host-pathogen interactions. As such, the code has been made publicly available. Overall, analysis of the CyCIF dataset generated by the host lab has led to the development of novel image analysis approaches that has potential for use in other high parameter image analysis.

While the analysis techniques here may be applied to a diverse range of high parameter imaging datasets, a major limitation is that these tools do not exist within a single

analysis environment. At present, there are multiple steps that currently take place in either ImageJ, MATLAB and R. This may restrict the ease of use of the tools developed within this thesis, and makes the analysis less streamlined. As such, an important future direction currently being undertaken is to synthesize the various tools used for image analysis within a single analysis environment; for our group that is R. This will include steps such as single-cell segmentation, cell-type clustering, and spatial analysis, allowing high parameter image datasets to be comprehensively analyzed within a single analysis environment.

Finally, while CyCIF was utilized in generating the HIV CyCIF images, these images did not employ the large number of cell markers that is generally available in multiplexed imaging technologies. While the major HIV target cell types were identified, specific cell subsets were excluded from the analysis. To expand on the results presented within this thesis, the main future direction underway is the generation of a 40 marker IMC panel. The images generated via IMC will employ the same experimental innovations presented in this thesis. A key difference will be the significantly increased number of markers used, allowing HIV interactions with specific cell subsets to be identified. This will provide a more comprehensive view of early HIV transmission to be captured. Furthermore, the increased complexity of the images generated will necessitate the need for continued optimization of image analysis strategies. IMC will produces images with lower resolution than fluorescent imaging that displays more markers, captures a smaller tissue area, and with a unique noise profile. Thus, the analysis of the IMC datasets will present a unique challenge, with many avenues for the exploration of development of computational approaches to high parameter image analysis.

Bibliography

[1] Werner Baschong, Rosmarie Suetterlin, and R. Hubert Laeng. Control of autofluorescence of archival formaldehyde-fixed, paraffin-embedded tissue in confocal laser scanning microscopy (clsm). *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society*, 49(12):1565–1571, 2001. doi: 10.1177/002215540104901210. URL https://doi.org/10.1177/002215540104901210.

[2] Nicholas Billinton and Andrew W. Knight. Seeing the wood through the trees: a review of techniques for distinguishing green fluorescent protein from endogenous autofluorescence. *Analytical biochemistry*, 291(2):175–197, 2001. doi: 10.1006/abio.2000.5006. URL https://doi.org/10.1006/abio.2000.5006.

[3] A. Sally Davis, Anke Richter, Steven Becker, Jenna E. Moyer, Aline Sandouk, Jeff Skinner, and Jeffery K. Taubenberger. Characterizing and diminishing autofluorescence in formalin-fixed paraffin-embedded human respiratory tissue. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society*, 62(6):405–423, 2014. doi: 10.1369/0022155414531549. URL https://doi.org/10.1369/0022155414531549.

[4] Michael Neumann and Detlef Gabel. Simple method for reduction of autofluorescence in fluorescence microscopy. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society*, 50(3):437–439, 2002. doi: 10.1177/002215540205000315. URL https://doi.org/10.1177/002215540205000315.

[5] Mary E. Dickinson, Gregory H. Bearman, S Tille, Rusty Lansford, and Scott E. Fraser. Multi-spectral imaging and linear unmixing add a whole new dimension to laser scanning fluorescence microscopy. *BioTechniques*, 31(6):1272–1278, 2001. doi: 10.2144/01316bt01. URL https://doi.org/10.2144/01316bt01.

[6] Michael J. Gerdes, Christopher J. Sevinsky, Anup Sood, Sudeshna Adak, Musodiq Bello, Alexander Bordwell, Ali Can, Corwin Alex D, Sean Richard Dinn, Robert John Filkins, Denise Hollman, Vidya Pundalik Kamath, Sireesha Kaanumalle, Kevin Bernard Kenny, Melinda Larsen, Michael Lazare, Qing Li, Christina Lowes, Colin Craig McCulloch, Elizabeth McDonough, Michael Christopher Montalto, Zhengyu Pang, Jens Rittscher, Alberto Santamaria-Pang, Brion Daryl Sarachan, Maximilian Lewis Seel, Antti Seppo, Kashan Shaikh, Yunxia Sui, Jingyu Zhang, and Fiona Ginty. Highly multiplexed single-cell analysis of formalin-fixed, paraffin-embedded cancer tissue. *Proceedings of the National Academy of Sciences of the United States of America*, 110(29):11982–11987, 2013. doi: 10.1073/pnas.1300136110. URL https://doi.org/10.1073/pnas.1300136110.

[7] Zhengyu Pang, Eugene Barash, Alberto Santamaria-Pang, Christopher J. Sevinsky, Qing Li, and Fiona Ginty. Autofluorescence removal using a customized filter set. *Microscopy research and technique*, 76(10):1007–1015, 2013. doi: 10.1002/jemt.22261. URL https://doi.org/10.1002/jemt.22261.

[8] Franco Woolfe, Michael J. Gerdes, Musodiq Bello, Xiaodong Tao, and Ali Can. Autofluorescence removal by non-negative matrix factorization. *IEEE transactions on*

image processing : a publication of the IEEE Signal Processing Society, 20(4):1085–1093, 2010. doi: 10.1109/tip.2010.2079810. URL https://doi.org/10.1109/tip.2010.2079810.

[9] Yuelin Kong, Yi Jing, Denise Allard, Marissa A Scavuzzo, Maran L Sprouse, Malgorzata Borowiak, Matthew L Bettini, and Maria Bettini. A dormant t-cell population with autoimmune potential exhibits low self-reactivity and infiltrates islets in type 1 diabetes. *European Journal of Immunology*, 2022.

[10] Marianna Zazhytska, Albana Kodra, Daisy A Hoagland, Justin Frere, John F Fullard, Hani Shayya, Natalie G McArthur, Rasmus Moeller, Skyler Uhl, Arina D Omer, et al. Non-cell-autonomous disruption of nuclear architecture as a potential cause of covid-19-induced anosmia. *Cell*, 185(6):1052–1064, 2022.

[11] Denis Schapiro, Hartland W. Jackson, Swetha Raghuraman, Jana R. Fischer, Vito R. T Zanotelli, Daniel Schulz, Charlotte Giesen, Raul Catena, Zsuzsanna Varga, and Bernd Bodenmiller. histocat: analysis of cell phenotypes and interactions in multiplex image cytometry data. *Nature methods*, 14(9):873–876, 2017. doi: 10.1038/nmeth.4391. URL https://doi.org/10.1038/nmeth.4391.

[12] Leeat Keren, Marc Bosse, Diana M. Marquez, Roshan Angoshtari, Samir Jain, Sushama Varma, Soo-Ryum Yang, Allison W. Kurian, David Van Valen, Robert B. West, Sean C. Bendall, and Michael Angelo. A structured tumor-immune microenvironment in triple negative breast cancer revealed by multiplexed ion beam imaging. *Cell*, 174(6):1373–1387, 2018. doi: 10.1016/j.cell.2018.08.039. URL https://doi.org/10.1016/j.cell.2018.08.039.

[13] Adrian Baddeley, Ege Rubak, and Rolf Turner. *Spatial Point Patterns*, volume NA of *NA*. 2015. doi: 10.1201/b19708. URL https://doi.org/10.1201/b19708.

[14] Nicolas Damond, Stefanie Engler, Vito R. T Zanotelli, Denis Schapiro, Clive Wasserfall, Irina Kusmartseva, Harry S. Nick, Fabrizio Thorel, Pedro Luis Herrera, Mark A. Atkinson, and Bernd Bodenmiller. A map of human type 1 diabetes progression by imaging mass cytometry. *Cell metabolism*, 29(3):755–768, 2019. doi: 10.1016/j.cmet.2018.11.014. URL https://doi.org/10.1016/j.cmet.2018.11.014.

[15] Angela L Ferguson, Ashleigh R Sharman, Ruth O Allen, Thomas Ye, Jenny H Lee, Tsu-Hui Hubert Low, Sydney Ch'ng, Carsten E Palme, Bruce Ashford, Marie Ranson, et al. High-dimensional and spatial analysis reveals immune landscape dependent progression in cutaneous squamous cell carcinoma. *bioRxiv*, 2022.

[16] Gohar Shaghayegh, Clare Cooksley, George Spyro Bouras, Beula Subashini Panchatcharam, Rejhan Idrizi, Metta Jana, Sarah Ellis, Alkis James Psaltis, Peter-John Wormald, and Sarah Vreugde. Chronic rhinosinusitis patients display an aberrant immune cell localization with enhanced s aureus biofilm metabolic activity and biomass. *Journal of Allergy and Clinical Immunology*, 2022.