



Image-Based Cell Profiling Enables Quantitative Tissue Microscopy in Gastroenterology

John W. Wills,^{1*}  Jack Robertson,¹ Huw D. Summers,² Michelle Minitier,¹ Claire Barnes,² Rachel E. Hewitt,¹  Åsa V. Keita,³ Johan D. Söderholm,³ Paul Rees,^{2,4†} Jonathan J. Powell^{1†}

¹Biomaterials Research, Cambridge University Department of Veterinary Medicine, School of Biological Sciences, Cambridge, UK

²Centre for Nanohealth, Swansea University College of Engineering, Swansea, UK

³Department of Surgery and Department of Clinical and Experimental Medicine, Linköping University, Linköping, Sweden

⁴Broad Institute of MIT and Harvard, Cambridge, Massachusetts, 02142

Received 6 January 2020; Revised 29 April 2020; Accepted 4 May 2020

Grant sponsor: Biotechnology and Biological Sciences Research Council, Grant numberBB/P026818/1; Grant sponsor: Engineering and Physical Sciences Research Council, Grant numberEP/H008683/1; Grant sponsor: UK Medical Research Council, Grant numberMR/R005699/1; Grant sponsor: University of Cambridge Herchel-Smith Fellowship; Grant sponsor: Girton College, University of Cambridge

Additional Supporting Information may be found in the online version of this article.

*Correspondence to: John W. Wills, Biomaterials Research, Cambridge University Department of Veterinary Medicine, School of Biological Sciences, Madingley Road, Cambridge CB3 0ES, UK Email: jw2020@cam.ac.uk

† These authors contributed equally to this study. Grant sponsor: UK Medical Research Council, Grant number: MR/R005699/1; Grant sponsor: UK Engineering and Physical Sciences Research Council, Grant number: EP/H008683/1; Grant sponsor: UK Biotechnology and Biological Sciences Research Council, Grant number: BB/P026818/1; Grant sponsor: Girton College, University of Cambridge; Grant sponsor: University of Cambridge Herchel-Smith Fund.

Published online 23 May 2020 in Wiley Online Library (wileyonlinelibrary.com)

DOI: 10.1002/cyto.a.24042

© 2020 The Authors. *Cytometry Part A* published by Wiley Periodicals LLC on behalf of International Society for Advancement of Cytometry.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

• Abstract

Immunofluorescence microscopy is an essential tool for tissue-based research, yet data reporting is almost always qualitative. Quantification of images, at the per-cell level, enables “flow cytometry-type” analyses with intact locational data but achieving this is complex. Gastrointestinal tissue, for example, is highly diverse: from mixed-cell epithelial layers through to discrete lymphoid patches. Moreover, different species (e.g., rat, mouse, and humans) and tissue preparations (paraffin/frozen) are all commonly studied. Here, using field-relevant examples, we develop open, user-friendly methodology that can encompass these variables to provide quantitative tissue microscopy for the field. Antibody-independent cell labeling approaches, compatible across preparation types and species, were optimized. Per-cell data were extracted from routine confocal micrographs, with semantic machine learning employed to tackle densely packed lymphoid tissues. Data analysis was achieved by flow cytometry-type analyses alongside visualization and statistical definition of cell locations, interactions and established microenvironments. First, quantification of *Escherichia coli* passage into human small bowel tissue, following Ussing chamber incubations exemplified objective quantification of rare events in the context of lumen-tissue crosstalk. Second, in rat jejunum, precise histological context revealed distinct populations of intraepithelial lymphocytes between and directly below enterocytes enabling quantification in context of total epithelial cell numbers. Finally, mouse mononuclear phagocyte–T cell interactions, cell expression and significant spatial cell congregations were mapped to shed light on cell–cell communication in lymphoid Peyer’s patch. Accessible, quantitative tissue microscopy provides a new window-of-insight to diverse questions in gastroenterology. It can also help combat some of the data reproducibility crisis associated with antibody technologies and over-reliance on qualitative microscopy. © 2020 The Authors. *Cytometry Part A* published by Wiley Periodicals LLC on behalf of International Society for Advancement of Cytometry.

• Key terms

intestinal tissue; cell segmentation; machine learning; immunofluorescence; confocal microscopy; processing tiles in CellProfiler | Getis-Ord spatial statistics

TISSUE microscopy provides powerful insights into biological processes across differing scales from subcellular to the macroscopic. For example, it enables distinct structural and substructural tissue regions to be defined as well as cell–cell spatial relationships to be observed (1–6). Indeed, in situ tissue-based research in gastroenterology is generally about scale, with compartment-specific analyses often desirable due to the specific physiology that occurs region-by-region. For example, crosstalk between intestinal tissue and the luminal environment, cell differentiation along the crypt-villus axis and immune cell maps of gut lymphoid tissues are all active research areas which are, or could be, facilitated by quantitative, in situ measures (7–11).

In terms of in situ microscopy-based bioclinical research, immunofluorescence labeling and confocal imaging is the current mainstay, as it permits sensitive,

quantifiable detection of multiple targets with subcellular localization (12). Automated imaging has become standard, while new advances in artificial intelligence promise increased throughput through restoration of noisy images obtained at higher scan speeds (13). Yet, despite substantial advances in hardware and software, the majority of reported tissue microscopy “data” remains qualitative and exemplified by the representative image. Typically, for tissue-based research, flow cytometry delivers the quantitative data and confocal microscopy is the visual means by which the spatial relationships and mechanics of biological processes are then conceptualized. There is, however, a clear advantage in combining these outputs to deliver quantification of cell types, their contents and their location, simultaneously. Indeed, the power of data mining from regular chromogen-based histology exemplifies such an approach even though the image data are lower resolution and less amenable to multilabel, per-cell quantification (4,5,12).

In fact, quantitative methodologies for the analysis of confocal microscope-derived tissue images have existed for at least 15 years (1) and yet there remains a huge disconnect between what is possible and what has translated through to the biomedical community for everyday usage. Reasons for this have not been formally established, but interdisciplinary capability is a chief suspect (14). Currently, joined-up approaches to deal with everything from optimal biological experimentation, through sample preparation and imaging, to the programming skills generally required for successful image analysis seldom reside under one roof within the biomedical community. There are also a number of philosophies as to what constitutes quantitative immunofluorescence microscopy, ranging from basic summation of fluorescence data across a given area, through integration within approximated cell-objects, to accurate per-cell identification and quantification (termed “cell segmentation”) (1,2,15). The latter has marked advantage as, within the limits of a microscope’s resolution, it permits per-cell quantification of information in a manner amenable to familiar, flow cytometry-type gated analyses (1–3,16,17). It also allows distances to be established accurately, meaning that not only can cells be counted, but their content and spatial relationship to other cells or histological features can also be quantified (6).

Despite these advantages, accurate cell segmentation in tissues is complex, surprisingly sample-specific and time consuming for the nonexpert (1,18). Much work to date has grown out of approaches established for cultured cells (19) as sample homogeneity facilitates image analysis. Generally, however, tissues are not at all homogenous. In the intestine, for example, a multicell epithelial layer with diffuse lymphoid tissue beneath (the lamina propria) may be juxtaposed to a dense B-cell dominant follicle with a different overlying epithelial layer (e.g., the Peyer’s patch). For these reasons, accurate, quantitative, cell-based image analysis, compatible with such varying structure and delivered in a manner that is accessible to bioclinical scientists has not yet been developed in gastroenterology.

Here, we demonstrate pragmatic methodology to enable per-cell immunofluorescence quantification from confocal microscopy-derived images of diverse gastrointestinal tissues, and we exemplify the approach with analyses of general interest to the field. We show how image-based cell profiling can take gastrointestinal tissue microscopy beyond representative images with quantification of (a) common or rare cellular events alongside; (b) their cell content; and (c) location, coupled with visualization and statistical definition of cell–cell interactions and tissue microenvironments. Importantly, we use open-source, user-friendly software platforms to carry out the work, and to construct quantitative pipelines, which similarly we provide here in open-access formats.

MATERIALS AND METHODS

Animal Tissue Collection

Mouse (9–12 week-old) and rat (13 week old) tissues were collected from surplus healthy animals sacrificed for husbandry purposes by CO₂ asphyxiation and cervical dislocation. Ileal draining mesenteric lymph nodes were removed alongside jejunal/ileal intestinal samples (the latter containing Peyer’s patches) in ~2 cm lengths. Upon excision, tissue samples were immediately plunge frozen into isopentane precooled on melting dry ice, transferred to labeled cryovials, and stored in liquid nitrogen until use. Tissue samples for paraffin embedding were fixed in neutral buffered formalin (≥4 h), before transfer to tissue cassette, and automatic processing by standard hospital protocol (dehydration by ethanol series, three changes of 100% xylene (at 30°C), then three changes of paraffin wax (at 62°C).

Human Tissue Collection and Ethics

Following informed consent and with approval from the Regional Ethical Review Board, Linköping, Sweden, specimens from the neo-terminal ileum next to the ileo-caecal valve were collected during surgery from one inflammatory bowel disease (IBD) patient with Crohn’s disease (49 years, female) and one patient with colonic cancer (68 years, female), as a noninflammatory bowel disease (non-IBD) control. The Crohn’s disease patient had no anti-inflammatory medication and indication of surgery was ileitis. The tissue was macroscopically noninflamed. The tissue from the colon cancer patient was free from cancer; the patient had no generalized disease and had not received preoperative chemo- or radiotherapy. Studies using human tissue were also approved by the UK NHS Health Research Authority, North West—Greater Manchester East Research Ethics Committee, REC reference 18/NW/0690.

Ussing Chamber Experiments

Human ex vivo tissue ileal samples were transported directly from the operating theater to the laboratory in Krebs buffer. Three tissue segments per individual were mounted in modified Ussing chambers (Harvard Apparatus) as previously described (20). Transepithelial resistance and potential difference was used to assess tissue viability. Crohn’s disease

associated adherent invasive *Escherichia (E.) coli* strain LF82 were transformed with a plasmid (pEGFP, BD Biosciences) for expression of enhance green fluorescence protein (EGFP) as described previously (21). Live LF82 were then added to the mucosal side of the tissues at a final concentration of 1×10^8 CFU/ml. After 20 min, tissues were fixed in chambers with 4% PBS-buffered paraformaldehyde for 12 h at 4°C. The tissue samples were then immersed in 30% sucrose until embedded in optimal cutting temperature compound (OCT) for cryostat sectioning according to the protocol outlined below.

Tissue Labeling and General Immunofluorescence Protocol

For cryostat sectioning, frozen tissue samples were transported on ice and transferred into the cryostat chamber (–20°C) to acclimatize for 30 min. Samples were trimmed with a safety razor and transferred to molds containing pre-chilled OCT (VWR, 00411243). Sections were cut at 12 µm thickness, picked up on superfrost plus coated slides (ThermoFisher, J1800AMNT) and rested at room temperature for at least 2 h prior to immunofluorescence labeling. Formalin fixed paraffin embedded (FFPE) sections were cut at 5 µm thickness, then fully dewaxed and rehydrated by baking at 60°C for 1 h, changing twice through xylene, a reverse ethanol series (100%, 70%, 50%, 10%), followed by 1 min in water. All sections were then ringed with hydrophobic barrier pen (Vector Laboratories, H-4000) and unfixed cryostat sections were additionally fixed in 4% 0.1 M phosphate buffered (pH 7.4) paraformaldehyde for 10 min. All sections were transferred to block buffer (10% goat serum (ThermoFisher, 16210064), 2% bovine serum albumin (Biosera, PM-T1726) diluted in 25 mM Tris-buffered (pH 7.4) saline (TBS) containing 25 mM glycine) for at least 1 h. The block buffer was removed, and 100 µl of the necessary primary antibodies in block buffer were added to each section (concentrations and manufacturer's codes specified, Table S1). Sections were incubated for 1 h at room temperature under gentle agitation on a rocking platform. Each section was then washed thoroughly with three, 100 µl changes of TBS. Nuclei were counterstained using a 1:2,500 dilution of Hoechst 33342 (ThermoFisher, H3570) in TBS. Sections were washed once with 100 µl TBS, prior to addition of the secondary antibodies (concentrations, manufacturer's codes and conjugated fluorophores shown in Table S1). In with the secondary antibodies, phalloidin-AlexaFluor 647 (ThermoFisher, A22287) was included at ~660 nM to label cell membranes in frozen sections, or, 20 µg/ml wheat-germ agglutinin (WGA)-AlexaFluor 647 (ThermoFisher, W32466) was used to label membranes in the FFPE sections. Secondary antibody and cell outlines stains were incubated with the tissue sections for 1 h on a rocking platform. Each section was then washed with three changes of TBS prior drying carefully around each section with absorbent paper and mounting with #1.5 coverslips in Prolong Diamond mountant (ThermoFisher, P36965).

General Microscopy

2D images (typically 2,048 × 2,048 pixels per tile) were collected using sequential scanning on a Leica SP8 confocal microscope equipped with 405, 488, 562, and 633 nm lasers using plan-apochromat 63X/1.4 or 40X/1.3 oil immersion objectives. Tiles were collected with 10% edge overlap using focus mapping to maximize throughput.

Immunofluorescence Controls

For each study, image data were obtained in a single run under identical settings, with supporting secondary-only, isotype and leave-one-out antibody controls included in tissue-matched serial sections to assess background fluorescence, nonspecific binding and spectral crosstalk, respectively. For the Ussing's chamber work involving *E. coli* exposures to ex-vivo human tissues, a biological negative control (i.e., images for tissue exposed to Krebs buffer alone without *E. coli*) was also included.

Tilescan Processing Code

Tilescans were stitched together using the “Mosaic Merge” function in the Leica LASX software. The registered images were then cut up into ~4,000 × 4,000 pixel tiles with edge overlaps for processing with the open source and freely available CellProfiler (15) (www.CellProfiler.org) software using a custom function called “TilescanToCellProfiler.” This function structures the image data for input, and also stores user choices in a side-car information file for subsequent automated reassembly. After extracting per-cell data using CellProfiler, a second function called “CellProfilerToTilescan” was written to reassemble the data. This reassembles the segmented cell masks (22) while removing “double-hits” on overlap edges. It also extracts and spatially reassembles all of the cell feature data, while assigning a unique, master cell identity number and the correct global cell position coordinates for every cell. These functions are provided for MATLAB and Python alongside example data and full instructions (screen-cast videos) at the BioStudies database (<http://www.ebi.ac.uk/biostudies>) under accession number S-BSST305.

Single-Cell Segmentation and Immunofluorescence Quantification

Cell segmentation results were obtained using CellProfiler and Ilastik (23) (www.ilastik.org) softwares. Example image data and analysis pipelines (built using CellProfiler version 3.1.9 and Ilastik 1.3.3) and accompanied by screen-cast video walkthroughs are available for download at the BioStudies database (<http://www.ebi.ac.uk/biostudies>) under accession number S-BSST305. Details of the section-type, species and tissue-type, objective lens and numerical aperture, image pixel density and the cell segmentation strategy used in every analysis are summarized in Table S2. In brief, villus mucosal tissues were segmented using a marker-controlled watershed approach, wherein nuclei were defined as primary objects before the actin (cryostat sections) or WGA (FFPE sections) delineated cell outlines were classified into cell-objects using a

“IdentifySecondaryObjects” module. As lymphoid tissues segmented poorly using this watershed approach, these images (i.e., Peyer’s patches, mesenteric lymph nodes [MLNs]) were first classified into “cell outline,” “intracellular environment” or “other/background” probability maps using pixel classification machine learning in the Ilastik software (feature selection shown, Fig. S1, method exemplified, Fig. S2). The resultant probability maps of each cell were then segmented to yield cell-objects via an “IdentifyPrimaryObjects” module in CellProfiler. Immunofluorescence channels were preprocessed by two-class Otsu thresholding with a manual lower threshold set (independently for each analysis) at the level required to remove $\geq \sim 95\%$ of fluorescence in tissue-matched, secondary antibody-only control images. Fluorescence intensity values per cell, alongside per-cell size and shape features were then measured for all channels by integration in each cell-object using the “MeasureObjectSizeShape” and “MeasureObjectIntensity” modules in CellProfiler. In the same way, integration of thresholded images outputted as binaries was used to measure the fluorescence area per-cell. Cell features were written to both text files (i.e., accessible via Excel spreadsheet), and MATLAB objects for subsequent analysis.

Scoring Segmentation Accuracies

The pixel overlap agreement between manually and automatically segmented cell-objects was scored using the widely used intersection over union metric (Jaccard index) (24,25).

$$J(P,G) = \frac{|P \cap G|}{|P \cup G|} = \frac{|P \cap G|}{|P| + |G| - |P \cap G|} \quad (1)$$

where P and G are two sets containing pixel positions for the prediction (P) and ground truth (G), respectively. A score of 0 represents no overlap (i.e., false negative) whereas 1 is a perfect, per-pixel overlap. With this approach, it is acknowledged that a value of ~ 0.7 is a good segmentation result, and values of ~ 0.9 lie close to human annotation accuracy (26). This benchmarking was carried out without first removing mis-segmented cells.

Single Cell Data: Preprocessing

To remove mis-segmented cells, plots of each cell-object’s integrated nuclei and cell outline (i.e., actin or WGA) scores were plotted according to data density using “dscatter” (27). A cell population for analysis was then gated manually from these scatterplots using the inbuilt MATLAB function “inpolygon” to trace the contour surrounding the main cell population. This selection was then held the same when processing all image-sets associated with an experiment (i.e., experimental data and tissue matched controls).

Intraepithelial Lymphocytes: Image Analysis

Pixel classification machine learning in the Ilastik software was used to project masks for the epithelium, lamina propria and lumen “tissue compartments” directly from the actin channel. In MATLAB, the epithelial mask was refined by filling isolated interior pixels using the inbuilt function

“bwmorph,” prior to performing an erosion followed by a dilation using disk structuring elements (5 and 10 pixels, respectively) to bridge gaps. To find the different intra-epithelial lymphocyte (IEL) subclasses, the resulting epithelial mask was skeletonized using “bwskel,” with spurs less than 500 pixels removed. Expanding the skeleton using “imdilate” with a disk-structuring element of 32 pixels then created a central path mask through each “loop” of epithelium. The IEL subclassifications IEL^{sub} and IEL^{inter} were subsequently defined as $CD3^+$ cells with centroids either inside the epithelial region, or inside this central path mask, respectively. The width of the central path was defined manually, by visually checking that IEL^{inter} events were consistently caught within the mask, while IEL^{sub} events were excluded outside.

Statistical Analyses

Nonparametric differences between data from different groups were analyzed by Wilcoxon Rank Sum test. Statistically significant congregations of cells (i.e., indicative of cellular zonation) were identified relative to what would be expected by random chance given the frequencies of different cell types present using the Getis-Ord G_i^* statistical approach (28). This measures the spatial concentration of values x_j associated with j values within a distance d of the value x_i . The ratio G_i^* is defined as:

$$G_i^*(d) = \frac{\sum_{j=1}^n w_{ij}(d)x_j}{\sum_{j=1}^n x_j} \quad (2)$$

where $w_{ij}(d)$ defines the contribution to the numerator of the ratio depending on the distance d , for example using, that is, $w_{ij}(d) = 1$ if $d_{ij} < d$ else; $w_{ij}(d) = 0$ if $d_{ij} > d$. The Getis-Ord statistic is then given by:

$$Z[G_i^*(d)] = \frac{[G_i^*(d) - E(G_i^*(d))]}{\sqrt{\text{var } G_i^*(d)}} \quad (3)$$

where $E(G_i(d))$ represents the expected fraction of items within d , assuming a completely random distribution calculated as:

$$E(G_i^*(d)) = \frac{\sum_j w_{ij}(d)}{n-1} \quad (4)$$

The value $Z[G_i(d)]$ now describes the difference in the fraction of values within the distance d from location i from the random expected value relative to the standard deviation. In our example, we discretize the field of view into a grid and value x_i is defined as the number of cells of a certain phenotype in the grid position i .

RESULTS

With a specific focus on intestinal tissues, this work aims to develop and demonstrate open, user-friendly methodologies that enable per-cell immunofluorescence quantification in situ

using routine, confocal microscopy-derived images. Here, we focus on analysis of 2D images, as qualitative display in this format is the current standard in the bioclinical sciences.

Labeling Gastrointestinal Tissues for Cell Segmentation

First, we sought simple fluorescence labeling strategies compatible across species (i.e., antibody independent) for the purpose of delineating individual nuclei and cell outlines for subsequent cell segmentation. For both human and murine ileal sections, cut from either regular paraffin-embedded (i.e., FFPE) or snap-frozen and OCT embedded tissues, the fluorescent nuclear dye, Hoechst 33342, provided a straightforward, reliable means to label cell nuclei (Fig. 1A–L). Different strategies were required, however, to clearly delineate cell outlines in the two different section-types. Frozen sections exhibited artifacts when cell membranes were directly labeled using phospholipid labeling with wheat germ agglutinin (WGA) conjugates. This was especially notable at goblet cell sites, and is likely explained by nonspecific binding to mucins (Fig. S3). To avoid this, actin cytoskeletal staining via fluorescent phalloidin conjugates was used, and provided good demarcation of cell outlines (Fig. 1A–D). In contrast, for FFPE sections the situation was reversed. The cell actin filaments labeled by phalloidin conjugates were destroyed by alcohol exposure during the formalin fixation process and thus could not be labeled for cell outline determination (Fig. S3). However, in FFPE sections, direct cell membrane labeling with WGA was a successful strategy (Fig. 1E–H) probably because mucins were cleared when exposed to the solvents during processing.

Cell Segmentation Strategies Using Open Source Tools

With approaches for per-cell labeling established, we next considered cell segmentation strategies. Once again, dual strategies were necessary but, this time, dependent upon tissue region rather than tissue processing. For villus regions where cells are not tightly packed but cell types vary greatly in shape, and cell outlines are not always clear, a routine seeded watershed approach, readily deployed in CellProfiler appeared best. With this, the nucleus of each cell is first segmented and then used as an anchor point from which to define each cell's outline (Fig. 1A–D). In densely packed, pure lymphoid tissue (e.g., MLN or Peyer's patch), however, there were difficulties in accurately resolving individual nuclei and the resulting watershed approach performed poorly (Fig. S4). To resolve this, pixel classification machine learning in the Ilastik software was used to convert these images into probability maps of "cell outlines," "intracellular environments" or "background/other" (shown, Fig. S2). The intracellular probability map was then directly segmented into cell objects in CellProfiler using a IdentifyPrimaryObjects module (Fig. 1E–L). Of note, this latter approach (a) only required cell outline information (i.e., actin or WGA) for effective segmentation, freeing up the nuclear channel for other targets, and (b) was compatible with lower-resolution input images

(e.g., Fig. 1I–L), as results depend not upon contrast boundaries in the source image but upon derived probability maps. Thus, in conjunction with the antibody independent, tissue labeling strategies outlined above, these strategies permit cell segmentation across diverse intestinal tissues and are readily transferable between species and section-types (e.g., mouse, rat, human; villus mucosa, Peyer's patch, MLN; frozen and paraffin embedded are demonstrated, Fig. 1). For all analyses, histological information alongside imaging specifics and the cell segmentation strategy used are summarized in Table S2.

Accuracy of Cell Segmentation

The automated cell segmentations presented in Figure 1, which are derived across varying species and tissue preparations, were benchmarked—cell-by-cell—against hand-drawn manual segmentations using the commonly employed intersection over union approach (Jaccard index) (24–26) (>1,000 cells scored; Fig. S5). This benchmarking was carried out without first removing mis-segmented cells. Median scores in terms of pixel overlap were consistently between 0.80 and 0.83, with scores of ~0.9 recognized as the maximum realistically feasible with this approach due to the inherent accuracy limits of the manual segmentation itself (i.e., due to line thickness, outline smoothing, etc.), and 0.8–0.9 considered strong agreement (26) (exemplified, Fig. S5).

Open Source Image Analysis

The source images and the complete CellProfiler/Ilastik image analysis pipelines, which are necessary to enable the segmentation strategies shown in Figure 1, are provided at the BioStudies database (<http://www.ebi.ac.uk/biostudies>) under accession number S-BSST305. Both the CellProfiler and Ilastik softwares are freely available, and no programming is required for implementation of the image analysis routines described. Results, for example, per-cell shape and immunofluorescence quantifications can be outputted as text files easily openable as Excel sheets, or saved as MATLAB or HDF5 objects.

Immunofluorescence Quantification and Exclusion of Debris

Following cell segmentation, per-cell immunofluorescence quantification was implemented by CellProfiler pipeline using Otsu thresholding and the "MeasureObjectIntensity" and "MeasureObjectSizeShape" modules—as described in the Methods. Here, we subsequently chose to process the outputted tables of per-cell measurements using MATLAB. One aspect in tissues that required a different approach from *in vitro* cells was the determination of mis-segmented cell-objects that should be discarded prior to analysis (i.e., the debris equivalent of flow cytometry). For cultured cells, a recommended approach involves discarding objects that lie outside of the 5% or 95% percentiles by size (19). In tissue, however, this approach is less effective due to the diversity of cross sectional cell shapes and sizes including the occurrence of infrequent cell types of irregular size. Instead, simple density plots (e.g., insets, Fig. 1B,F,J) of each cell-object's

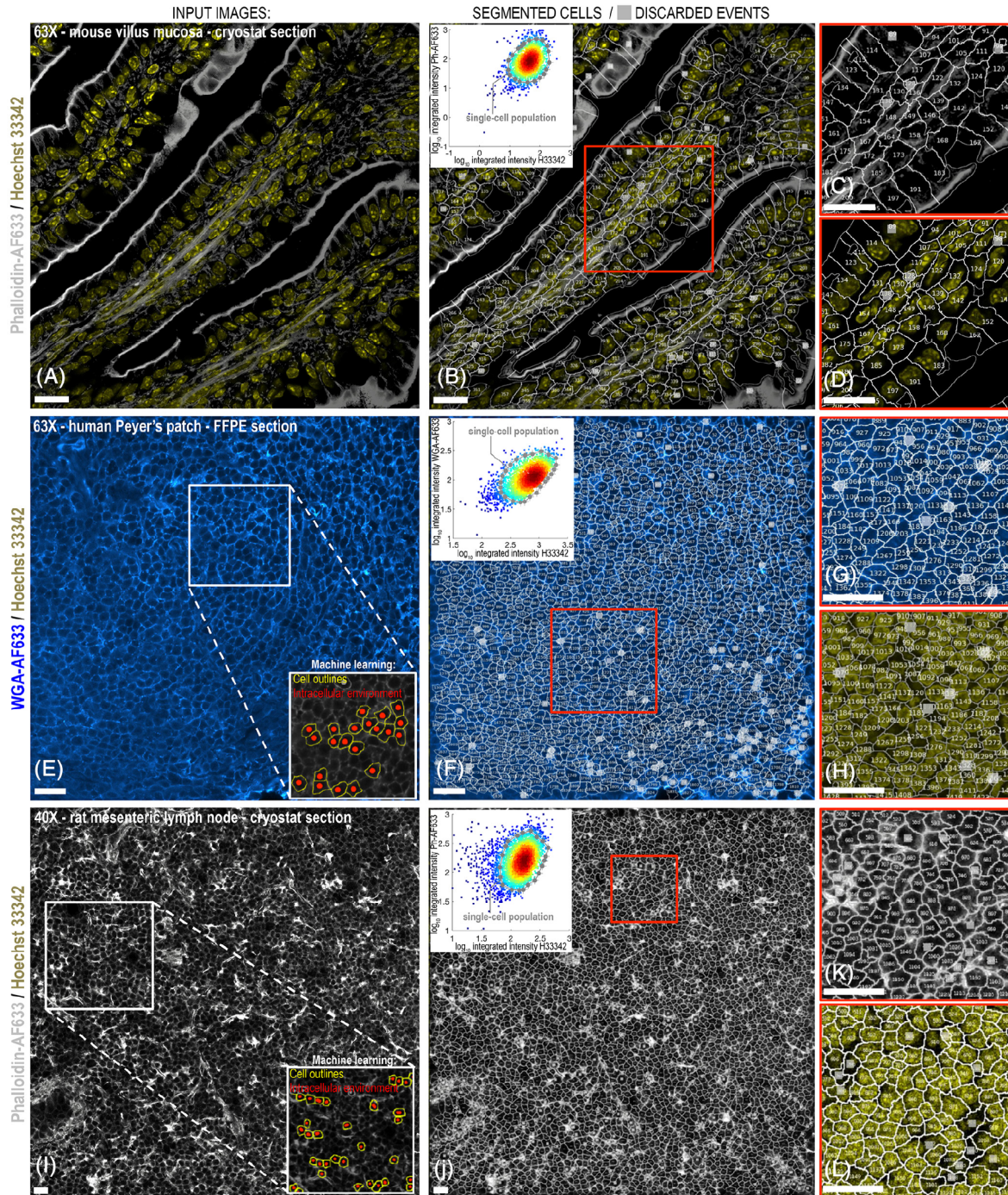


Fig 1. Open-source cell segmentation strategies for diverse intestinal tissues. (A–D) mouse villus mucosa frozen section. Here, we used a simple, marker-controlled watershed approach that first defines the nucleus (gold, Hoechst 33342) of each cell, and then uses this as an anchor point from which to find each cell’s actin-delineated boundary (gray, actin-AF633). (E–H) human Peyer’s patch lymphoid tissue; formalin fixed paraffin embedded (FFPE) section. Exposure to alcohol during the FFPE process destroys the actin microfilaments (see Fig. S3) so, instead, cell membranes were labeled using wheat germ agglutinin (WGA-AF633, blue). The marker-controlled watershed algorithm performed poorly in such densely packed tissue types (shown, Fig. S4), and so machine learning via the Ilastik software was instead used to produce probability maps of the cell outlines to enable segmentation (training shown in E, inset/process fully described, Fig. S2). (I–L) rat mesenteric lymph node frozen section. Despite lower magnification and image resolution, the same machine learning based, Ilastik-CellProfiler process enables accurate cell segmentation. B, F, and J—inserts, density plotting each cell’s nuclear and cell outline fluorescence provides a straightforward approach to “gate out” incorrectly segmented cell objects with abnormally high (e.g., doublets) or low (e.g., debris) signals. Example discarded events that lie outside of the indicated “single-cell population” are indicated with gray squares on the tissue images. For all examples, segmentation accuracy scores are provided in Figure S5. Scale bars = 20 μm . [Color figure can be viewed at wileyonlinelibrary.com]

integrated nuclear and cell outline fluorescence (i.e., WGA or actin) provided a route to gate out poorly segmented cells. Events that fell outside of the main population due to abnormally high (e.g., doublets) or low (e.g., true debris) signals were excluded (discarded events exemplified, Fig. 1—gray squares). A further advantage of this approach is that cells just partially clipped by the optical section tend to get removed, providing more consistent sampling of cells' cross-sectional immunofluorescence data.

Rare Events: *E. coli* Passage into Ileal Tissue

To demonstrate how image-based cell profiling can tackle rare event analysis of intestinal tissue, the passage of GFP-labeled *E. coli* strain LF82 into human ileum was considered (Fig. 2). Three tissue samples taken from one non-IBD patient with colon cancer, and one IBD patient with macroscopically noninflamed Crohn's disease, were investigated. A fourth tissue sample from the Crohn's patient was exposed to Krebs buffer alone (i.e., without *E. coli*) as a biological negative immunofluorescence control (Fig. 2A–C).

Images were collected from villus tissue regions across ~6–8 tissue sections taken at random intervals throughout each biopsy. This approach enabled rapid sampling from across the full dimensions of each tissue sample. As expected, no punctate spots of anti-GFP fluorescence were observed in the tissue biopsies exposed to Krebs buffer only (Fig. 2A). In each of the three cancer control non-IBD tissue biopsies, the few *E. coli* that were observed were bound to the apical side of the epithelium (indicated, Fig. 2B). Contrastingly, in all three tissue biopsies from the patient with Crohn's disease, transmucosal *E. coli* were identified within both the epithelial layer and lamina propria (Fig. 2C).

The aim of this work, however, was to move beyond careful qualitative observation—as described above—to objective quantification. To this end, the watershed approach developed for mucosal tissue rapidly allowed per-cell assessment of ~5,000 cells per tissue sample. The background fluorescence distribution was then established on the tissue sample exposed to Krebs buffer alone by plotting a cell-number normalized histogram of the signal in the anti-GFP channel (total cells analyzed = 5,475). When this step was repeated for the non-IBD tissue samples that had been exposed to *E. coli*, virtually no signal—above the established background—was observed (Fig. 2D, 14,671 cells analyzed). This demonstrated that the *E. coli* were not readily able to achieve transmucosal passage within the exposure timeframe in the non-IBD tissues. In contrast, when this was repeated in the Crohn's disease tissue samples, a positive increase in the per-cell fluorescence distribution was observed (Fig. 2D, 15,226 cells analyzed). Comparison of this increase relative to the non-IBD group showed significance at the $P < 0.001$ level (Wilcoxon rank sum, Fig. 2E).

Oftentimes it is convenient to call a cell as simply “positive” or “negative”—in this case meaning cells with anti-GFP fluorescence indicative of ≥ 1 *E. coli* event or none. As with flow cytometry, gating is required to determine this cut off and, again as for flow cytometry, there is a degree of

subjectivity relating to the stringency of specificity versus sensitivity. Here, when a gate was applied above the defined background fluorescence (indicated in Fig. 2D), then the number of anti-GFP positive cells in the Crohn's disease tissue was just 282 or 1.85%. The data therefore demonstrate how the image-based cell profiling approach can quantify rare events objectively, substantiating the representative images shown.

Processing Large Unbroken Image-Fields: Working with TileScans

Working with sets of individual images, obtained randomly across multiple tissue sections, as above, is one approach in image-based cell profiling. However, under other circumstances it may be desirable to work with high resolution, unbroken fields (i.e., tileScans) in which per-cell immunofluorescence analyses can be augmented by histological context (tissue mapping). CellProfiler does not currently possess dedicated modules for processing tileScans, and it is often not possible to directly process input images much larger than $\sim 4,000 \times 4,000$ pixels due to memory limitations on the local machine. For this reason, here we developed two software functions specifically aimed at processing immunofluorescence tileScans. The first, which we call “TileScanToCellProfiler,” takes stitched tileScans directly in most proprietary microscopy formats and cuts them into a series of user-defined, manageably-sized tiles for CellProfiler input. After processing, a second function called “CellProfilerToTileScan” seamlessly reassembles the cell segmentation and spatial positions of the extracted, per-cell data. These functions can be deployed with a single line of code in the programming environments MATLAB or Python. Example images, code and full instructions (screen-cast videos) for the nonexpert are provided at the BioStudies database (<http://www.ebi.ac.uk/biostudies>) under accession number S-BSST305.

Machine-Learning Tissue Compartments: The Intestinal Epithelium

The highly convoluted shape of the gastrointestinal mucosa makes accurate, region-of-interest selections for different tissue “compartments” (e.g., epithelium, lamina propria, etc.) complicated and time consuming to perform. At the same time, compartment-specific analyses are often desirable due to the specific physiology that occurs region-by-region. To demonstrate the automation of compartment-specific gastrointestinal analysis, we set out to profile intraepithelial T lymphocytes in longitudinal frozen sections of rat jejunum—just using a single CD marker and the histological context afforded by *in situ* microscopy. To accurately identify the epithelium, one of a pair of serial frozen sections was immunolabeled for epithelial cell adhesion molecule (EPCAM)—alongside nuclei and actin. This precisely pinpointed the location of the epithelial region between the basement membrane and the apical enterocyte surface (29–31) (Fig. 3A). Using this EPCAM labeling as a guide to inform pixel annotation, we then trained an Ilastik machine learning model to mask the

INPUT IMAGES:

Nuclei / Actin / anti-GFP

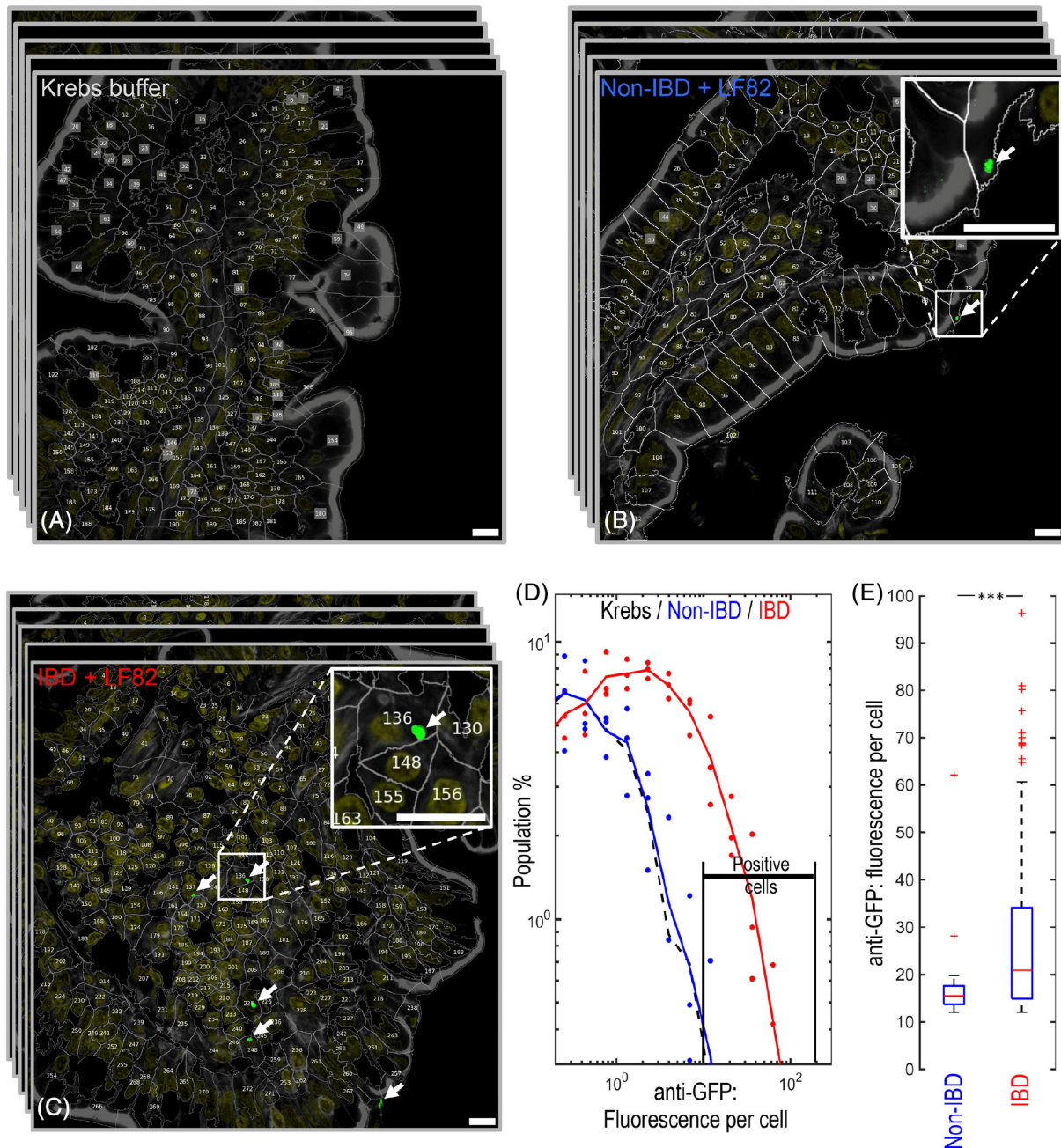


Fig 2. Conquering the representative image: *E. coli* passage into human control or IBD ileal mucosa. (A–C) Mucosal tissue samples from the distal ileum of either a non-IBD cancer control patient (B) or an IBD patient (Crohn's disease) (A and C) were maintained ex vivo in Ussing chambers. The apical sides of the living tissues were exposed to either Krebs buffer (A) or GFP- expressing, adherent invasive *E. coli* strain LF82 (B and C) for 20 min. Images of the mucosa were then collected randomly across 6–8 frozen sections per tissue sample prior to per-cell analysis for anti-GFP fluorescence to identify LF82. (D and E) Instead of relying on representative images, the image-based cell profiling approach allowed quantification and display of all of the collected data (>35,000 cells analyzed in total). (D) Concordance between the tissue samples taken per patient was observed (filled circles). While the fluorescence distributions (lines represent averages) from the non-IBD cancer control samples directly overlaid the Krebs buffer negative control (dashed line) suggesting no transmucosal uptake, the samples from the IBD patient all showed elevated fluorescence. (E) Comparison between non-IBD and IBD groups for gated cells (indicated, D) with anti-GFP values greater than present in Krebs buffer control. (***) indicates statistical significance at $P < 0.001$ (Wilcoxon rank sum). Scale bars: A–C = 10 μ m. [Color figure can be viewed at wileyonlinelibrary.com]

epithelium, as well as the lumen and lamina propria tissue compartments, directly from the actin channel itself. In this way, the EPCAM labeling was no longer required (Fig. 3B)

(process exemplified stepwise, Fig. S6). Of note, we also found that the same approach worked with WGA labeling in FFPE sections (demonstrated, Fig. S6E).

Utilizing Locational and Cellular Information: Profiling Intraepithelial Lymphocytes

Next, we set out to utilize both tissue compartment and per-cell image-data to profile intraepithelial T lymphocytes in the jejunal mucosa. In the second serial section, a 112-tile tilescan containing a wide region of villous mucosa was collected with anti-CD3 labeling to identify T cells. As both the EPCAM and CD3 antibodies were raised in the same host, instead of dual-labeling, the tissue compartment model was deployed in the mucosa to provide a mask for the epithelium (Fig. 3C).

To understand and quantify background fluorescence, as well as the nonspecific binding capacity of the CD3 antibody in the rat jejunal tissue, set of 10 image-fields for either the secondary antibody alone (i.e., 2° only control), or the secondary plus an irrelevant primary antibody of the same isotype (i.e., an isotype control) were collected for the CD3 channel in adjacent, serial sections. Per-cell immunofluorescence data was then extracted from the CD3 tilescan (~60,000 cells) and control image-sets (~6,000 cells) using the watershed cell segmentation pipeline optimized for mucosal tissue (above). A CD3⁺ cell population was then formed by gating cells with per-cell fluorescence values greater than those observed in the 2°-only and isotype controls (i.e., as is typical in flow cytometry) (Fig. 3D). Cell centroid markers were displayed on each gated cell, to help pinpoint CD3⁺ T lymphocytes both visually and for subsequent locational categorization. Interfacing this gated cell population with the epithelial mask allowed further division of the CD3⁺ cell population into intraepithelial lymphocyte (IEL) and lamina propria T cell subpopulations by identification of cells with centroids inside or outside of the mask (Fig. 3D). Upon close study of the defined IEL CD3⁺ cells in context of the masked epithelium, it was clear that this cell population existed in two distinct forms. IEL events were either observed in close association with the basal aspect of enterocytes (hereafter termed “IEL^{sub}”), or, were truly between individual enterocytes (hereafter termed “IEL^{inter}”). To split the IELs into these two classes, the epithelial mask was subjected to a morphological process called skeletonization. This reduced the epithelial mask to yield a central path through each “loop” of villus epithelium (process exemplified, Fig. S6). Inclusion within this submask allowed the central, IEL^{inter} population to be separated out, leaving behind the IEL^{sub} cells (Fig. 4A–E).

In this way, harnessing per-cell fluorescence data in combination with the precise histological context provided by the high-resolution tilescan allowed the identified CD3⁺ cells to be subdivided into three distinct subpopulations (i.e., lamina propria CD3⁺ (LP^{CD3+}), IEL^{sub} and IEL^{inter}). This, alongside the segmentation of all cells, whether immunolabeled or not, provided data well suited to automated cell counting in the context of a tissue map. Hence, we measured the areas occupied by the different designated compartments—alongside their cell counts—in total, per 100 cells, and as ratios between the different tissue compartments (Fig. 4F–I). Interestingly, while not so apparent visually, the epithelium occupied a greater area (Fig. 4F) and contained more total cells (Fig. 4G,I) than the underlying

lamina propria. CD3⁺ cells were also determined more abundant per-cell in the lamina propria than in the epithelium (Fig. 4I). Meanwhile, whereas IELs were quite common, the IEL^{inter} subclass were rare events (~4 per 100 epithelial cells). This was especially true when compared to the IEL^{sub} class, at ~13 per 100 epithelial cells (Fig. 4H,I).

Cell Interactions and Expression: Mapping in the Peyer’s Patch

Access to per-cell immunofluorescence data collected in situ provides the opportunity to consider both cell expression and physical cellular interactions via nearest-cell neighbor analyses. Lymphoid tissues represent one such environment in which interaction and expression data are of key importance. Here we considered CD11c⁺ mononuclear phagocyte—CD3⁺ T cell expression and interactions in a transverse section of mouse Peyer’s patch (24-tile tilescan, ~16,000 cells) (Fig. 5). A basic overview of the structure and cellular zonation of the murine Peyer’s patch is provided in Figure S7.

Image-data were collected for six channels: fluorescence data were collected for nuclei, actin, CD11c (for mononuclear phagocytes; i.e., antigen presenting cells) and CD3 as a pan T-lymphocyte marker. Alongside, transmitted and reflected light were also collected to inform on overall histology and section quality (Fig. 5A). As before, data for the respective 2°-only and isotype controls were also collected alongside in tissue-matched serial sections. As per-cell immunofluorescence quantification was to be carried out on two of the channels (i.e., CD3 and CD11c), leave-one-out control image-sets were also taken to check for any fluorescence crosstalk between channels. This involved labeling additional serial sections with either CD11c or CD3, yet collecting the respective fluorescence data for both channels. In this way, any crosstalk into the “empty” channel could be detected in the resultant per-cell fluorescence distributions.

Using the Ilastik/CellProfiler machine learning cell segmentation pipeline, alongside the software reassembly (tilescan) functions described above, the lymphoid tissue was segmented seamlessly across the entire Peyer’s patch (Fig. 5B). A region-of-interest (ROI) was then set around the lymphoid tissue, and just the CD11c and CD3 immunofluorescence data were shown on top of the segmented-cell outlines inside the ROI. Outside of the ROI, just the actin staining was displayed, to provide histological context (Fig. 5C,D). This visualization approach was found to dramatically reduce the visual complexity of the six-channel image, permitting display of most important information in a per-cell and visually intuitive manner—across the scale of the entire lymphoid follicle.

To build CD3⁺ and CD11c⁺ cell populations, after debris removal (discussed above), gating was first used to select cells with fluorescence values above those observed in the 2°-only and “leave-one-out” controls. The fluorescence distributions of the isotype controls were also used to inform gating. Here, while we gated above values high enough to remove >~99% of cells from the isotype distributions, gating at the maximum was avoided for fear of building highly specific, yet poorly

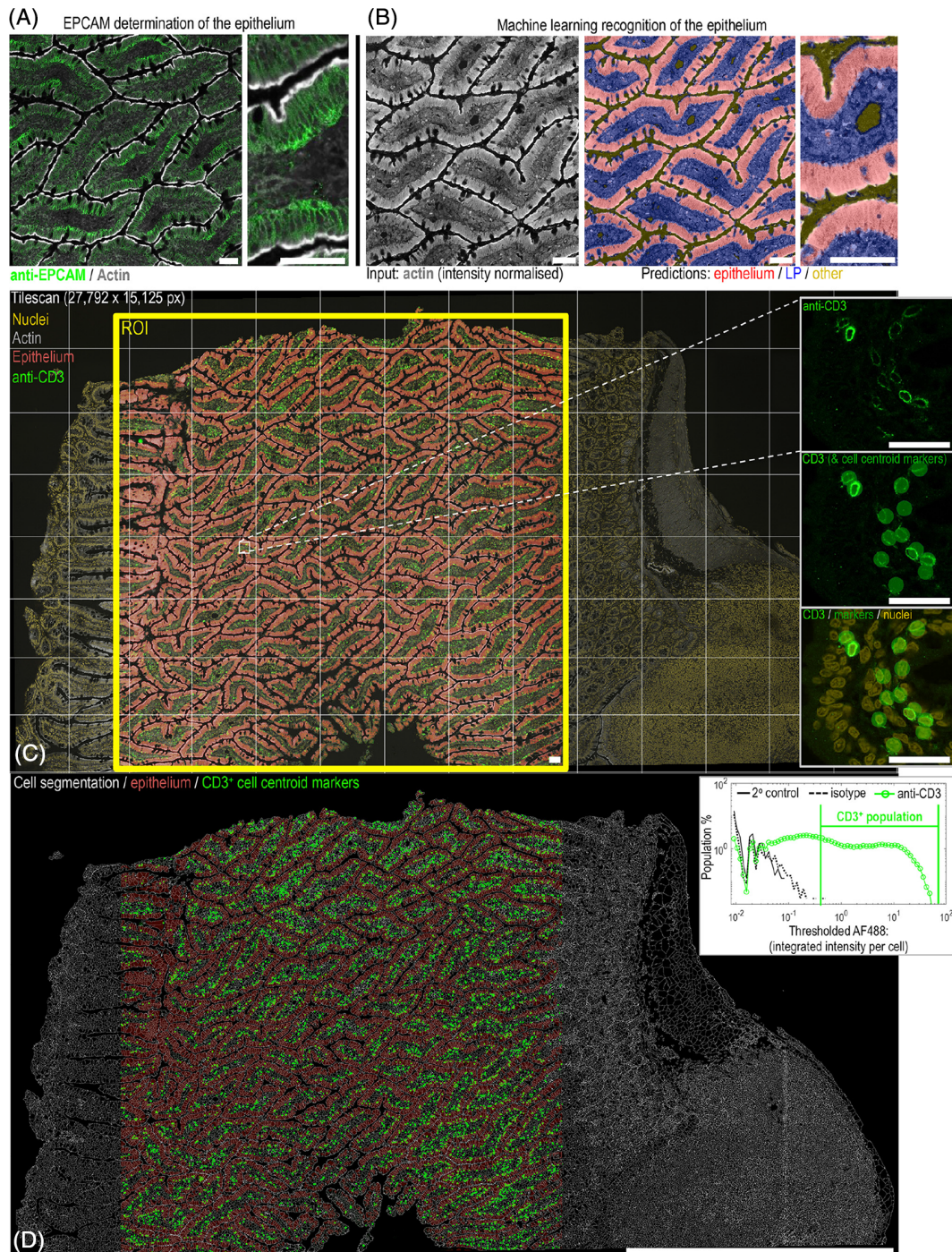


Fig 3. Identifying intraepithelial T lymphocytes in large tilescons using a single CD marker. Rat jejunal longitudinal tissue section. **(A)** First, anti-epithelial cell adhesion molecule (EPCAM) immunofluorescence labeling was used to delineate the epithelium (i.e., cells lying between the basement membrane and the apical enterocyte surface). **(B)** As the anti-EPCAM antibody was raised in the same host species as the desired lymphocyte marker, the epithelium, lamina propria and lumen “compartments” were directly-detected from the actin channel using pixel-classification machine learning in Ilastik (process outlined in Fig. S6). **(C)** A 112-tile confocal tilescan labeled for nuclei, actin and anti-CD3 was collected. Each individual field was segmented into individual cells, and a software function was developed to spatially reassemble the images, segmentation masks and cell positions (> 60,000 cells). **(D)** A region-of-interest (ROI) was placed around the tissue region containing optimally cross-sectioned villi, and the Ilastik model was used to predict and mask the epithelium (pink). D—inset, CD3⁺ cells inside or outside of this epithelial mask were then identified by gating against the secondary-only and isotype control per-cell fluorescence distributions (i.e., as is typical in flow cytometry). Cell centroid markers were placed on each positive event. This approach permitted sensitive and accurate pinpointing of CD3⁺ lymphocytes (C, inset). Scale bars: A and B = 25 μ m; C = 50 μ m; D = 1 mm. [Color figure can be viewed at wileyonlinelibrary.com]

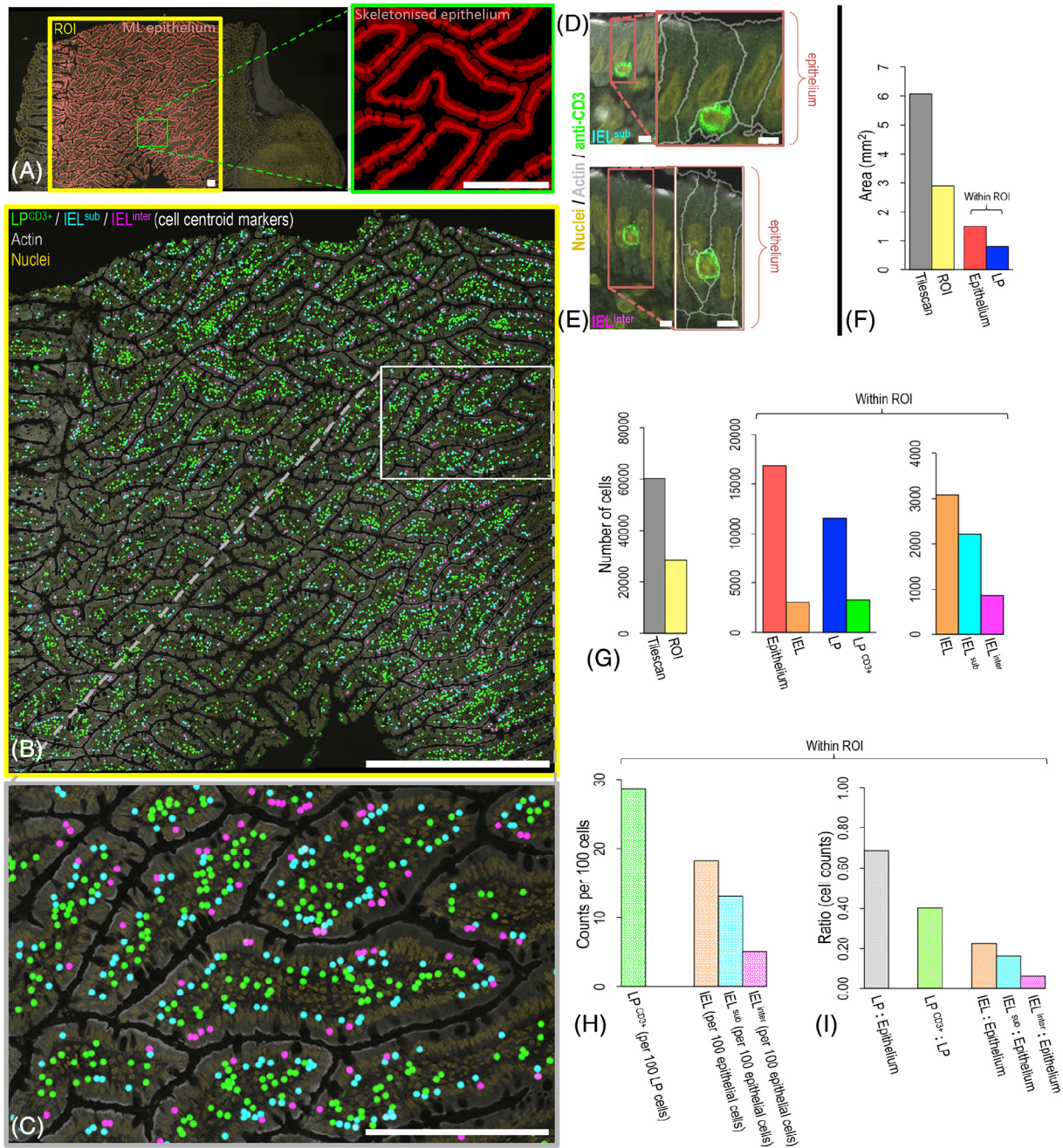


Fig 4. Defining and counting T lymphocyte subpopulations on the basis of spatial location. Rat jejunum, continued from Figure 3. **(A)** To differentiate intraepithelial (IEL) CD3⁺ cells in close association with the apical side of the epithelium (i.e., “IEL^{sub}”) versus those truly in between enterocytes (i.e., “IEL^{int}”), the epithelial mask defined in Figure 3 was skeletonized and dilated to form a central path enabling detection of IEL^{int} cells in each “loop” of epithelium (method shown, Fig. S6). **(B and C)** The CD3⁺ cell population was then split into three subpopulations: Lamina propria CD3⁺ (LP^{CD3+}), IEL^{sub} or IEL^{int}, and locations were displayed by marker placement on each cell’s centroid. **(D and E)** Typical examples of IEL^{sub} and IEL^{int} events. **(F)** Area measurements for the different tissue compartments. **(G)** Total cell counts according to tissue compartment. **(H)** Tissue compartment cell counts per 100 cells. **(I)** Cell counts expressed as ratios between the different tissue compartments. Scale bars: A and C = 100 μm; B = 500 μm; D and E = 5 μm. [Color figure can be viewed at wileyonlinelibrary.com]

sensitive cell populations (Fig. 5E). Due to the closely packed cells, and in conjunction with the expression of CD11c and CD3 immunofluorescence on the cell membrane, it was found that adding a second sequential gate on the area of fluorescence within each cell helped to reduce “bystander-positive”

events caused by small amounts of fluorescence spanning the segmented-cell outlines and manifesting in immediately adjacent neighboring cells (further discussion/exemplification provided, Fig. S8). In this way, cells exhibiting CD marker fluorescence all around their perimeters were better isolated

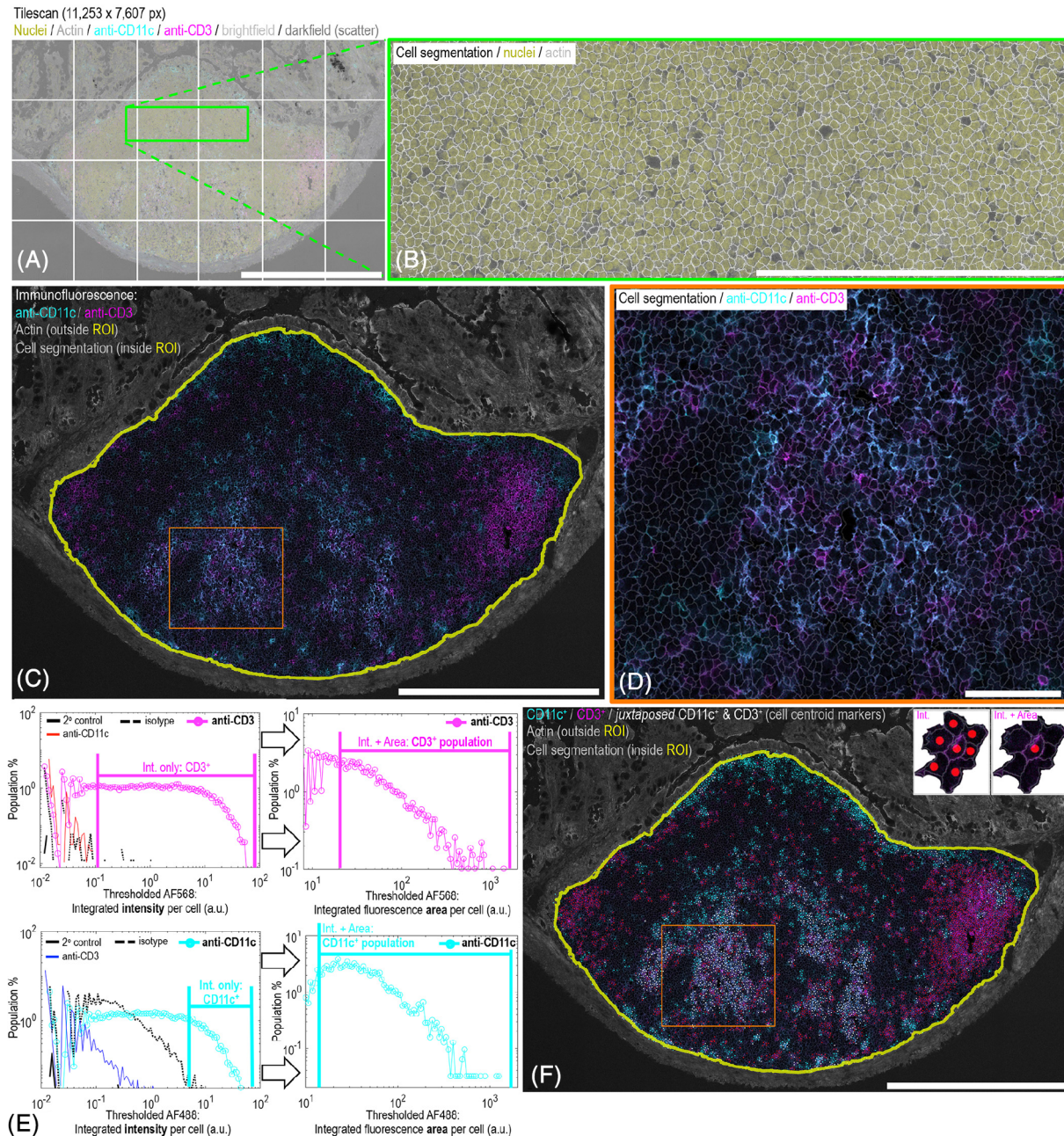


Fig 5. CD11c and CD3 expression in Peyer's patches. Mouse ileal transverse section. (A) The section was imaged as a 24-tile tilescan labeled for nuclei, actin, anti-CD11c, and anti-CD3. Transmitted and reflected light were also collected. (B) The densely packed, lymphoid tissue segmented accurately into cells (~16,000) using the Ilastik/CellProfiler machine learning approach, and the software reassembly function spatially reassembled the data without artifacts at the overlap boundaries between the tiled images. (C and D) The complexity of multichannel image (A) is dramatically reduced, using an example region-of-interest (yellow), and displaying the immunofluorescence markers (cyan = CD11c, magenta = CD3) on the segmented-cell outlines. (E and F) CD3⁺ and CD11c⁺ cell populations were gated against secondary-only, leave-one-out, and isotype control per-cell fluorescence distributions. In the densely packed tissue, a second sequential gate on fluorescence area per cell-object helped to reduce "bystander-positive" events caused by fluorescence overlap into neighboring cells (shown, F—Inset) (further discussion/exemplification provided, Fig. S8). (F) Marker placement on the cell populations identified by the gating strategy in (E). Highly juxtaposed, and thus indicative of communication, CD11c-CD3 neighboring cells (i.e., region shown in D) that sequentially identified positive in both gates are shown with white markers. Scale bars: A, C, and F = 500 μ m; B = 250 μ m; D = 50 μ m. [Color figure can be viewed at wileyonlinelibrary.com]

from their immediate neighbors, while maintaining sensitivity (Fig. 5F, inset). To aid this second gating step, cell-centroid

markers for the identified cell populations were placed onto the immunofluorescence images, as described above, providing visual feedback (Fig. 5F). As expected, the subepithelial dome (SED) was rich in mononuclear phagocytes and the

interfollicular region (IFR) at the right of the image contained large numbers of T-cells. Surprisingly, however, a population of highly juxtaposed, CD11c-CD3 neighboring cells (i.e., region shown in Fig. 5D) that still identified positive in both gates after bystander removal were identified, indicating an interaction (17) relative to other cells, and suggesting a likelihood of cell–cell communication (Fig. 5F).

In addition to placing markers on cell centroids to delineate the gated cell populations (Fig. 6A), other methods capable of clearly visualizing the single-cell data and consequent spatial relationships across the scale of the complete Peyer's Patch were sought. In Figure 6B, the marker-placement view was simplified further by flood-filling the individual segmented cell masks to clearly show the populations in a manner that could be effectively visualized at small size. The absence of immunofluorescence labeling (i.e., black, CD11c⁻/CD3⁻ regions) was also informative, as within the patch, the vast majority of these double-negative cells will be B lymphocytes (32). Next, the flood-filled view was simplified further to only show CD11c cells with touching CD3 nearest-cell neighbors (including juxtaposed CD11c/CD3 cells) (Fig. 6C). In this way, the view gives a sense of the spatial distribution of APCs within interactive distances of T lymphocytes. Interestingly, it was observed that the majority of these events were predominantly congregated around B cell follicles in the germinal center (GC) region, and were much less apparent in the SED where, probably, MNP–B cell interactions may predominate (33).

Having successfully identified populations of cells, next we moved forwards to consider quantification of per-cell fluorescence (i.e., related to protein expression). To do this, we made use of the ~800 nm optical Z plane afforded by the confocal optics and high numerical aperture objective (63X/1.4) to isolate a thin plane through individual cells. The analysis was also aided by the ability to select for cell objects optimally cross-sectioned through their central plane during the debris removal step (discussed above), as this improved measurement consistency by sampling data from similar, central regions in each cell. To clearly visualize the data from across the whole lymphoid follicle within a reasonable figure-size, the per-cell expression of CD11c and CD3 was displayed in four intensity bands (i.e., dim, low, intermediate, and high) (Fig. 6D,E). Perhaps unsurprisingly given the highly mixed population of mononuclear phagocytes delineated by CD11c, no clear spatial patterning according to CD11c expression was observed (Fig. 6D). For CD3, however, the IFR at the right of the patch, in addition to the APC and T-cell zones around the GC were rich in CD3^{int/hi} events, while the marginal zone and SED where predominated by CD3^{dim/lo}. This may be related to T cell subtypes, or activation, and deserves further scrutiny (34).

Finally, we also sought a method to statistically identify significant spatial congregations of cells so that regions of cellular zonation/established cellular microenvironments could be defined across the lymphoid follicle. To do this we harnessed both the cell location and CD11c or CD3 per-cell expression data and used these to calculate the Getis-Ord GI*

spatial statistic (28). This provided a heat map identifying where statistically significant, spatial congregations of different cell types occurred relative to what should be expected by random chance—given the frequencies of the different cell types involved (Fig. 6F). As expected, the SED was significantly rich for CD11c, as was the IFR for CD3. For both cell types, however, the maps also revealed a wealth of complex microstructure surrounding B-cell follicles in the GC. Under the “steady state” normal biology depicted here, it was also noted that the SED was sparse in terms of congregating CD3⁺ T-lymphocytes.

DISCUSSION

Here, with a specific focus on intestinal tissues, we develop open, user-friendly methodology that enables per-cell quantification using routine confocal micrographs. As a methodological advancement, it is important that the findings here are seen as a range of examples around capability, rather than individually-powered biological studies. Notwithstanding, we flag areas where the technique revealed interesting findings, including measures of spatially distinct IEL subpopulations—being either between or beneath enterocytes in the villus mucosa—and the complex microstructure of cellular zonation in the Peyer's patch, including spatial distributions of APC-T cell interactions.

Our image-based cell profiling approach delivers data in three key ways: (a) it enumerates different cell types, as in flow cytometry, but it also (b) provides precise cellular locational data with histological context and (c) resolution and quantification of cell contents. To achieve this, a number of novel approaches had to be developed or bridged together. First, we provide routes in MATLAB or Python to enable tilescan processing with CellProfiler—to include spatial reassembly of the mined per-cell data and the production of global segmentation masks with unique cell identities to enable visualizations. Second, to permit accurate cell segmentation, antibody-independent cell labeling was employed, which we optimized for both FFPE and frozen intestinal tissue sections. Importantly, this approach does not use up antibody hosts and transfers easily and directly between species. Next, effective per-cell immunofluorescence analysis requires accurately segmented cells with mis-segmented cells (debris) excluded. We show how this can be achieved using density plots to refine a consistently sampled cell population, with the outliers (partial cells or doublets) excluded. Alongside, to tackle the difficult issue of dense cell packing in lymphoid tissues, we use semantic machine learning within a fast, user-friendly framework (23) to yield accurate cell segmentations. Finally, we demonstrate how inadvertent bystander-positive cells can be obviated through sequentially gating on fluorescence intensity followed by fluorescence area, and suggest that any remaining bystanders are indicative of cell–cell interactions. We also demonstrate how spatial statistics can be employed to better define tissue microenvironments in terms of identifying significant cellular congregations.

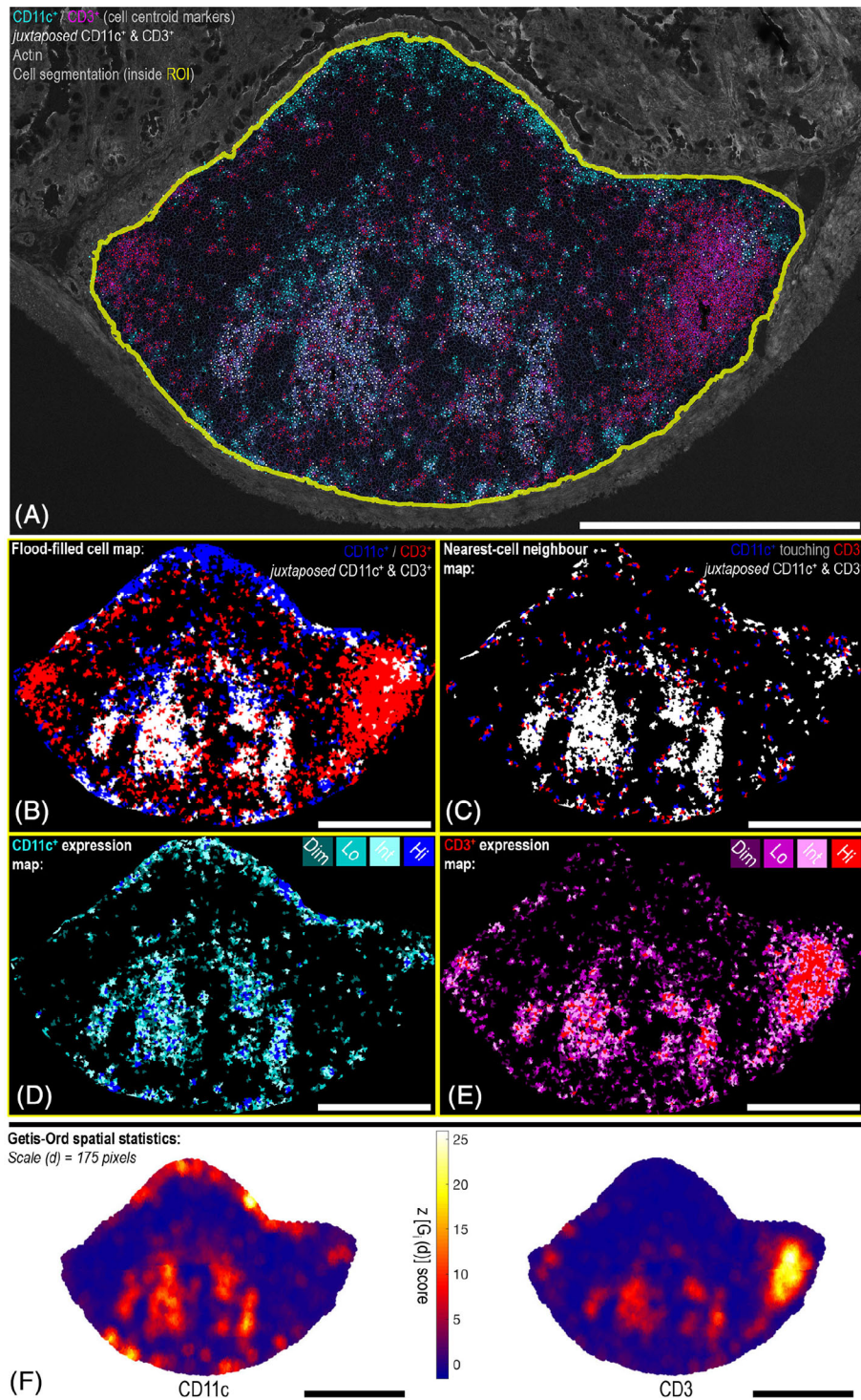


Fig 6. Mapping CD11c⁺ and CD3⁺ cell interactions, expression, and spatial congregations. Mouse ileum, continued from Figure 5. (A) Cell-marker placement on the basis of immunofluorescence labeling. Highly juxtapsed CD11c-CD3 neighboring cells that consequently identified positive for both markers are shown with white markers. (B) Flood-filling segmented cell objects provides a visually intuitive version of the data shown in (A) that can be displayed at a much smaller size. (C) Here, the view shown in (B) is simplified to only show CD11c cells with touching CD3 nearest-cell neighbors, or juxtapsed cells that identify positive for both markers due to close spatial association. (D and E) CD11c and CD3 expression maps with cell-objects colored in four levels (i.e., dim, lo, intermediate (int), hi) according to each segmented cell's level of immunofluorescence. (F) Getis-Ord statistical map: This shows—as a probability heat map—where statistically significant congregations of cells are found relative to what would be expected by random chance given the frequencies of the different cell types. Scale bars: A–F = 500 μ m. [Color figure can be viewed at wileyonlinelibrary.com]

Here we use the popular CellProfiler software as a “backbone” to enable per-cell quantification. Importantly, its pipeline-based style is extremely flexible, and can use original, deconvolved and/or spectrally unmixed input images from almost any microscope or upstream software package. Moreover, the pipelines provided here can also utilize probability maps to enable cell segmentation from any source, including, where necessary, more advanced machine learning approaches such as deep convolutional neural networks (26). Of note, a delivery of the popular Unet architecture within the user-friendly environment of Ilastik is planned for release in spring 2020 (35).

In embracing such a technical approach to tissue analysis, it is critical that the fundamentals of robust immunofluorescence methodology are not overlooked. In our experience, best possible tissue orientation helps greatly in interpretation of outputs. Moreover, our approach does not obviate good practice in labeling and imaging: rather, success relies upon it. Controls, to assess background autofluorescence and non-specific antibody binding, are extremely important, alongside assurance that fluorescence signals do not cross between channels. Here, we have demonstrated the use of 2°-only, isotype, leave-one-out and biological negative controls collected in tissue-matched, serial sections to assess these parameters. We then use Otsu thresholding in conjunction with gated analyses to accurately isolate cell populations and measure cell expression/contents. Of course, these controls in themselves do not ensure that the correct target is being labeled and, as always in such work, proper validation of antibodies remains essential (36,37).

Our accessible approach to per-cell analysis of tissue sections contrasts with other techniques. While imaging mass cytometry (e.g., “CyTOF”), enables the use of dozens of antibody markers, it has lower spatial resolution and necessitates highly specialist instrumentation for detection (6,18). Although high throughput and extremely powerful within diagnostic pathology, packages that permit the analysis of chromogen stained slide-scans lack the resolution, sensitivity, ability to multiplex and quantitate immunolabeled targets in a way that is often required for precision research (12). Meanwhile, commercial “all-in-one” solutions, such as those employing fluorescence slide scanners or spinning disk confocal techniques, are (a) expensive, especially when highly capable confocal microscopes are already available at most research institutions and (b) rely on software with the unenviable task of enabling the analysis of all conceivable tissue types. In our experience, this results in approximate per-cell measurements. In contrast, focusing in one field and interfacing different strategies (18), as we do here with the intestine, enables precision cell segmentation to be achieved and thus accurate analyses of cellular localization, per-cell content and cell–cell interactions.

Finally, some of the original, pioneering, work in quantitative, flow cytometry-type immunofluorescence analysis of tissues (e.g., histocytometry, Refs. 1–3,16,17) relies upon commercial software for implementation, limiting accessibility. Moreover, while the histocytometry approach utilizes both

2D and 3D confocal images, analyses have primarily focused on the spatial relationships of just the CD-marker delineated cells. Our approach, supported by machine learning segmentation where necessary, enables precision analyses of all cells and hence highly accurate cell counting and per-cell quantifications within an entire section or region-of-interest. Notwithstanding, here we focus on open-source, intestinal-specific 2D delivery, as qualitative display in this format is today’s gold standard, and because volumetric (i.e., 3D) immunofluorescence quantification is extremely challenging for routine usage, given the time requirements and increasing nonuniformities that manifest with imaging depth. With this in mind, the use of the confocal optical section provides 2D immunofluorescence data that is consistently sampled, and thus well-suited for summation within cell-objects and for fair comparison across experimental samples. In turn, an important question for future work may involve addressing how far regular fluorescence images (i.e., nonconfocal) can be taken toward producing similar, quantitative results. To this end, herein we show how per-cell data can be extracted from FFPE sections through the use of WGA staining to delineate cell outlines. This approach may prove important to success with regular fluorescence microscopy because FFPE sections can be cut much thinner than cryostat sections, and this physical section thickness itself may enable reliable extraction of per-cell information.

To conclude, here we have developed open, user-friendly methodology that delivers per-cell quantifications using routine, confocal microscopy-derived images of diverse gastrointestinal tissues. In combination, the presented approaches take the field of gastroenterology far beyond the representative image, and should now help to combat some of the data reproducibility issues that are associated with antibody technologies and over-reliance on qualitative tissue microscopy (36,37).

ACKNOWLEDGMENTS

The authors would like to acknowledge the UK Medical Research Council (grant number MR/R005699/1), the UK Engineering and Physical Sciences Research Council (grant EP/H008683/1), and the UK Biotechnology and Biological Sciences Research Council (grant number BB/P026818/1) for supporting the work. J.W.W. is extremely grateful to Girton College and the University of Cambridge Herchel-Smith Fund for supporting him with fellowships.

CONFLICT OF INTERESTS STATEMENT

The authors declare no conflicts of interest.

LITERATURE CITED

1. Ecker RC, Steiner GE. Microscopy-based multicolor tissue cytometry at the single-cell level. *Cytometry Part A* 2004;59A:182–190.
2. Harnett MM. Laser scanning cytometry: Understanding the immune system in situ. *Nat Rev Immunol* 2007;7:897–904.
3. Gerner MY, Kastenmuller W, Ifrim I, Kabat J, Germain RN. Histo-cytometry: A method for highly multiplex quantitative tissue imaging analysis applied to dendritic cell subset microanatomy in lymph nodes. *Proc Natl Acad Sci U S A* 2012;37:364–376.

4. Bankhead P, Loughrey MB, Fernandez JA, Dombrowski Y, McArt DG, Dunne PD. QuPath: Open source software for digital pathology image analysis. *Sci Rep* 2017;7:16878.
5. McQuaid S, Gray RT, Murray LJ, Coleman HG, James JA, Salto-Tellez M, Hamilton PW, Blom S. Systems pathology by multiplexed immunohistochemistry and whole-slide digital image analysis. *Sci Rep* 2017;7:15580.
6. Schapiro D, Jackson HW, Raghuraman S. histoCAT: Analysis of cell phenotypes and interactions in multiplex image cytometry data. *Nat Methods* 2017;14:873–876.
7. Morita N, Umemoto E, Fujita S, Hayashi A, Kikuta J, Kimura I, Haneda T, Imai T, Inoue A, Mimuro H, et al. GPR31-dependent dendrite protrusion of intestinal CX3CR1(+) cells by bacterial metabolites. *Nature* 2019;566:110–114.
8. Moor AE, Harnik Y, Ben-Moshe S, Massasa EE, Rozenberg M, Eilam R, Bahar Halpern K, Itzkovitz S. Spatial reconstruction of single enterocytes uncovers broad zonation along the intestinal villus axis. *Cell* 2018;175:1156–1167.e15.
9. Eisenhoffer GT, Loftus PD, Yoshigi M, Otsuna H, Chien CB, Morcos PA, Rosenblatt J. Crowding induces live cell extrusion to maintain homeostatic cell numbers in epithelia. *Nature* 2012;484:546–549.
10. Da Silva C, Wagner C, Bonnardel J, Gorvel JP, Lelouard H. The Peyer's patch mononuclear phagocyte system at steady state and during infection. *Front Immunol* 2017;8:1254.
11. Cummings RJ, Barbet G, Bongers G, Hartmann BM, Gettler K, Muniz L, Furtado GC, Cho J, Lira SA, Blander JM. Different tissue phagocytes sample apoptotic cells to direct distinct homeostasis programs. *Nature* 2016;539:565–569.
12. Robertson D, Savage K, Reis-Filho JS, Isacke CM. Multiple immunofluorescence labelling of formalin-fixed paraffin-embedded (FFPE) tissue. *BMC Cell Biol* 2008;9:13.
13. Weigert M, Schmidt U, Boothe T, Müller A, Dibrov A, Jain A, Wilhelm B, Schmidt D, Broaddus C, Culley S, et al. Content-aware image restoration: Pushing the limits of fluorescence microscopy. *Nat Methods* 2018;15:1090–1097.
14. Carpenter AE, Kametsky L, Eliceiri KW. A call for bioimaging software usability. *Nat Methods* 2012;9:666–670.
15. Carpenter AE, Jones TR, Lamprecht MR, Clarke C, Kang IH, Friman O, Guertin DA, Chang JH, Lindquist RA, Moffat J, et al. CellProfiler: Image analysis software for identifying and quantifying cell phenotypes. *Genome Biol* 2006;7:R100.
16. Li W, Germain RN, Gerner MY. Multiplex, quantitative cellular analysis in large tissue volumes with clearing-enhanced 3D microscopy (Ce3D). *Proc Natl Acad Sci U S A* 2017;114:e7321–e7330.
17. Kotov DI, Pengo T, Mitchell JS, Gastinger MJ, Jenkins MK. Chrysalis: A new method for high-throughput histo-cytometry analysis of images and movies. *J Immunol* 2019;202:300–308.
18. Schuffler PJ, Schapiro D, Giesen C, Wang HA, Bodenmiller B, Buhmann JM. Automatic single cell segmentation on highly multiplexed tissue images. *Cytometry Part A* 2015;87A:936–942.
19. Caicedo JC, Cooper S, Heigwer F, Warchal S, Qiu P, Molnar C, Vasilevich AS, Barry JD, Bansal HS, Kraus O, et al. Data-analysis strategies for image-based cell profiling. *Nat Methods* 2017;14:849.
20. Keita AV, Gullberg E, Ericson AC, Salim SY, Wallon C, Kald A, Artursson P, Soderholm JD. Characterization of antigen and bacterial transport in the follicle-associated epithelium of human ileum. *Lab Invest* 2006;86:504–516.
21. Moussata D, Goetz M, Gloeckner A, Kerner M, Campbell B, Hoffman A, Biesterfeld S, Flourie B, Saurin JC, Galle PR, et al. Confocal laser endomicroscopy is a new imaging modality for recognition of intramucosal bacteria in inflammatory bowel disease in vivo. *Gut* 2011;60:26–33.
22. kwikteam/npymatlab: Code to read and write NumPy's NPY format in MATLAB. GitHub repository. Available at <https://github.com/kwikteam/npymatlab>. Accessed March 25th, 2020.
23. Sommer C, Straehle C, Köthe U, Hamprecht FA. Ilastik: Interactive learning and segmentation toolkit. 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2011; p. 230–233. Chicago, IL: IEEE.
24. Maska M, Ulman V, Svoboda D, Matula P, Matula P, Ederra C, Urbíola A, Espana T, Venkatesan S, Balak DM, et al. A benchmark for comparison of cell tracking algorithms. *Bioinformatics* 2014;30:1609–1617.
25. Everingham M, Van Gool L, Williams CKI, Winn J, Zisserman A. The Pascal visual object classes (VOC) challenge. *Int J Comput* 2010;88:303–338.
26. Falk T, Mai D, Bensch R, Çiçek Ö, Abdulkadir A, Marrakchi Y, Böhm A, Deubner J, Jäckel Z, Seiwald K, et al. U-Net: Deep learning for cell counting, detection, and morphometry. *Nat Methods* 2019;16:67–70.
27. Eilers PHC, Goeman JJ. Enhancing scatterplots with smoothed densities. *Bioinformatics* 2004;20:623–628.
28. Ord JK, Getis A. Local spatial autocorrelation statistics: Distributional issues and an application. *Geograph Anal* 1995;27:286–306.
29. Dobbins WO 3rd. Human intestinal intraepithelial lymphocytes. *Gut* 1986;27:972–985.
30. Pellegrino S, Villanacci V, Sansotta N, Scarfi R, Bassotti G, Vieni G, Princiotta A, Sferlazzas C, Magazzu G, Tuccari G. Redefining the intraepithelial lymphocytes threshold to diagnose gluten sensitivity in patients with architecturally normal duodenal histology. *Aliment Pharmacol Ther* 2011;33:697–706.
31. Rostami K, Marsh MN, Johnson MW, Mohaghegh H, Heal C, Holmes G, Ensari A, Aldulaimi D, Bancel B, Bassotti G, et al. ROC-king onwards: Intraepithelial lymphocyte counts, distribution & role in coeliac disease mucosal interpretation. *Gut* 2017;66:2080–2086.
32. De Jesus M, Ahlawat S, Mantis NJ. Isolating and immunostaining lymphocytes and dendritic cells from murine Peyer's patches. *J Vis Exp* 2013; e50167.17(73).
33. Reboldi A, Arnon TI, Rodda LB, Atakilit A, Sheppard D, Cyster JG. IgA production requires B cell interaction with subepithelial dendritic cells in Peyer's patches. *Science* 2016;352:aaf4822.
34. Valle A, Barbaggioanni G, Jofra T, Stabilini A, Perol L, Baeyens A, Anand S, Cagnard N, Gagliani N, Piaggio E, et al. Heterogeneous CD3 expression levels in differing T cell subsets correlate with the in vivo anti-CD3-mediated T cell modulation. *J Immunol* 2015;194:2117–2127.
35. Kreshuk A. Automating image analysis with machine learning. *Microscience Microscopy Congress*. Manchester, UK; 2019.
36. Uhlen M, Bandrowski A, Carr S, Edwards A, Ellenberg J, Lundberg E, Rimm DL, Rodriguez H, Hiltke T, Snyder M, et al. A proposal for validation of antibodies. *Nat Methods* 2016;13:823–827.
37. Baker M. Reproducibility crisis: Blame it on the antibodies. *Nature* 2015;521:274–276.