1   **TITLE PAGE**

2   **Image-based Cell Profiling Enables Quantitative Tissue Microscopy in Gastroenterology**

3   **Authors:**

4   *John W. Wills[1], Jack Robertson[1], Huw D. Summers[2], Michelle Miniter[1], Claire Barnes[2], Rachel E.

5   Hewitt[1], Åsa V. Keita[3], Johan D. Söderholm[3], §Paul Rees[2,4], §Jonathan J. Powell[1].

6   § Contributed equally.

7

8   **Affiliations:**

9   [1]Biominerals Research, Cambridge University Department of Veterinary Medicine, School of

10   Biological Sciences, Madingley Road, Cambridge CB3 0ES, UK.

11   [2]Centre for Nanohealth, Swansea University College of Engineering, Fabian Way, Crymlyn Burrows,

12   Swansea SA1 8EN, UK.

13   [3]Department of Surgery and Department of Clinical and Experimental Medicine, Linköping

14   University, 58193 Linköping, Sweden.

15   [4]Broad Institute of MIT and Harvard, 415 Main Street, Cambridge, MA 02142, USA.

16

17   **Running Headline:**

18   Image-Based Cell Profiling in Gastroenterology

19

20   ***Corresponding Author Information:**

21   John W. Wills

22   Email: jw2020@cam.ac.uk

23   Tel: +447843 824042

24   [1]Biominerals Research, Cambridge University Department of Veterinary Medicine, School of

25   Biological Sciences, Madingley Road, Cambridge CB3 0ES, UK.

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 labelling approaches, compatible across preparation types and species, were optimised. 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 visualisation and statistical definition of cell locations, interactions and established microenvironments. First, quantification of *E. coli* passage into human small bowel tissue, following Ussing chamber incubations exemplified objective quantification of rare events in the context of lumen-tissue crosstalk. Secondly, in rat jejunum, precise histological context revealed distinct populations of intra-epithelial 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.

KEYWORDS

Intestinal tissue, cell segmentation, machine learning, immunofluorescence, confocal microscopy, Processing tilescans in CellProfiler.

INTRODUCTION

Tissue microscopy provides powerful insights into biological processes across differing scales from sub-cellular to the macroscopic. For example, it enables distinct structural and sub-structural 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, cross talk 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 labelling and confocal imaging is the current mainstay, as it permits sensitive, quantifiable detection of multiple targets with subcellular localisation (12). Automated imaging has become standard, whilst 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 conceptualised. 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 multi-label, per-cell quantification (4,5,12).

In fact, quantitative methodologies for the analysis of confocal microscope-derived tissue images have existed for at least fifteen 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

81 suspect (14). Currently, joined-up approaches to deal with everything from optimal biological

82 experimentation, through sample preparation and imaging, to the programming skills generally

83 required for successful image analysis seldom reside under one roof within the biomedical

84 community. There are also a number of philosophies as to what constitutes quantitative

85 immunofluorescence microscopy, ranging from basic summation of fluorescence data across a

86 given area, through integration within approximated cell-objects, to accurate per-cell identification

87 and quantification (termed 'cell segmentation') (1,2,15). The latter has marked advantage as, within

88 the limits of a microscope's resolution, it permits per-cell quantification of information in a manner

89 amenable to familiar, flow cytometry-type gated analyses (1-3,16,17). It also allows distances to be

90 established accurately, meaning that not only can cells be counted, but their content and spatial

91 relationship to other cells or histological features can also be quantified (6).

92

93 Despite these advantages, accurate cell segmentation in tissues is complex, surprisingly sample-

94 specific and time consuming for the non-expert (1,18). Much work to date has grown out of

95 approaches established for cultured cells (19) as sample homogeneity facilitates image analysis.

96 Generally, however, tissues are not at all homogenous. In the intestine, for example, a multi-cell

97 epithelial layer with diffuse lymphoid tissue beneath (the lamina propria) may be juxtaposed to a

98 dense B-cell dominant follicle with a different overlying epithelial layer (*e.g.,* the Peyer's patch). For

99 these reasons, accurate, quantitative, cell-based image analysis, compatible with such varying

100 structure and delivered in a manner that is accessible to bio-clinical scientists has not yet been

101 developed in gastroenterology.

102

103 Here, we demonstrate pragmatic methodology to enable per-cell immunofluorescence quantification

104 from confocal microscopy-derived images of diverse gastrointestinal tissues, and we exemplify the

105 approach with analyses of general interest to the field. We show how image-based cell profiling can

106 take gastrointestinal tissue microscopy beyond representative images with quantification of (i)

107 common or rare cellular events alongside (ii) their cell content and (iii) location, coupled with

108 visualisation and statistical definition of cell-cell interactions and tissue microenvironments.

109 Importantly, we use open-source, user-friendly software platforms to carry out the work, and to

110 construct quantitative pipelines, which similarly we provide here in open-access formats.

111

112

113 **MATERIALS & METHODS**

114 **ANIMAL TISSUE COLLECTION**

115 Mouse (9-12 week-old) and Rat (13 week old) tissues were collected from surplus healthy animals

116 sacrificed for husbandry purposes by CO_2 asphyxiation and cervical dislocation. Ileal draining

117 mesenteric lymph nodes were removed alongside jejunal / ileal intestinal samples (the latter

118 containing Peyer's patches) in ~ 2 cm lengths. Upon excision, tissue samples were immediately

119 plunge frozen into isopentane pre-cooled on melting dry ice, transferred to labeled cryovials, and

120 stored in liquid nitrogen until use. Tissue samples for paraffin embedding were fixed in neutral

121 buffered formalin (≥ 4 h), before transfer to tissue cassette, and automatic processing by standard

122 hospital protocol (dehydration by ethanol series, three changes of 100% xylene (at 30 °C), then

123 three changes of paraffin wax (at 62 °C).

124

125 **HUMAN TISSUE COLLECTION & ETHICS**

126 Following informed consent and with approval from the Regional Ethical Review Board, Linköping,

127 Sweden, specimens from the neo-terminal ileum next to the ileo-caecal valve were collected during

128 surgery from one inflammatory bowel disease (IBD) patient with Crohn's disease (49 years, female)

129 and one patient with colonic cancer (68 years, female), as a non-inflammatory bowel disease (non-

130 IBD) control. The Crohn's disease patient had no anti-inflammatory medication and indication of

131 surgery was ileitis. The tissue was macroscopically non-inflamed. The tissue from the colon cancer

132 patient was free from cancer; the patient had no generalised disease and had not received

133 preoperative chemo- or radiotherapy. Studies using human tissue were also approved by the UK

134 NHS Health Research Authority, North West – Greater Manchester East Research Ethics

135 Committee, REC reference 18/NW/0690.

136

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 LABELLING & GENERAL IMMUNOFLUORESCENCE PROTOCOL

For cryostat sectioning, frozen tissue samples were transported on ice and transferred into the cryostat chamber (- 20 °C) to acclimatise for 30 min. Samples were trimmed with a safety razor and transferred to moulds containing pre-chilled OCT (VWR, 00411243). Sections were cut at 12 micron thickness, picked up on superfrost plus coated slides (ThermoFisher, J1800AMNT) and rested at room temperature for at least 2 h prior to immunofluorescence labelling. Formalin fixed paraffin embedded (FFPE) sections were cut at 5 micron thickness, then fully dewaxed and rehydrated by baking at 60 °C for 1h, 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% 0.1M 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 hour. 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, **Supplementary Table 1**).

165 Sections were incubated for 1 hour at room temperature under gentle agitation on a rocking

166 platform. Each section was then washed thoroughly with three, 100 µL changes of TBS. Nuclei were

167 counterstained using a 1:2500 dilution of Hoechst 33342 (ThermoFisher, H3570) in TBS. Sections

168 were washed once with 100 µL TBS, prior to addition of the secondary antibodies (concentrations,

169 manufacturer's codes and conjugated fluorophores shown in **Supplementary Table 1).** In with the

170 secondary antibodies, phalloidin-AlexaFluor 647 (ThermoFisher, A22287) was included at ~ 660 nM

171 to label cell membranes in frozen sections, or, 20 µg/mL wheat-germ agglutinin (WGA)-AlexaFluor

172 647 (ThermoFisher, W32466) was used to label membranes in the FFPE sections. Secondary

173 antibody and cell outlines stains were incubated with the tissue sections for 1 hour on a rocking

174 platform. Each section was then washed with three changes of TBS prior drying carefully around

175 each section with absorbent paper and mounting with #1.5 coverslips in Prolong Diamond mountant

176 (ThermoFisher, P36965).

177

178 **GENERAL MICROSCOPY**

179 2-D images (typically 2048x2048 pixels per tile) were collected using sequential scanning on a

180 Leica SP8 confocal microscope equipped with 405 nm, 488 nm, 562 nm and 633 nm lasers using

181 plan-apochromat 63X/1.4 or 40X/1.3 oil immersion objectives. Tilescans were collected with 10%

182 edge overlap using focus mapping to maximize throughput.

183

184 **IMMUNOFLUORESCENCE CONTROLS**

185 For each study, image data were obtained in a single run under identical settings, with supporting

186 secondary-only, isotype and leave-one-out antibody controls included in tissue-matched serial

187 sections to assess background fluorescence, non-specific binding and spectral cross-talk,

188 respectively. For the Ussing's chamber work involving *E. coli* exposures to *ex-vivo* human tissues, a

189 biological negative control (*i.e.,* images for tissue exposed to Krebs buffer alone without *E. coli*) was

190 also included.

191

192 **TILESCAN PROCESSING CODE**

193   Tilescans were stitched together using the 'Mosaic Merge' function in the Leica LASX software. The

194   registered images were then cut up into ~ 4000x4000 pixel tiles with edge overlaps for processing

195   with the open source and freely available CellProfiler (15) (www.CellProfiler.org) software using a

196   custom function called 'TilescanToCellProfiler'. This function structures the image data for input,

197   and also stores user choices in a side-car information file for subsequent automated reassembly.

198   After extracting per-cell data using CellProfiler, a second function called 'CellProfilerToTilescan' was

199   written to reassemble the data. This reassembles the segmented cell masks (22) whilst removing

200   'double-hits' on overlap edges. It also extracts and spatially reassembles all of the cell feature data,

201   whilst assigning a unique, master cell identity number and the correct global cell position

202   coordinates for every cell. These functions are provided for MATLAB and Python alongside example

203   data and full instructions at the BioStudies database (http://www.ebi.ac.uk/biostudies) under

204   accession number S-BSST305.

205

206   **SINGLE-CELL SEGMENTATION AND IMMUNOFLUORESCENCE QUANTIFICATION**

207   Cell segmentation results were obtained using CellProfiler and Ilastik (23) (www.ilastik.org)

208   softwares. Example image data and analysis pipelines (built using CellProfiler version 3.1.9 and

209   Ilastik 1.3.3) and accompanied by screen-cast video walkthroughs are available for download at the

210   BioStudies database (http://www.ebi.ac.uk/biostudies) under accession number S-BSST305. Details

211   of the section-type, species and tissue-type, objective lens and numerical aperture, image pixel

212   density and the cell segmentation strategy used in every analysis are summarised in **Table S2**. In

213   brief, villus mucosal tissues were segmented using a watershed approach, wherein nuclei were

214   defined as primary objects before the actin (cryostat sections) or WGA (FFPE sections) delineated

215   cell outlines were classified into cell-objects using a 'IdentifySecondaryObjects' module. As

216   lymphoid tissues segmented poorly using the watershed approach, these images (*i.e.,* Peyer's

217   patches, mesenteric lymph nodes (MLNs)) were first classified into 'cell outline', 'intracellular

218   environment' or 'other / background' probability maps using pixel classification machine learning in

219   the Ilastik software (feature selection shown, **Figure S1**, method exemplified, **Figure S2**). The

220   resultant probability maps of each cell were then segmented to yield cell-objects via an

221 'IdentifyPrimaryObjects' module in CellProfiler. Immunofluorescence channels were preprocessed

222 by two-class Otsu thresholding with a manual lower threshold set (independently for each analysis)

223 at the level required to remove $\geq \sim 95\%$ of fluorescence in tissue-matched, secondary antibody-only

224 control images. Fluorescence intensity values per cell, alongside per-cell size and shape features

225 were then measured for all channels by integration in each cell-object using the

226 'MeasureObjectSizeShape' and 'MeasureObjectIntensity' modules in CellProfiler. In the same way,

227 integration of thresholded images outputted as binaries was used to measure the fluorescence area

228 per-cell. Cell features were written to both text files (*i.e.,* accessible via Excel spreadsheet), and

229 MATLAB objects for subsequent analysis.

230

231 **SCORING SEGMENTATION ACCURACIES**

232 The pixel overlap agreement between manually and automatically segmented cell-objects was

233 scored using the widely used intersection over union metric (Jaccard index) (24,25).

234

235
$$J(P, G) = \frac{|P \cap G|}{|P \cup G|} = \frac{|P \cap G|}{|P| + |G| - |P \cap G|} \qquad (1)$$

236

237

238 Where *P* and *G* are two sets containing pixel positions for the prediction (*P*) and ground truth (*G*),

239 respectively. A score of 0 represents no overlap (*i.e.,* false negative) whereas 1 is a perfect, per-

240 pixel overlap. With this approach, it is acknowledged that a value of ~ 0.7 is a good segmentation

241 result, and values of ~ 0.9 lie close to human annotation accuracy (26). This benchmarking was

242 carried out without first removing mis-segmented cells.

243

244

245 **SINGLE CELL DATA: PREPROCESSING**

246 To remove mis-segmented cells, plots of each cell-object's integrated nuclei and cell outline (*i.e.,*

247 actin or WGA) scores were plotted according to data density using 'dscatter' (27). A cell population

248 for analysis was then gated manually from these scatterplots using the inbuilt MATLAB function

249   'inpolygon' to trace the contour surrounding the main cell population. This selection was then held

250   the same when processing all image-sets associated with an experiment (*i.e.,* experimental data

251   and tissue matched controls).

252

253   **INTRAEPITHELIAL LYMPHOCYTES: IMAGE ANALYSIS**

254   Pixel classification machine learning in the Ilastik software was used to project masks for the

255   epithelium, lamina propria and lumen 'tissue compartments' directly from the actin channel. In

256   MATLAB, the epithelial mask was refined by filling isolated interior pixels using the inbuilt function

257   'bwmorph', prior to performing an erosion followed by a dilation using disk structuring elements (5

258   and 10 pixels, respectively) to bridge gaps. To find the different intraepithelial lymphocyte (IEL)

259   subclasses, the resulting epithelial mask was skeletonised using 'bwskel', with spurs less than 500

260   pixels removed. Expanding the skeleton using 'imdilate' with a disk-structuring element of 32 pixels

261   then created a central path mask through each 'loop' of epithelium. The IEL sub-classifications

262   IEL^{sub} and IEL^{inter} were subsequently defined as $CD3^{+}$ cells with centroids either inside the epithelial

263   region, or inside this central path mask, respectively. The width of the central path was defined

264   manually, by visually checking that IEL^{inter} events were consistently caught within the mask, whilst

265   IEL^{sub} events were excluded outside.

266

267   **STATISTICAL ANALYSES**

268   Non-parametric differences between data from different groups were analysed by Wilcoxon Rank

269   Sum test. Statistically significant congregations of cells (*i.e.,* indicative of cellular zonation) were

270   identified relative to what would be expected by random chance given the frequencies of different

271   cell types present using the Getis-Ord GI* statistical approach (28). This measures the spatial

272   concentration of values x_j associated with *j* values within a distance *d* of the value x_i. The ratio *G* is

273   defined as:

274

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

276

277   where $w_{ij}(d)$ defines the contribution to the numerator of the ratio depending on the distance d, for

278   example using i.e., $w_{ij}(d)$ = 1 if d_{ij} < d else; $w_{ij}(d)$ = 0 if d_{ij} > d. The Getis-Ord statistic is then given

279   by:

280

281    $$Z[G_i(d)] = \frac{[G_i(d) - E(G_i(d))]}{\sqrt{var\ G_i(d)}} \qquad (3)$$

282

283   Where, $E(G_i(d))$ represents the expected fraction of items within d, assuming a completely random

284   distribution calculated as:

285

286    $$E\big(G_i(d)\big) = \frac{\sum_j \omega_{ij}(d)}{n-1} \qquad (4)$$

287

288   The value $Z[G_i(d)]$ now describes the difference in the fraction of values within the distance d from

289   location i from the random expected value relative to the standard deviation. In our example we

290   discretise the field of view into a grid and value x_i is defined as the number of cells of a certain

291   phenotype in the grid position i.

292

293

294   **RESULTS**

295   With a specific focus on intestinal tissues, this works aims to develop and demonstrate open, user-

296   friendly methodologies that enable per-cell immunofluorescence quantification *in situ* using routine,

297   confocal microscopy-derived images. Here, we focus on analysis of 2-D images, as qualitative

298   display in this format is the current standard in the bio-clinical sciences.

299

300   **Labelling Gastrointestinal Tissues for Cell Segmentation**

301   First, we sought simple fluorescence labelling strategies compatible across species (*i.e.,* antibody

302   independent) for the purpose of delineating individual nuclei and cell outlines for subsequent cell

303   segmentation. For both human and murine ileal sections, cut from either regular paraffin-embedded

304   (*i.e.,* FFPE) or snap-frozen and OCT embedded tissues, the fluorescent nuclear dye, Hoechst

305   33342, provided a straightforward, reliable means to label cell nuclei (**Figure 1A-L**). Different

306   strategies were required however to clearly delineate cell outlines in the two different section-types.

307   Frozen sections exhibited artefacts when cell membranes were directly labelled using phospholipid

308   labelling with wheat germ agglutinin (WGA) conjugates. This was especially notable at goblet cell

309   sites, and is likely explained by non-specific binding to mucins (**Figure S3**). To avoid this, actin

310   cytoskeletal staining via fluorescent phalloidin conjugates was used, and provided good

311   demarcation of cell outlines (**Figure 1A-D**). In contrast, for FFPE sections the situation was

312   reversed. The cell actin filaments labelled by phalloidin conjugates were destroyed by alcohol

313   exposure during the formalin fixation process and thus could not be labelled for cell outline

314   determination (**Figure S3**). However, in FFPE sections, direct cell membrane labelling with WGA

315   *was* a successful strategy (**Figure 1E-H**) probably because mucins were cleared when exposed to

316   the solvents during processing.

317

318   **Cell Segmentation Strategies Using Open Source Tools**

319   With approaches for per-cell labelling established, we next considered cell segmentation strategies.

320   Once again, dual strategies were necessary but, this time, dependent upon tissue region rather than

321   tissue processing. For villus regions where cells are not tightly packed but cell types vary greatly in

322   shape, and cell outlines are not always clear, a routine seeded watershed approach, readily

323   deployed in CellProfiler appeared best. With this, the nucleus of each cell is first segmented and

324   then used as an anchor point from which to define each cell's outline (**Figure 1A-D**). In densely

325   packed, pure lymphoid tissue (*e.g.,* MLN or Peyer's patch), however, there were difficulties in

326   accurately resolving individual nuclei and the resulting watershed approach performed poorly

327   (**Figure S4**). To resolve this, pixel classification machine learning in the Ilastik software was used to

328   convert these images into probability maps of 'cell outlines', 'intracellular environments' or

329   'background / other' (shown, **Figure S2**). The intracellular probability map was then directly

330   segmented into cell objects in CellProfiler using a IdentifyPrimaryObjects module (**Figure 1E-L**). Of

331   note, this latter approach (i) only required cell outline information (*i.e.,* actin or WGA) for effective

segmentation, freeing up the nuclear channel for other targets and (ii) was compatible with lower-resolution input images (*e.g.,* **Figure 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 labelling 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, **Figure 1**). For all analyses, histological information alongside imaging specifics and the cell segmentation strategy used are summarised 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) (>1000 cells scored; **Figure S5**). This benchmarking was carried out without first removing mis-segmented cells. Median scores in terms of pixel overlap were consistently between 0.80 – 0.83, with scores of ~ 0.9 recognised 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, **Figure 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, **Figure 1 B/F/J**) of each cell-object's 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, **Figure 1** -

grey 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-labelled *E. coli* strain LF82 into human ileum was considered (**Figure 2**). Three

tissue samples taken from one non-IBD patient with colon cancer, and one IBD patient with

macroscopically non-inflamed 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 (**Figure 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 (**Figure 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, **Figure 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 (**Figure 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 normalised histogram of the signal in the anti-GFP channel (total cells analysed = 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 (**Figure 2D**, 14,671 cells analysed). 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 (**Figure 2D**, 15,226 cells analysed). Comparison of this increase relative to the non-IBD group showed significance at the $p < 0.001$ level (Wilcoxon rank sum, **Figure 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 **Figure 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.

416

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 ~ 4,000 x 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 for the non-expert 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 intra-epithelial 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 immunolabelled for epithelial cell adhesion molecule (EPCAM) –

alongside nuclei and actin. This precisely pinpointed the location of the epithelial region between the

444   basement membrane and the apical enterocyte surface (29-31) (**Figure 3A**). Using this EPCAM

445   labelling as a guide to inform pixel annotation, we then trained an Ilastik machine learning model to

446   mask the epithelium, as well as the lumen and lamina propria tissue compartments, *directly* from the

447   actin channel itself. In this way, the EPCAM labelling was no long required (**Figure 3B**) (process

448   exemplified stepwise, **Figure S6**). Of note, we also found that the same approach worked with WGA

449   labelling in FFPE sections (demonstrated, **Figure S6E**).

450

451   **Utilising Locational and Cellular Information: Profiling Intraepithelial Lymphocytes**

452   Next, we set out to utilise both tissue compartment and per-cell image-data to profile intraepithelial

453   T lymphocytes in the jejunal mucosa. In the second serial section, a 112-image tilescan containing a

454   wide region of villous mucosa was collected with anti-CD3 labelling to identify T cells. As both the

455   EPCAM and CD3 antibodies were raised in the same host, instead of dual-labelling, the tissue

456   compartment model was deployed in the mucosa to provide a mask for the epithelium (**Figure 3C**).

457

458   To understand and quantify background fluorescence, as well as the non-specific binding capacity

459   of the CD3 antibody in the rat jejunal tissue, set of ten image-fields for either the secondary-

460   antibody alone (*i.e.,* 2° only control), or the secondary plus an irrelevant primary antibody of the

461   same isotype (*i.e.,* an isotype control) were collected for the CD3 channel in adjacent, serial

462   sections. Per-cell immunofluorescence data was then extracted from the CD3 tilescan (~ 60,000

463   cells) and control image-sets (~ 6,000 cells) using the watershed cell segmentation pipeline

464   optimised for mucosal tissue (above). A CD3$^+$ cell population was then formed by gating cells with

465   per-cell fluorescence values greater than those observed in the 2°-only and isotype controls (*i.e.,* as

466   is typical in flow cytometry) (**Figure 3D**). Cell centroid markers were displayed on each gated cell, to

467   help pinpoint CD3$^+$ T lymphocytes both visually and for subsequent locational categorisation..

468   Interfacing this gated cell population with the epithelial mask allowed further division of the CD3$^+$ cell

469   population into intra-epithelial lymphocyte (IEL) and lamina propria T cell subpopulations by

470   identification of cells with centroids inside or outside of the mask (**Figure 3D**). Upon close study of

471   the defined IEL CD3$^+$ cells in context of the masked epithelium, it was clear that this cell population

472 existed in two distinct forms. IEL events were either observed in close association with the basal

473 aspect of enterocytes (hereafter termed 'IEL^{sub}'), or, were *truly* between individual enterocytes

474 (hereafter termed 'IEL^{inter}'). To split the IELs into these two classes, the epithelial mask was

475 subjected to a morphological process called skeletonisation. This reduced the epithelial mask to

476 yield a central path through each 'loop' of villus epithelium (process exemplified, **Figure S6**).

477 Inclusion within this sub-mask allowed the central, IEL^{inter} population to be separated out, leaving

478 behind the IEL^{sub} cells (**Figure 4A-E**).

479

480 In this way, harnessing per-cell fluorescence data in combination with the precise histological

481 context provided by the high-resolution tilescan allowed the identified CD3$^+$ cells to be subdivided

482 into three distinct subpopulations (*i.e.,* lamina propria CD3$^+$ (LP^{CD3+}), IEL^{sub} and IEL^{inter}). This,

483 alongside the segmentation of all cells, whether immunolabelled or not, provided data well suited to

484 automated cell counting in the context of a tissue map. Hence, we measured the areas occupied by

485 the different designated compartments – alongside their cell counts – in total, per 100 cells, and as

486 ratios between the different tissue compartments (**Figure 4F-I**). Interestingly, whilst not so apparent

487 visually, the epithelium occupied a greater area (**Figure 4F**) and contained more total cells (**Figure**

488 **4G/I**) than the underlying lamina propria. CD3$^+$ cells were also determined more abundant per-cell

489 in the lamina propria than in the epithelium (**Figure 4I**). Meanwhile, whereas IELs were quite

490 common, the IEL^{inter} sub-class were rare events (~ 4 per 100 epithelial cells). This was especially

491 true when compared to the IEL^{sub} class, at ~ 13 per 100 epithelial cells (**Figure 4H/I**).

492

493 **Cell Interactions and Expression: Mapping in the Peyer's Patch**

494 Access to per-cell immunofluorescence data collected *in situ* provides the opportunity to consider

495 both cell expression and physical cellular interactions via nearest-cell neighbour analyses.

496 Lymphoid tissues represent one such environment in which interaction and expression data are of

497 key importance. Here we considered CD11c$^+$ mononuclear phagocyte – CD3$^+$ T cell expression

498 and interactions in a transverse section of mouse Peyer's patch (24 image tilescan, ~ 16,000 cells)

499 (**Figure 5**). A basic overview of the structure and cellular zonation of the murine Peyer's patch is

500 provided in **Figure S7**.

501

502 Image-data were collected for six channels: fluorescence data were collected for nuclei, actin,

503 CD11c (for mononuclear phagocytes; *i.e.,* antigen presenting cells) and CD3 as a pan T-lymphocyte

504 marker. Alongside, transmitted and reflected light were also collected to inform on overall histology

505 and section quality (**Figure 5A**). As before, data for the respective 2°-only and isotype controls were

506 also collected alongside in tissue-matched serial sections. As per-cell immunofluorescence

507 quantification was to be carried out on two of the channels (*i.e.,* CD3 and CD11c), leave-one-out

508 control image-sets were also taken to check for any fluorescence cross-talk between channels. This

509 involved labelling additional serial sections with either CD11c or CD3, yet collecting the respective

510 fluorescence data for both channels. In this way, any cross-talk into the 'empty' channel could be

511 detected in the resultant per-cell fluorescence distributions.

512

513 Using the Ilastik / CellProfiler machine learning cell segmentation pipeline, alongside the software

514 reassembly (tilescan) functions described above, the lymphoid tissue was segmented seamlessly

515 across the entire Peyer's patch (**Figure 5B**). A region-of-interest (ROI) was then set around the

516 lymphoid tissue, and just the CD11c and CD3 immunofluorescence data were shown on top of the

517 segmented-cell outlines inside the ROI. Outside of the ROI, just the actin staining was displayed, to

518 provide histological context (**Figure 5C/D**). This visualisation approach was found to dramatically

519 reduce the visual complexity of the six-channel image, permitting display of most important

520 information in a per-cell and visually intuitive manner – across the scale of the entire lymphoid

521 follicle.

522

523 To build $CD3^+$ and $CD11c^+$ cell populations, after debris removal (discussed above), gating was first

524 used to select cells with fluorescence values above those observed in the 2°-only and 'leave-one-

525 out' controls. The fluorescence distributions of the isotype controls were also used to inform gating.

526 Here, whilst we gated above values high enough to remove $> \sim 99\%$ of cells from the isotype

527   distributions, gating at the maximum was avoided for fear of building highly specific, yet poorly

528   sensitive cell populations (**Figure 5E**). Due to the closely packed cells, and in conjunction with the

529   expression of CD11c and CD3 immunofluorescence on the cell membrane, it was found that adding

530   a second sequential gate on the area of fluorescence within each cell helped to reduce 'bystander-

531   positive' events caused by small amounts of fluorescence spanning the segmented-cell outlines and

532   manifesting in immediately adjacent neighbouring cells (further discussion/exemplification provided,

533   **Figure S8**). In this way, cells exhibiting CD marker fluorescence all around their perimeters were

534   better isolated from their immediate neighbours, whilst maintaining sensitivity (**Figure 5F, inset**). To

535   aid this second gating step, cell-centroid markers for the identified cell populations were placed onto

536   the immunofluorescence images, as described above, providing visual feedback (**Figure 5F**). As

537   expected, the sub-epithelial dome (SED) was rich in mononuclear phagocytes and the inter-follicular

538   region (IFR) at the right of the image contained large numbers of T-cells. Surprisingly however, a

539   population of highly juxtaposed, CD11c-CD3 neighbouring cells (*i.e.,* region shown in **Figure 5D**)

540   that *still* identified positive in both gates after bystander removal were identified, indicating an

541   interaction (17) relative to other cells, and suggesting a likelihood of cell-cell communication (**Figure**

542   **5F**).

543

544   In addition to placing markers on cell centroids to delineate the gated cell populations (**Figure 6A**),

545   other methods capable of clearly visualising the single-cell data and consequent spatial

546   relationships across the scale of the complete Peyer's Patch were sought. In **Figure 6B**, the

547   marker-placement view was simplified further by flood-filling the individual segmented cell masks to

548   clearly show the populations in a manner that could be effectively visualised at small size. The

549   absence of immunofluorescence labelling (*i.e.,* black, CD11c$^-$ / CD3$^-$ regions) was also informative,

550   as within the patch, the vast majority of these double-negative cells will be B lymphocytes (32).

551   Next, the flood-filled view was simplified further to only show CD11c cells with touching CD3

552   nearest-cell neighbours (including juxtaposed CD11c/CD3 cells) (**Figure 6C**). In this way, the view

553   gives a sense of the spatial distribution of APCs within interactive distances of T lymphocytes.

554   Interestingly, it was observed that the majority of these events were predominantly congregated

555   around B cell follicles in the germinal center (GC) region, and were much less apparent in the SED

556   where, probably, MNP – B cell interactions may predominate (33).

557

558   Having successfully identified populations of cells, next we moved forwards to consider

559   quantification of per-cell fluorescence (*i.e.,* related to protein expression). To do this, we made use

560   of the ~ 800 nm optical Z plane afforded by the confocal optics and high numerical aperture

561   objective (63X/1.4) to isolate a thin plane through *individual* cells. The analysis was also aided by

562   the ability to select for cell objects optimally cross-sectioned through their central plane during the

563   debris removal step (discussed above), as this improved measurement consistency by sampling

564   data from similar, central regions in each cell. To clearly visualise the data from across the whole

565   lymphoid follicle within a reasonable figure-size, the per-cell expression of CD11c and CD3 was

566   displayed in four intensity bands (*i.e.,* dim, low, intermediate and high) (**Figure 6D/E**). Perhaps

567   unsurprisingly given the highly mixed population of mononuclear phagocytes delineated by CD11c,

568   no clear spatial patterning according to CD11c expression was observed (**Figure 6D**). For CD3

569   however, the IFR at the right of the patch, in addition to the APC and T-cell zones around the GC

570   were rich in CD3$^{int/hi}$ events, whilst the marginal zone and SED where predominated by CD3$^{dim/lo}$.

571   This may be related to T cell sub-types, or activation, and deserves further scrutiny (34).

572

573   Finally, we also sought a method to statistically identify significant spatial congregations of cells so

574   that regions of cellular zonation / established cellular microenvironments could be defined across

575   the lymphoid follicle. To do this we harnessed both the cell location and CD11c or CD3 per-cell

576   expression data and used these to calculate the Getis-Ord GI* spatial statistic (28). This provided a

577   heat map identifying where statistically significant, spatial congregations of different cell types

578   occurred relative to what should be expected by random chance – given the frequencies of the

579   different cell types involved (**Figure 6F**). As expected, the SED was significantly rich for CD11c, as

580   was the IFR for CD3. For both cell types however, the maps also revealed a wealth of complex

581   microstructure surrounding B-cell follicles in the GC. Under the 'steady state' normal biology

582 depicted here, it was also noted that the SED was sparse in terms of congregating CD3$^+$ T-

583 lymphocytes.

584

585 **DISCUSSION**

586 Here, with a specific focus on intestinal tissues, we develop open, user-friendly methodology that

587 enables per-cell quantification using routine confocal micrographs. As a methodological

588 advancement, it is important that the findings here are seen as a range of examples around

589 capability, rather than individually-powered biological studies. Notwithstanding, we flag areas where

590 the technique revealed interesting findings, including measures of spatially distinct IEL sub-

591 populations – being either between or beneath enterocytes in the villus mucosa – and the complex

592 microstructure of cellular zonation in the Peyer's patch, including spatial distributions of APC-T cell

593 interactions.

594

595 Our image-based cell profiling approach delivers data in three key ways: (i) it enumerates different

596 cell types, as in flow cytometry, but it also (ii) provides precise cellular locational data with

597 histological context and (iii) resolution and quantification of cell contents. To achieve this, a number

598 of novel approaches had to be developed or bridged together. Firstly, we provide routes in MATLAB

599 or Python to enable tilescan processing with CellProfiler – to include spatial reassembly of the

600 mined per-cell data and the production of global segmentation masks with unique cell identities to

601 enable visualisations. Secondly, to permit accurate cell segmentation, antibody-independent cell

602 labelling was employed, which we optimised for both FFPE and frozen intestinal tissue sections.

603 Importantly this approach does not use up antibody hosts and transfers easily and directly between

604 species. Next, effective per-cell immunofluorescence analysis requires accurately segmented cells

605 with mis-segmented cells (debris) excluded. We show how this can be achieved using density plots

606 to refine a consistently sampled cell population, with the outliers (partial cells or doublets) excluded.

607 Alongside, to tackle the difficult issue of dense cell packing in lymphoid tissues, we use semantic

608 machine learning within a fast, user-friendly framework (23) to yield accurate cell segmentations.

609 Finally, we demonstrate how inadvertent bystander-positive cells can be obviated through

610 sequentially gating on fluorescence intensity followed by fluorescence area, and suggest that any

611 remaining bystanders are indicative of cell-cell interactions. We also demonstrate how spatial

612 statistics can be employed to better define tissue microenvironments in terms of identifying

613 significant cellular congregations.

614

615 Here we use the popular CellProfiler software as a 'backbone' to enable per-cell quantification.

616 Importantly, its pipeline-based style is extremely flexible, and can use original, deconvolved and / or

617 spectrally unmixed input images from almost any microscope or upstream software package.

618 Moreover, the pipelines provided here can also utilise probability maps to enable cell segmentation

619 from any source, including, where necessary, more advance machine learning approaches such as

620 deep convolutional neural networks (26). Of note, a delivery of the increasingly popular Unet

621 architecture within the user-friendly environment of Ilastik is planned for release in spring 2020 (35).

622

623 In embracing such a technical approach to tissue analysis, it is critical that the fundaments of robust

624 immunofluorescence methodology are not overlooked. In our experience, best possible tissue

625 orientation helps greatly in interpretation of outputs. Moreover, our approach does not obviate good

626 practice in labelling and imaging: rather, success relies upon it. Controls, to assess background

627 autofluorescence and non-specific antibody binding, are extremely important, alongside assurance

628 that fluorescence signals do not cross between channels. Here, we have demonstrated the use of

629 2°-only, isotype, leave-one-out and biological negative controls collected in tissue-matched, serial

630 sections to assess these parameters. We then use Otsu thresholding in conjunction with gated

631 analyses to accurately isolate cell populations and measure cell expression / contents. Of course,

632 these controls in themselves do not ensure that the correct target is being labelled and, as always in

633 such work, proper validation of antibodies remains essential (36,37).

634

635 Our accessible approach to per-cell analysis of tissue sections contrasts with other techniques.

636 Whilst imaging mass cytometry (*e.g.,* 'CyTOF'), enables the use of dozens of antibody markers, it

637 has lower spatial resolution and necessitates highly specialist instrumentation for detection (6,18).

638   Although high throughput and extremely powerful within diagnostic pathology, packages that permit

639   the analysis of chromogen stained slide-scans lack the resolution, sensitivity, ability to multiplex and

640   quantitate immunolabelled targets in a way that is often required for precision research (12).

641   Meanwhile, commercial 'all-in-one' solutions, such as those employing fluorescence slide scanners

642   or spinning disk confocal techniques, are (i) expensive, especially when highly capable confocal

643   microscopes are already available at most research institutions and (ii) rely on software with the

644   unenviable task of enabling the analysis of all conceivable tissue types. In our experience, this

645   results in approximate per-cell measurements. In contrast, focussing in one field and interfacing

646   different strategies (18), as we do here with the intestine, enables precision cell segmentation to be

647   achieved and thus accurate analyses of cellular localisation, per-cell content and cell-cell

648   interactions.

649

650   Finally, some of the original, pioneering, work in quantitative, flow cytometry-type

651   immunofluorescence analysis of tissues (*e.g.,* histocytometry (1-3,16,17)) relies upon commercial

652   software for implementation, limiting accessibility. Moreover, whilst the histocytometry approach

653   utilises both 2-D and 3-D confocal images, analyses have primarily focussed on the spatial

654   relationships of just the CD-marker delineated cells. Our approach, supported by machine learning

655   segmentation where necessary, enables precision analyses of *all* cells and hence highly accurate

656   cell counting and per-cell quantifications within an entire section or region-of-interest.

657   Notwithstanding, here we focus on open-source, intestinal-specific 2-D delivery, as qualitative

658   display in this format is today's gold standard, and because volumetric (*i.e.,* 3-D)

659   immunofluorescence quantification is extremely challenging for routine usage, given the time

660   requirements and increasing non-uniformities that manifest with imaging depth. With this in mind,

661   the use of the confocal optical section provides 2-D immunofluorescence data that is *consistently*

662   sampled, and thus well-suited for summation within cell-objects and for fair comparison across

663   experimental samples. In turn, an important question for future work may involve addressing how far

664   regular fluorescence images (*i.e.,* non-confocal) can be taken towards producing similar,

665   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).

694   **BIBLIOGRAPHY**

695

696   1.   Ecker RC, Steiner GE. Microscopy-based multicolor tissue cytometry at the single-cell level.

697   Cytometry A 2004;59:182-90.

698

699   2.   Harnett MM. Laser scanning cytometry: understanding the immune system in situ. Nat Rev

700   Immunol 2007;7:897-904.

701

702   3.   Gerner MY, Kastenmuller W, Ifrim I, Kabat J, Germain RN. Histo-cytometry: a method for

703   highly multiplex quantitative tissue imaging analysis applied to dendritic cell subset

704   microanatomy in lymph nodes. Proc Natl Acad Sci U S A 2012;37:364-76.

705

706   4.   Bankhead P, Loughrey MB, Fernandez JA, Dombrowski Y, McArt DG, Dunne PD. QuPath:

707   Open source software for digital pathology image analysis. Sci Rep 2017;7:16878.

708

709   5.   McQuaid S, Gray RT, Murray LJ, Coleman HG, James JA, Salto-Tellez M, Hamilton PW,

710   Blom S. Systems pathology by multiplexed immunohistochemistry and whole-slide digital

711   image analysis. Sci Rep 2017;7:15580.

712

713   6.   Schapiro D, Jackson HW, Raghuraman S. histoCAT: analysis of cell phenotypes and

714   interactions in multiplex image cytometry data. Nat Methods 2017;14:873-876.

715

716   7.   Morita N, Umemoto E, Fujita S, Hayashi A, Kikuta J, Kimura I, Haneda T, Imai T, Inoue A,

717   Mimuro H and others. GPR31-dependent dendrite protrusion of intestinal CX3CR1(+) cells

718   by bacterial metabolites. Nature 2019;566:110-114.

719

720   8.   Moor AE, Harnik Y, Ben-Moshe S, Massasa EE, Rozenberg M, Eilam R, Bahar Halpern K,

721   Itzkovitz S. Spatial Reconstruction of Single Enterocytes Uncovers Broad Zonation along the

722   Intestinal Villus Axis. Cell 2018;175:1156-1167.e15.

723

724   9.   Eisenhoffer GT, Loftus PD, Yoshigi M, Otsuna H, Chien CB, Morcos PA, Rosenblatt J.

725   Crowding induces live cell extrusion to maintain homeostatic cell numbers in epithelia.

726   Nature 2012;484:546-9.

727

728   10.   Da Silva C, Wagner C, Bonnardel J, Gorvel JP, Lelouard H. The Peyer's Patch Mononuclear

729   Phagocyte System at Steady State and during Infection. Front Immunol 2017;8:1254.

730

731   11.   Cummings RJ, Barbet G, Bongers G, Hartmann BM, Gettler K, Muniz L, Furtado GC, Cho J,

732   Lira SA, Blander JM. Different tissue phagocytes sample apoptotic cells to direct distinct

733   homeostasis programs. Nature 2016;539:565-569.

734

735   12.   Robertson D, Savage K, Reis-Filho JS, Isacke CM. Multiple immunofluorescence labelling of

736   formalin-fixed paraffin-embedded (FFPE) tissue. BMC Cell Biol 2008;9:13.

737

738   13.   Weigert M, Schmidt U, Boothe T, Müller A, Dibrov A, Jain A, Wilhelm B, Schmidt D,

739   Broaddus C, Culley S and others. Content-aware image restoration: pushing the limits of

740   fluorescence microscopy. Nat Methods 2018;15:1090-1097.

741

742   14.   Carpenter AE, Kamentsky L, Eliceiri KW. A call for bioimaging software usability. Nat

743   Methods 2012;9:666-70.

744

745   15.   Carpenter AE, Jones TR, Lamprecht MR, Clarke C, Kang IH, Friman O, Guertin DA, Chang

746   JH, Lindquist RA, Moffat J and others. CellProfiler: image analysis software for identifying

747   and quantifying cell phenotypes. Genome Biol 2006;7:R100.

748

749 16. Li W, Germain RN, Gerner MY. Multiplex, quantitative cellular analysis in large tissue

750   volumes with clearing-enhanced 3D microscopy (Ce3D). Proc Natl Acad Sci U S A

751   2017;114:E7321-e7330.

752

753 17. Kotov DI, Pengo T, Mitchell JS, Gastinger MJ, Jenkins MK. Chrysalis: A New Method for

754   High-Throughput Histo-Cytometry Analysis of Images and Movies. J Immunol 2019;202:300-

755   308.

756

757 18. Schuffler PJ, Schapiro D, Giesen C, Wang HA, Bodenmiller B, Buhmann JM. Automatic

758   single cell segmentation on highly multiplexed tissue images. Cytometry A 2015;87:936-42.

759

760 19. Caicedo JC, Cooper S, Heigwer F, Warchal S, Qiu P, Molnar C, Vasilevich AS, Barry JD,

761   Bansal HS, Kraus O and others. Data-analysis strategies for image-based cell profiling. Nat

762   Methods 2017;14:849.

763

764 20. Keita AV, Gullberg E, Ericson AC, Salim SY, Wallon C, Kald A, Artursson P, Soderholm JD.

765   Characterization of antigen and bacterial transport in the follicle-associated epithelium of

766   human ileum. Lab Invest 2006;86:504-16.

767

768 21. Moussata D, Goetz M, Gloeckner A, Kerner M, Campbell B, Hoffman A, Biesterfeld S,

769   Flourie B, Saurin JC, Galle PR and others. Confocal laser endomicroscopy is a new imaging

770   modality for recognition of intramucosal bacteria in inflammatory bowel disease in vivo. Gut

771   2011;60:26-33.

772

773 22. kwikteam / npy-matlab: code to read and write NumPy's NPY format in MATLAB. GitHub

774   repository. Available at https://github.com/kwikteam/npy-matlab Accessed March 25th, 2020.

775

776

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 2011:230-3.

24. Maska M, Ulman V, Svoboda D, Matula P, Matula P, Ederra C, Urbiola A, Espana T, Venkatesan S, Balak DM and others. A benchmark for comparison of cell tracking algorithms. Bioinformatics 2014;30:1609-17.

25. Everingham M, Van Gool L, Williams CKI, Winn J, Zisserman A. The Pascal Visual Object Classes (VOC) Challenge. Int J Compt 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 and others. 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. Geographical Analysis 1995;27:286-306.

29. Dobbins WO, 3rd. Human intestinal intraepithelial lymphocytes. Gut 1986;27:972-85.

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.

804

805   31.   Rostami K, Marsh MN, Johnson MW, Mohaghegh H, Heal C, Holmes G, Ensari A, Aldulaimi

806   D, Bancel B, Bassotti G and others. ROC-king onwards: intraepithelial lymphocyte counts,

807   distribution & role in coeliac disease mucosal interpretation. Gut 2017;66:2080-2086.

808

809   32.   De Jesus M, Ahlawat S, Mantis NJ. Isolating and immunostaining lymphocytes and dendritic

810   cells from murine Peyer's patches. J Vis Exp 2013:e50167.

811

812   33.   Reboldi A, Arnon TI, Rodda LB, Atakilit A, Sheppard D, Cyster JG. IgA production requires B

813   cell interaction with subepithelial dendritic cells in Peyer's patches. Science

814   2016;352:aaf4822.

815

816   34.   Valle A, Barbagiovanni G, Jofra T, Stabilini A, Perol L, Baeyens A, Anand S, Cagnard N,

817   Gagliani N, Piaggio E and others. Heterogeneous CD3 expression levels in differing T cell

818   subsets correlate with the in vivo anti-CD3-mediated T cell modulation. J Immunol

819   2015;194:2117-27.

820

821   35.   Kreshuk A. Automating image analysis with machine learning. Microscience Microscopy

822   Congress. Manchester, UK; 2019.

823

824   36.   Uhlen M, Bandrowski A, Carr S, Edwards A, Ellenberg J, Lundberg E, Rimm DL, Rodriguez

825   H, Hiltke T, Snyder M and others. A proposal for validation of antibodies. Nat Methods

826   2016;13:823-827.

827

828   37.   Baker M. Reproducibility crisis: Blame it on the antibodies. Nature 2015;521:274-6.

829

830

831

832   **FIGURE LEGENDS**

833   **Figure 1 - Open-source cell segmentation strategies for diverse intestinal tissues**. **A-D**,

834   Mouse villus mucosa frozen section. Here, we used a simple watershed approach that first defines

835   the nucleus (gold, Hoechst 33342) of each cell, and then uses this as an anchor point from which to

836   find each cell's actin-delineated boundary (grey, actin-AF633). **E-H**, Human Peyer's patch lymphoid

837   tissue; formalin fixed paraffin embedded (FFPE) section. Exposure to alcohol during the FFPE

838   process destroys the actin microfilaments (see **Figure S3**) so, instead, cell membranes were

839   labelled using wheat germ agglutinin (WGA-AF633, blue). Watershed algorithms perform poorly in

840   such densely-packed tissue types (shown, **Figure S4**), and so machine learning via the Ilastik

841   software was instead used to produce probability maps of the cell outlines to enable segmentation

842   (training shown in **E,** inset / process fully described, **Figure S2**). **I- L**, Rat mesenteric lymph node

843   frozen section. Despite lower magnification and image resolution, the same machine learning

844   based, Ilastik-CellProfiler process enables accurate cell segmentation. **B**/**F**/**J** - insets, Density

845   plotting each cell's nuclear and cell outline fluorescence provides a straightforward approach to

846   'gate out' incorrectly segmented cell objects with abnormally high (*e.g.,* doublets) or low (*e.g.,*

847   debris) signals. Example discarded events that lie outside of the indicated 'single-cell population'

848   are indicated with gray squares on the tissue images. For all examples, segmentation accuracy

849   scores are provided in **Figure S5**. Scale bars = 20 microns.

850

851   **Figure 2 - Conquering the representative image: *E. coli* passage into human control or IBD**

852   **ileal mucosa. A-C**, Mucosal tissue samples from the distal ileum of either a non-IBD cancer control

853   patient (**B**) or an IBD patient (Crohn's disease) (**A** and **C**) were maintained *ex-vivo* in Ussing

854   chambers. The apical sides of the living tissues were exposed to either Krebs buffer (**A**) or GFP-

855   expressing, adherent invasive *E. coli* strain LF82 (**B/C**) for 20 minutes. Images of the mucosa were

856   then collected randomly across 6-8 frozen sections per tissue sample prior to per-cell analysis for

857   anti-GFP fluorescence to identify LF82. **D/E**, Instead of relying on representative images, the image-

858   based cell profiling approach allowed quantification and display of *all* of the collected data (> 35,000

859   cells analysed in total). **D**, Concordance between the tissue samples taken per patient was

860   observed (filled circles). Whilst the fluorescence distributions (lines represent averages) from the

861   non IBD cancer control samples directly overlaid the Krebs buffer negative control (dashed line)

862   suggesting no transmucosal uptake, the samples from the IBD patient all showed elevated

863   fluorescence. **E**, Comparison between non-IBD and IBD groups for gated cells (indicated, **D**) with

864   anti-GFP values greater than present in Krebs buffer control. (***) Indicates statistical significance at

865    $p < 0.001$ (Wilcoxon rank sum). Scale bars: **A-C** = 10 µm.

866

867   **Figure 3 - Identifying intraepithelial T lymphocytes in large tilescans using a single CD**

868   **marker.** Rat jejunal longitudinal tissue section. **A**, First, anti-epithelial cell adhesion molecule

869   (EPCAM) immunofluorescence labelling was used to delineate the epithelium (*i.e.,* cells lying

870   between the basement membrane and the apical enterocyte surface). **B**, As the anti-EPCAM

871   antibody was raised in the same host species as the desired lymphocyte marker, the epithelium,

872   lamina propria and lumen 'compartments' were directly-detected from the actin channel using pixel-

873   classification machine learning in Ilastik (process outlined in **Figure S6**). **C**, A 112-image confocal

874   tilescan labelled for nuclei, actin and anti-CD3 was collected. Each individual field was segmented

875   into individual cells, and a software function was developed to spatially reassemble the images,

876   segmentation masks and cell positions (> 60,000 cells). **D**, A region-of-interest (ROI) was placed

877   around the tissue region containing optimally cross-sectioned villi, and the Ilastik model was used to

878   predict and mask the epithelium (Pink). **D** - inset, CD3[+] cells inside or outside of this epithelial mask

879   were then identified by gating against the secondary-only and isotype control per-cell fluorescence

880   distributions (*i.e.,* as is typical in flow cytometry). Cell centroid markers were placed on each positive

881   event. This approach permitted sensitive and accurate pinpointing of CD3[+] lymphocytes (**C**, inset).

882   Scale bars: **A/B** = 25 µm; **C** = 50 µm; **D** = 1 mm.

883

884   **Figure 4 - Defining and counting T lymphocyte subpopulations on the basis of spatial**

885 **location**. Rat jejenum, continued from **Figure 3**. **A**, To differentiate intraepithelial (IEL) CD3$^+$ cells

886 in close association with the apical side of the epithelium (*i.e.,* 'IEL^{sub}') versus those *truly* in

887 between enterocytes (*i.e.,* 'IEL^{inter}'), the epithelial mask defined in **Figure 3** was skeletonised and

888 dilated to form a central path enabling detection of IEL^{inter} cells in each 'loop' of epithelium (method

889 shown, **Figure S6**). **B/C**, The CD3$^+$ cell population was then split into three subpopulations: lamina

890 propria CD3$^+$ (LP^{CD3+}), IEL^{sub} or IEL^{inter}, and locations were displayed by marker placement on

891 each cell's centroid. **D/E,** Typical examples of IEL^{sub} and IEL^{inter} events. **F**, Area measurements

892 for the different tissue compartments. **G**, Total cell counts according to tissue compartment. **H**,

893 Tissue compartment cell counts per 100 cells. **I**, Cell counts expressed as ratios between the

894 different tissue compartments. Scale bars: **A/C** = 100 µm; **B** = 500 µm; **D/E** = 5 µm.
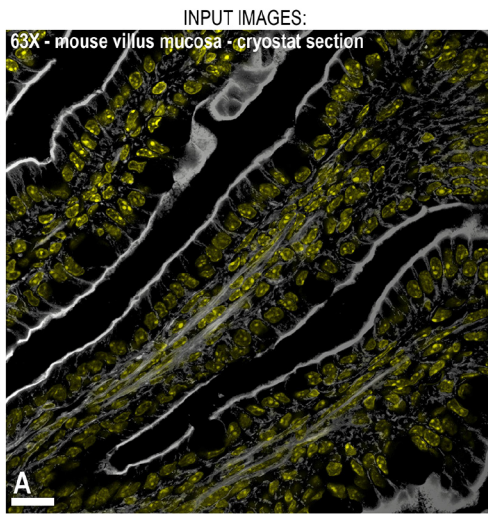
895

896 **Figure 5 - CD11c and CD3 expression in Peyer's patches**. Mouse ileal transverse section. **A**,

897 The section was imaged as a 24- image tilescan labelled for nuclei, actin, anti-CD11c and anti-CD3.

898 Transmitted and reflected light were also collected. **B**, The densely packed, lymphoid tissue

899 segmented accurately into cells (~ 16,000) using the Ilastik/CellProfiler machine learning approach,

900 and the software reassembly function spatially reassembled the data without artifacts at the overlap

901 boundaries between the tiled images. **C/D**, The complexity of multichannel image (**A**) is dramatically

902 reduced, using an example region-of-interest (yellow), and displaying the immunofluorescence

903 markers (cyan = CD11c, magenta = CD3) on the segmented-cell outlines. **E/F**, CD3$^+$ and CD11c$^+$

904 cell populations were gated against secondary-only, leave-one-out and isotype control per-cell

905 fluorescence distributions. In the densely packed tissue, a second sequential gate on fluorescence

906 area per cell-object helped to reduce 'bystander-positive' events caused by fluorescence overlap

907 into neighbouring cells (shown, **F** – inset) (further discussion/exemplification provided, **Figure S8**).

908 **F**, Marker placement on the cell populations identified by the gating strategy in (**E**). Highly

909 juxtaposed, and thus indicative of communication, CD11c-CD3 neighbouring cells (*i.e.,* region

910 shown in **D**) that consequentially identified positive in *both* gates are shown with white markers.

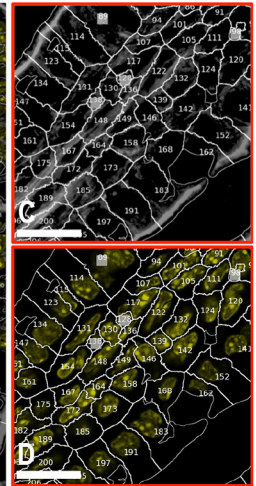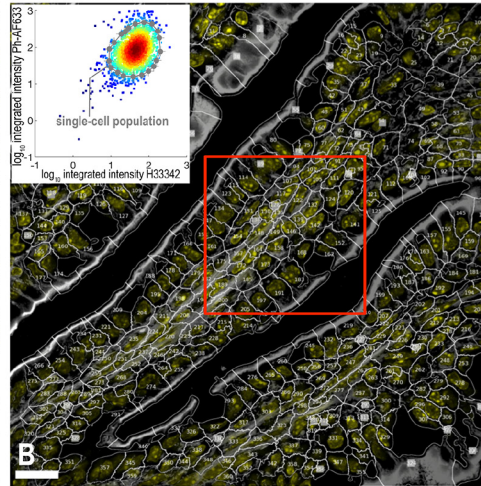911   Scale bars: **A/C/F** = 500 µm; **B** = 250 µm; **D** = 50 µm.

912

913   **Figure 6 - Mapping CD11c$^+$ and CD3$^+$ cell interactions, expression and spatial**

914   **congregations.** Mouse ileum, continued from Figure 5. **A**, Cell-marker placement on the basis of

915   immunofluorescence labelling. Highly juxtaposed CD11c-CD3 neighbouring cells that

916   consequentially identified positive for both markers are shown with white markers. **B**, Flood-filling

917   segmented cell objects provides a visually intuitive version of the data shown in (**A**) that can be

918   displayed at a much smaller size. **C**, Here, the view shown in (**B**) is simplified to only show CD11c

919   cells with touching CD3 nearest-cell neighbours, or juxtaposed cells that identify positive for both

920   markers due to close spatial association. **D/E**, CD11c and CD3 expression maps with cell-objects

921   coloured in four levels (*i.e.,* dim, lo, intermediate (int), hi) according to each segmented cell's level of

922   immunofluorescence. **F**. Getis-Ord statistical map: this shows – as a probability heat map – where

923   statistically significant congregations of cells are found relative to what would be expected by

924   random chance given the frequencies of the different cell types. Scale bars: **A-F** = 500 µm.
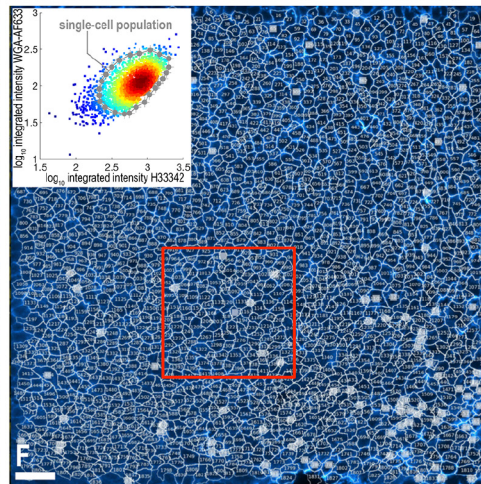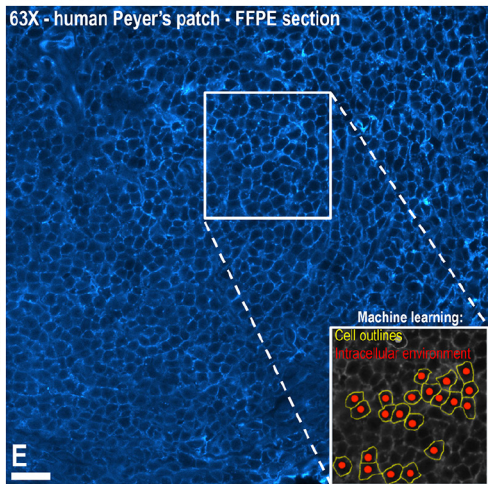
INPUT IMAGES:

63X - mouse villus mucosa - cryostat section

Phalloidin-AF633 / Hoechst 33342

SEGMENTED CELLS / ▢ DISCARDED EVENTS

63X - human Peyer's patch - FFPE section

WGA-AF633 / Hoechst 33342

Machine learning:
Cell outlines
Intracellular environment

40X - rat mesenteric lymph node - cryostat section

Phalloidin-AF633 / Hoechst 33342

Machine learning:
Cell outlines
Intracellular environment

single-cell population
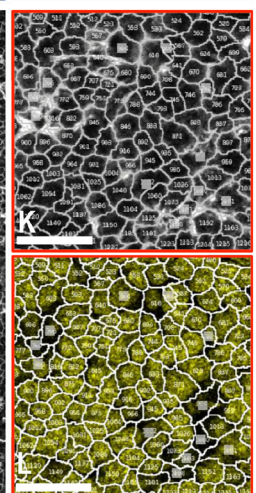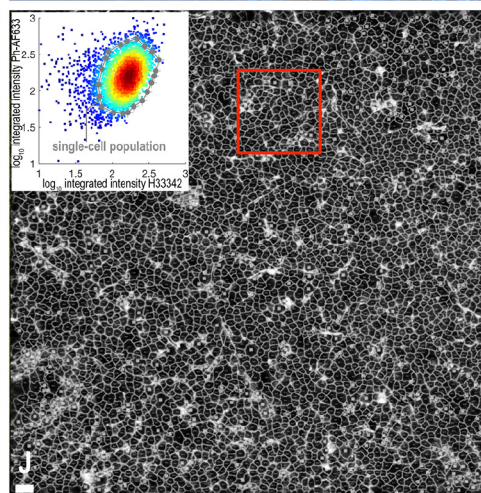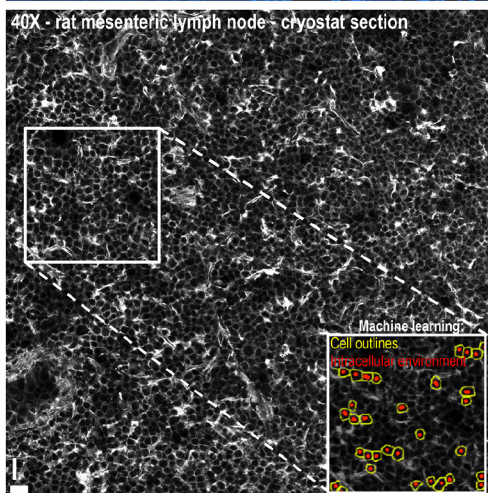
INPUT IMAGES:

Nuclei / Actin / **anti-GFP**

A — Krebs buffer

B — Non-IBD + LF82

C — IBD + LF82

D — Krebs / Non-IBD / IBD

Population %

anti-GFP: Fluorescence per cell

Positive cells

E — anti-GFP: fluorescence per cell

Non-IBD   IBD

***

A EPCAM determination of the epithelium

anti-EPCAM / Actin

B Machine learning recognition of the epithelium

Input: actin (intensity normalised)   Predictions: epithelium / LP / other
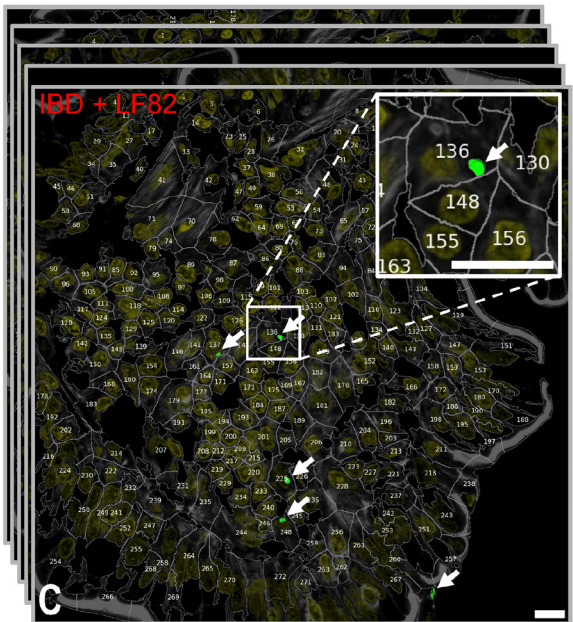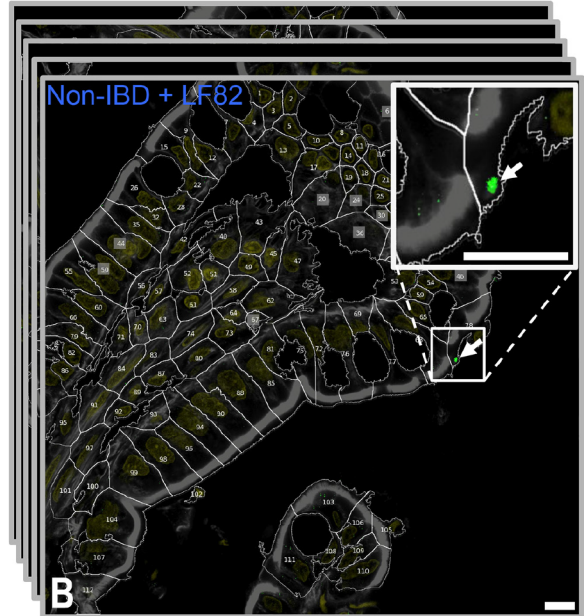
C Tilescan (27,792 x 15,125 px)

Nuclei
Actin
Epithelium
anti-CD3

ROI

anti-CD3

CD3 (& cell centroid markers)

CD3 / markers / nuclei

D Cell segmentation / epithelium / CD3+ cell centroid markers

Population %

— 2° control   --- isotype   —○— anti-CD3

CD3+ population

Thresholded AF488:
(integrated intensity per cell)

A ROI ML epithelium Skeletonised epithelium

B LP^CD3+ / IEL^sub / IEL^inter (cell centroid markers)
Actin
Nuclei

C

D Nuclei / Actin / anti-CD3
IEL^sub
epithelium

E IEL^inter
epithelium

F Area (mm²)
Within ROI
Tilescan ROI Epithelium LP

G Number of cells
Within ROI
Tilescan ROI
Epithelium IEL LP LP^CD3+
IEL IEL^sub IEL^inter

H Within ROI
Counts per 100 cells
LP^CD3+ (per 100 LP cells)
IEL (per 100 epithelial cells)
IEL^sub (per 100 epithelial cells)
IEL^inter (per 100 epithelial cells)

I Ratio (cell counts)
LP : Epithelium
LP^CD3+ : LP
IEL^sub : Epithelium
IEL : Epithelium
IEL^inter : Epithelium

Tilescan (11,253 x 7,607 px)
Nuclei / Actin / anti-CD11c / anti-CD3 / brightfield / darkfield (scatter)

A

Cell segmentation / nuclei / actin

B

Immunofluorescence:
anti-CD11c / anti-CD3
Actin (outside ROI)
Cell segmentation (inside ROI)

C

Cell segmentation / anti-CD11c / anti-CD3

D

E

Int. only: CD3+
Int. + Area: CD3+ population
2° control
isotype
anti-CD11c
anti-CD3
Population %
Thresholded AF568:
Integrated **intensity** per cell (a.u.)
Thresholded AF568:
Integrated fluorescence **area** per cell (a.u.)

Int. + Area:
CD11c+ population
Int. only:
CD11c+
2° control
isotype
anti-CD3
anti-CD11c
Population %
Thresholded AF488:
Integrated **intensity** per cell (a.u.)
Thresholded AF488:
Integrated fluorescence **area** per cell (a.u.)

CD11c+ / CD3+ / *juxtaposed* CD11c+ & CD3+ (cell centroid markers)
Actin (outside ROI)
Cell segmentation (inside ROI)

Int.   Int. + Area

F

A

CD11c⁺ / CD3⁺ (cell centroid markers)
juxtaposed CD11c⁺ & CD3⁺
Actin
Cell segmentation (inside ROI)

B Flood-filled cell map:
CD11c⁺ / CD3⁺
juxtaposed CD11c⁺ & CD3⁺

C Nearest-cell neighbour map:
CD11c⁺ touching CD3⁺
juxtaposed CD11c⁺ & CD3⁺

D CD11c⁺ expression map:
Dim | Lo | Int | Hi

E CD3⁺ expression map:
Dim | Lo | Int | Hi

F Getis-Ord spatial statistics:
Scale (d) = 175 pixels

z [Gᵢ (d)] score

CD11c   CD3