A SIMPLI (Single-cell Identification from MultiPLexed Images) approach for spatially resolved tissue phenotyping at single-cell resolution.

Michele Bortolomeazzi[1,2], Lucia Montorsi[1,2], Damjan Temelkovski[1,2], Mohamed Reda Keddar[1,2], Amelia Acha-Sagredo[1,2], Michael J. Pitcher[3], Gianluca Basso[4,5], Luigi Laghi[4,6], Manuel Rodriguez-Justo[7], Jo Spencer[3], Francesca D. Ciccarelli[1,2*]

[1] Cancer Systems Biology Laboratory, The Francis Crick Institute, London NW1 1AT, UK

[2] School of Cancer and Pharmaceutical Sciences, King's College London, London SE11UL, UK

[3] School of Immunology and Microbial Sciences, King's College London, London SE19RT, UK

[4] Laboratory of Molecular Gastroenterology, IRCCS Humanitas Research Hospital, Rozzano 20089 MI, Italy

[5] Genomic Unit, IRCCS Humanitas Research Hospital, Rozzano 20089 MI, Italy

[6] Department of Medicine and Surgery, University of Parma, Parma 43121 PR, Italy

[7] Department of Histopathology, University College London Cancer Institute, London WC1E 6JJ, UK

* = corresponding author (francesca.ciccarelli@crick.ac.uk)

ABSTRACT

Multiplexed imaging technologies enable to study biological tissues at single-cell resolution while preserving spatial information. Currently, the analysis of these data is technology-specific and requires multiple tools, restricting the scalability and reproducibility of results. Here we present SIMPLI (Single-cell Identification from MultiPlexed Images), a novel, technology-agnostic software that unifies all steps of multiplexed imaging data analysis. After processing raw images, SIMPLI performs a spatially resolved, single-cell analysis of the tissue as wells as cell-independent quantifications of marker expression to investigate features undetectable at the cell level. SIMPLI is highly customisable and can run on desktop computers as well as high-performance computing environments, enabling workflow parallelisation for the analysis of large datasets. It produces multiple outputs at each step, including tabular text files and visualisation plots. The containerised implementation and minimum configuration requirements make SIMPLI a portable and reproducible solution for multiplexed imaging data analysis. SIMPLI is available at: https://github.com/ciccalab/SIMPLI.

MAIN

A detailed investigation of tissue composition and function in health and disease requires spatially resolved single-cell approaches that precisely quantify cell types and states as well as their interactions *in situ*. Recent technological advances have enabled to stain histological sections simultaneously with multiple tagged antibodies that are subsequently detected using fluorescence microscopy or mass spectrometry[1]. High-dimensional imaging approaches such as imaging mass cytometry (IMC)[2], multiplexed ion beam imaging (MIBI)[3], co-detection by indexing (CODEX)[4], and multiplexed immunofluorescence (mIF, including cycIF)[5] enable quantification and localisation of cells in sections from formalin-fixed paraffin-embedded (FFPE) tissues, including clinical diagnostic samples. This is of particular value for the analysis of tissue-level characteristics of disease conditions and for predicting the outcome of therapies that depend on the tissue environment, such as cancer immunotherapy. For example, a recent IMC phenotypic screen of breast cancer subtypes associated the heterogeneity of somatic mutations with that of the tumour microenvironment[6]. Similarly, a CODEX-based profile of FFPE tissue microarrays from high-risk colorectal cancer patients correlated $PD1^{+}CD4^{+}$ T cells with patient survival[7].

The analysis of multiplexed images requires the conversion of pixel intensity data into single-cell data, which can then be identified phenotypically, quantified and localised in the tissue. Currently, each step of the analysis is accomplished by independent tools. For example, several computational approaches allow processing of raw images and extraction of single-cell data either interactively (Ilastick[8], CellProfiler4[9], CODEX Toolkit[4]) or via command line (imcyto[10], ImcSegmentationPipeline[11]). Different sets of tools can perform cell phenotyping (CellProfiler Analyst[12], Cytomapper[13], Immunocluster[14]) and analyse cell spatial

interactions (CytoMap[15], ImaCytE[16], SPIAT[17], neighbouRhood[18]). Each of these tools relies on specific configuration and input formats, making the analysis challenging for users with limited computational skills and restricting the scalability, portability and reproducibility in different computing environments. histoCAT++[19] and QuPath[20] can perform most of the required analytical steps for IMC and mIF, respectively. However, these platforms have been developed specifically for interactive use and are not well suited for the analysis of large datasets.

Here we introduce SIMPLI (Single-cell Identification from MultiPLexed Images), a tool that combines processing of raw images, extraction of cell data, and spatially resolved quantification of cell types or functional states into a single pipeline. SIMPLI is highly customisable allowing the adaptation of each step to *ad hoc* user requirements. It can be run on desktop computers as well as on high-performance-computing environments, where it can be easily applied to large datasets due to automatic workflow parallelisation. To demonstrate the flexibility of SIMPLI to work with different technologies and experimental conditions, we study the phenotypes and spatial distribution of cells in human colon mucosa, appendix and colorectal cancer images obtained with IMC and mIF.

RESULTS

Overview of the SIMPLI analytical workflow

SIMPLI divides the analysis of multiplexed imaging data into three steps (Methods, Fig. 1 and Supplementary Fig. 1), the first of which consists of processing the raw data from the experiments (Fig. 1a). Data from single-channel images are extracted, normalised and thresholded to reduce the background noise and derive

masks of tissue compartments or cell populations (Fig. 1a). The resulting images can then be analysed at the cell (Fig. 1b) or pixel (Fig. 1c) levels using two complementary approaches.

The cell-based analysis aims to investigate the qualitative and quantitative cell representation through (1) single-cell data extraction, (2) cell phenotyping and (3) spatial analysis of cell distances (Fig. 1b). To extract single-cell data, SIMPLI performs single-cell segmentation and derives the masks of the single cells together with the expression values for each marker. Cells can then be assigned to tissue compartments or predefined cell populations based on their overlap with previously derived masks. Moreover, their functional phenotype can be further characterised with two alternative approaches. The first applies unsupervised clustering to all cells or subsets of cells (for example those mapping to specific tissue compartments or belonging to predefined populations) using their marker expression levels. This leads to the unbiased identification of clusters of cells with similar expression profiles that correspond to specific functional phenotypes. The second approach identifies cells with designated phenotypes by combining user-defined thresholds to marker expression values. Finally, a spatial analysis of the distance between cells within the imaged tissue can be performed. This can involve the identification of aggregations of cells of the same (homotypic) or different (heterotypic) types. In the case of homotypic aggregations, SIMPLI identifies clusters of cells within a user-defined distance and visually localises them in the tissue image. In the case of heterotypic aggregations, SIMPLI computes the distance distribution between distinct cell types and compares them across cell types and experimental conditions.

The pixel-based approach implemented in SIMPLI enables the quantification of areas positive for a specific marker or combination of markers, independently of their

association with cells (Fig. 1c). These marker-positive areas can be normalised over the area of the whole image, or those of specific tissue compartments or cell populations using the predefined masks, to allow comparisons across samples. The pixel-based analysis is useful for the investigation of tissue features that are not detectable at the cell level. For instance, extracellular or secreted proteins cannot be quantified with approaches dependent on cell segmentation. In addition, being completely cell agnostic, the pixel-based analysis can provide an independent validation of cell-based observations.

Quantification of secreted and cell-associated IgA in human colon mucosa

As a first case study, we stained six sections of human normal colon mucosa (CLN1-CLN6, Supplementary Table 1) with 26 antibodies marking four immune cell populations (T cells, macrophages, dendritic cells and B cells) as well as stromal components (Supplementary Table 2). Single regions of interest (ROI) in stained sections were ablated using the Hyperion Imaging System to obtain IMC data that we analysed with SIMPLI.

We extracted and normalised the 26 single channel images for each of the six ROI and combined them into a single image per ROI (Fig. 2a). By applying a threshold to the E-cadherin and vimentin channels, we obtained the masks for the epithelium and the lamina propria, respectively (Fig. 2b). We used these masks to assign cells to the two compartments and normalise marker expression values or positive pixel areas.

We then applied the pixel-based approach to measure the levels of immunoglobulin A (IgA), the major immunoglobulin isotype in intestinal mucosa[21]. This returned a quantification of both the IgA expressed by the plasma cells resident in the diffuse lymphoid tissue of the lamina propria as well as the secreted IgA undergoing

transcytosis to traverse the epithelial compartment (Fig. 2c). Normalised IgA$^+$ areas in the two compartments, as expected, confirmed higher IgA$^+$ levels in the lamina propria than in the epithelium in each sample (Fig. 2d).

Next, we quantified the IgA$^+$ plasma cells in the lamina propria using SIMPLI's cell-based approach. First, we performed single-cell segmentation and retained only cells overlapping with the lamina propria mask (Fig. 2e). We then identified IgA$^+$ plasma cells, T cells, macrophages, and dendritic cells resident in the lamina propria according to the highest overlap between the cell area and the mask of each immune cell population (Fig. 2f). Finally, we quantified the four immune cell populations across the six samples and observed that IgA$^+$ plasma cells constitute approximately 25% of all identified immune cells (Fig. 2g). This is consistent with previous quantifications of the fraction of plasma cells over the total mononucleated cells in the lamina propria of healthy individuals[22].

The relative proportion of IgA$^+$ plasma cells positively correlated with the normalised IgA$^+$ area in the lamina propria, demonstrating that the quantification from the single-cell analysis is supported by the cell agnostic measurements at the pixel-level (Fig. 2h).

Localisation of T follicular helper cells within a human germinal centre

As a second case study, we used SIMPLI to characterise the immune cell populations in a FFPE section of healthy human appendix (APP1, Supplementary Table 1). After staining the tissue section with the 26 antibodies used previously (Supplementary Table 2), we performed IMC and used SIMPLI to extract and normalise the single channel images from the raw IMC data for the ROI. The resulting

combined image revealed a germinal centre in the B cell area and follicle-associated epithelium forming the boundary with the appendiceal lumen (Fig.3a).

We performed single-cell segmentation and classified immune and epithelial cells based on the highest overlap with the corresponding masks (Fig. 3b). Using both phenotyping approaches available in SIMPLI, we further characterised the phenotypes of T cells, that overall constituted approximately 27% of all cells (Fig. 3c). First, we applied unsupervised clustering using seven markers of T cell function (Supplementary Table 2) obtaining five distinct cell clusters (Fig. 3d). Based on the marker expression profiles, we assigned cluster 1 to $CD4^+$ T cells, cluster 2 to $CD8^+CD45RO^+$ T cells, cluster 3 to $CD4^+CD45RA^+$ T cells, cluster 4 to $CD4^+CD45RO^+$ T cells and cluster 5 to $CD4^+PD1^+$ T cells (Fig. 3e). The latter likely represented a set of $PD1^+$ T follicular helper cells known to be located in the germinal centre[23].

We re-identified these $PD1^+$ T follicular helper cells with the second phenotyping approach implemented in SIMPLI. Starting from all T cells, we first extracted $CD4^+$ cells (≥0.1 CD4 expression, Fig. 3f) and, among those, we further identified $PD1^+$ cells (≥0.15 PD1 expression, Fig. 3g). The expression profile of the resulting $CD4^+PD1^+$ T cells (Fig 3h) closely recapitulated that of cluster 5 (Fig. 3e), indicating that cell phenotypes identified with unsupervised clustering can be confirmed through user-guided thresholding of marker expression.

Next, we investigated the spatial localisation of $PD1^+$ T follicular helper cells within the ROI by performing an analysis of homotypic aggregations. This allowed us to localise a single high-density cluster containing 84% of $CD4^+PD1^+$ T cells within the germinal centre (Fig. 3i). This distribution of $CD4^+PD1^+$ T cells was in accordance with the localisation of T helper cells in the in the follicles of secondary lymphoid organs[23] and was confirmed by the histological inspection of the tissue image (Fig. 3j).

Spatially resolved characterisation of cell-cell interactions in colorectal cancer

As a third case study, we applied SIMPLI to the analysis of mIF-derived images of a colorectal cancer sample (CRC1, Supplementary Table 1) stained with anti CD8, PD1, Ki67, PDL1, CD68, GzB and DAPI antibodies (Supplementary Table 2). We focused on a 5mm^2 ROI that was rich in T cells at the invasive margins of the tumour (Fig. 4a). This allowed us to characterise the cell-cell interactions between PDL1$^+$ cells and CD8$^+$PD1$^+$ T cells.

After image pre-processing and single-cell segmentation, we identified PDL1$^+$ cells and CD8$^+$PD1$^+$ by applying expert-defined thresholds to PDL1 (≥0.1), CD8 (≥0.1), and PD1 (≥0.005) expression levels, respectively. We extracted 2026 PDL1$^+$ cells (Fig. 4b) and 3177 CD8$^+$ cells, 94 of which also expressed PD1 (Fig. 4c). The two sets of PDL1$^+$ and CD8$^+$PD1$^+$ cells constituted 3.7% and 0.2% of all cells in the analysed region, respectively (Fig. 4d).

We characterised these PDL1$^+$ and CD8$^+$PD1$^+$ cells further, focusing on the ones in close proximity to each other. Using the Euclidean distances between their centroids, we identified 35 PDL1$^+$ cells and 21 CD8$^+$PD1$^+$ cells at a distance lower than 12μm apart, which corresponded to twice the maximum cell radius length. We considered these cells proximal enough to be possibly engaging in PD1-PDL1 mediated interactions. By comparing 21 CD8$^+$PD1$^+$ cells proximal to PDL1$^+$ cells to 73 CD8$^+$PD1$^+$ cells distal to PDL1$^+$ cells, we found no difference in the expression of cytotoxicity (GzB) or proliferation (ki67) markers (Fig. 4e). On the contrary 35 PDL1$^+$ cells proximal to CD8$^+$PD1$^+$ cells expressed higher levels of CD68 than 1991 PDL1$^+$ cells distal to CD8$^+$PD1$^+$ cells (Fig. 4f). By inspecting the imaged tissue at 40x

magnification, we confirmed that indeed $CD8^+PD1^+$ macrophages were in close proximity to $CD8^+PD1^+$ T cells (Fig.4g).

DISCUSSION

SIMPLI is an open-source, customisable and technology-independent tool for the analysis of multiplexed imaging data. It enables the processing of raw images, the extraction of cell data and the spatially resolved quantification of cell types or functional states as well as a cell-independent analysis of tissues at the pixel level, all within a single platform.

In comparison to currently available software, SIMPLI increases the portability, scalability and reproducibility of the analysis. Moreover, it can easily accommodate specific analytical requirements across a wide range of tissues, experimental conditions and imaging technologies through user-friendly configuration files. SIMPLI interoperates with multiple software and programming languages by leveraging workflow management and containerisation. This makes the inclusion of new algorithms, features and imaging data formats easy to implement, such as alternative methods of cell-segmentation, pixel and cell classification or a Graphical User Interface for interactive data visualisation

Multiplexed imaging methods have proven to be a powerful approach for the study of tissues through the in-depth characterisation of cell phenotypes and interactions. SIMPLI represents an effort to make these analyses more accessible to a wider community. This will enable exploitation of highly multiplexed imaging technologies for multiple applications, ranging from basic life science and pharmaceutical research to precision medical use in the clinics.

ONLINE METHODS

SIMPLI description and implementation

SIMPLI's workflow is divided into three steps (raw image processing; cell-based analysis; pixel-based analysis), which are in turn constituted of multiple tasks (Fig. 1 and Supplementary Fig. S1). Each of these can be selected for execution sequentially or independently from the command-line or through a configuration file, giving the user high flexibility.

Raw data from IMC experiments (.mcd or .txt files) are converted into single or multi-channel .tiff images with imctools[24]. Data from other multiplexed imaging platforms may be supplied directly as raw single or multi-channel tiff images (Supplementary Fig. 1a). Images are then normalised by rescaling all pixel values of each channel up to the 99th percentile of the distribution using the EBImage[25] package with custom R scripts. The resulting normalised images are processed with CellProfiler4[9] to generate thresholded images and masks that will be used as input for the two subsequent steps, the pixel-based and the cell-based analyses. These can be run as single workflows or in parallel within the same run. Both of them provide multiple outputs of the various tasks, including tabular text files, visualisation plots and various comparisons across samples (Supplementary Fig. 1).

The cell-based analysis is composed of cell data extraction, cell phenotyping and spatial analysis (Supplementary Fig. 1b). The extraction of cell data starts with single-cell segmentation using CellProfiler4[9] and returns (1) single-cell data consisting of the marker expression values and the coordinates of each cell in the ROI and (2) the ROI segmentation mask marking all the pixels belonging to each cell with its unique identifier. Cells can be assigned to tissue compartments or specific populations

according to their overlap with the previously defined masks and visualised in the ROI as outlines, while their proportions are quantified in barplots and boxplots.

All cells, or only those in specific tissue compartments or populations can be phenotyped using two approaches. The first consists of unsupervised clustering based on the marker expression values using Seurat[26]. Cells are represented as nodes in a k-nearest neighbour graph based on their Euclidean distances in a principal component analysis space. This graph is then partitioned into clusters using the Louvain algorithm[27] at user-defined levels of resolution leading to the unsupervised identification of cell phenotypes. Clusters of cell phenotypes are plotted as scatterplots in Uniform Manifold Approximation and Projection (UMAP)[28] space. The second phenotyping approach is based on user-defined thresholds of marker expression values that can be combined using logical operators for the identification of designated cell phenotypes. The distributions of cells are represented as density plots based on the marker expression levels. In both phenotyping approaches, the expression profiles of the cell types are plotted as heatmaps, their proportions quantified in barplots and boxplots and their locations in the ROI visualised as cell outlines.

Once cell populations and phenotypes have been identified, the spatial analysis investigates the distance between cells of the same (homotypic aggregations) or different (heterotypic aggregations) types. In the homotypic analysis, clusters of cells of the same type within a user-defined distance are identified with DBSCAN[3] as implemented in the fpc[29] R package. These homotypic cell aggregations are visualised as position maps, reporting cell location and high-density clusters in the ROI. In the heterotypic analysis, the cell distances, defined as the Euclidean distances between cell centroids, are computed using custom R scripts and visualised as density plots.

The homotypic and heterotypic spatial analyses can be run in parallel or singularly on one or more sets of cells.

The pixel-based analysis quantifies areas positive for user-defined combination of markers using the EBImage[25] package with custom R scripts (Supplementary Fig. 1c). These measurements are performed starting from the thresholded images produced in the raw image processing step (Supplementary Fig. 1a). The marker positive areas obtained in this way are then normalised over the area of the whole image, specific tissue compartments or cell populations. The resulting normalised positive areas can then be quantified in barplots and boxplots.

SIMPLI is implemented as a Nextflow[30] pipeline employing Singularity containers[31] hosted on Singularity Hub[32] to manage all the libraries and software tools. This allows SIMPLI to automatically manage all dependencies, irrespective of the running platform. Nextflow also manages automatic parallelisation of all tasks while still allowing the selection of parts of the analysis to execute.

Sample description

Six FFPE blocks of normal (non-cancerous) colon mucosa (CLN1-CLN6), one of normal appendix (APP1), and one of colorectal cancer (CRC1) were obtained from eight individuals who underwent surgery for the removal of colorectal cancers (Supplementary Table 1). All blocks were reviewed by an expert pathologist (MRJ). All patients provided written informed consent in accordance with approved institutional guidelines (University College London Hospital, REC Reference: 20/YH/0088; Istituto Clinico Humanitas, REC Reference: ICH-25-09).

Staining and IMC ablation of human colon mucosa and appendix

Four µm-thick sections were cut from each block of samples CLN1-CLN6 and APP1 with a microtome and used for staining with a panel of 26 antibodies targeting the main immune, stromal and epithelial cell populations of the gastrointestinal tract (Supplementary Table 2). The optimal dilution of each antibody in the panel was identified by staining and ablating FPPE appendix sections. The resulting images were reviewed by an mucosal immunologist (J.S.) and the dilution giving the best signal to background ratio was selected for each antibody (Supplementary Table 2). To perform the staining for IMC, slides were dewaxed after a one-hour incubation at 60°C, rehydrated and heat-induced antigen retrieval was performed with a pressure cooker in Antigen Retrieval Reagent-Basic (R&D Systems). Slides were incubated in a 10% BSA (Sigma), 0.1% Tween (Sigma), and 2% Kiovig (Shire Pharmaceuticals) Superblock Blocking Buffer (Thermo Fisher) blocking solution at room temperature for two hours. Each antibody was added to a primary antibody mix at the selected concentration in blocking solution and incubated overnight at 4°C. After two washes in PBS and PBS-0.1% Tween, the slides were treated with the DNA intercalator Cell-ID™ Intercalator-Ir (Fluidigm) (containing the two iridium isotopes 191Ir and 193Ir) 1.25 mM in a PBS solution. After a 30-minute incubation, the slides were washed once in PBS and once in MilliQ water and air-dried. The stained slides were then loaded in the Hyperion Imaging System (Fluidigm) imaging module to obtain light-contrast high resolution images of approximately four mm^2. These images were used to select the ROI in each slide. For CLN1-CLN6, 1 mm^2 ROIs were selected to contain the full thickness of the colon mucosa in a longitudinal orientation. For APP1, a one mm^2 ROI containing a lymphoid follicle in its whole depth alongside a portion of lamina propria and of epithelium was selected. ROIs were ablated at a o µm/pixel resolution and 200 Hz frequency.

IMC data analysis

Twenty-eight images from 26 antibodies and two DNA intercalators were obtained from the raw .txt files of the ablated regions in CLN1-CLN6 and APP1. Pixel intensities for each channel were normalised to the 99th percentile in all samples and background pixels were removed.

For CLN1-CLN6, the masks corresponding to the lamina propria were obtained using the Vimentin channel including all <75-pixel large negative areas. Epithelial masks were generated starting from the Pan-keratin and E-cadherin channels, dilatating the images with a three-pixel disk and the filling up of all <75-pixel large negative areas. These masks were then added into a sum image, which underwent dilatation with a three-pixel disk and filling up of all <25-pixel large negative areas. Positive features outside of the lamina and epithelium were removed with an opening operation using a 150-pixel radius and the lamina propria mask was subtracted from the sum image to generate the final mask for the epithelial compartment. These masks and the thresholded images were used as input for the pixel-based and cell-based analysis tasks.

Single-cell segmentation started from DNA1 segmentation with global thresholding to identify the cell nuclei. For CLN1-CLN6, cells were identified by radially expanding each nucleus for up to 10 pixels over a membrane mask derived from the IgA, CD3, CD68, CD11c and E-cadherin channels. Only cells overlapping with the lamina propria mask by at least 30% were retained. For APP1, cells were isolated through watershed segmentation with the nuclei as seeds on a membrane mask using the CD45, Pan-keratin and E-cadherin channels.

Cell identities were assigned according to the highest overlap of the cell area with marker-specific thresholds defined by an expert histologist (JS. For CLN1-CLN6: ≥15% of the IgA^+ mask for IgA^+ cells; ≥15% of the $CD3^+$ mask for T cells; ≥25% of the $CD68^+$ mask for macrophages; ≥15% of $CD11c^+$ mask for dendritic cells. For APP1: ≥10% of the $CD3^+$ mask for T cells; ≥10% of the sum of the $CD20^+$ and $CD27^+$ masks for B cells; ≥10% of the $CD68^+$ mask for macrophages; ≥10% of the $CD11c^+$ mask for dendritic cells; ≥10% of E-cadherin$^+$ and Pan-keratin$^+$ masks for epithelial cells). Cells that did not overlap with any marker were left as unassigned.

In APP1, T cells were further phenotyped using unsupervised clustering at resolutions 0.1, 0.15, 0.2, 0.25 and 0.3, based on the cell marker intensity for CD3, CD45RA, CD45RO, CD4, CD8, Ki67 and PD1. The resulting clusters were manually inspected and the clustering with the highest number of biologically meaningful clusters (resolution = 0.25) was chosen. $CD4^+PD1^+$ cells within the T cells were re-identified using mean intensity thresholds defined by an expert histologist (JS) of 0.1 for CD4 and 0.15 for PD1. The homotypic aggregations of $CD4^+PD1^+$ T cells were computed using a minimum of five points per cluster and a reachability parameter corresponding to a density of at least 5 cells/mm^2.

CD3 staining and mIF of human colorectal cancer

Two 4 µm thick serial sections were cut from CRC1 FFPE block using a microtome. The first slide was dewaxed and rehydrated before carrying out HIER with Antigen Retrieval Reagent-Basic (R&D Systems). The tissue was then blocked and incubated with the anti-CD3 antibody (Dako, Supplementary Table 2) followed by horseradish peroxidase (HRP) conjugated anti-rabbit antibody (Dako) and stained with 3,3' diaminobenzidine (DAB) substrate (Abcam) and haematoxylin. Areas with $CD3^+$

infiltration in the proximity of the tumour invasive margin were identified by a clinical pathologist (M. R-J.)

The second slide was stained with a panel of six antibodies (CD8, PD1, Ki67, PDL1, CD68, GzB, Supplementary Table 2), Opal fluorophores and 4',6-diamidino-2-phenylindole (DAPI) on a Ventana Discovery Ultra automated staining platform (Roche). Expected expression and cellular localisation of each marker as well as fluorophore brightness were used to minimise fluorescence spillage upon antibody-Opal pairing. Following a one-hour incubation at a 60°C, the slide was subjected to an automated staining protocol on an autostainer. The protocol involved deparaffinisation (EZ-Prep solution, Roche), HIER (DISC. CC1 solution, Roche) and seven sequential rounds of: one hour incubation with the primary antibody, 12 minutes incubation with the HRP-conjugated secondary antibody (DISC. Omnimap anti-Ms HRP RUO or DISC. Omnimap anti-Rb HRP RUO, Roche) and 16 minute incubation with the Opal reactive fluorophore (Akoya Biosciences). For the last round of staining, the slide was incubated with Opal TSA-DIG reagent (Akoya Biosciences) for 12 minutes followed by Opal 780 reactive fluorophore for our hour (Akoya Biosciences). A denaturation step (100°C for 8 minutes) was introduced between each staining round in order to remove the primary and secondary antibodies from the previous cycle without disrupting the fluorescent signal. The slide was counterstained with DAPI (Akoya Biosciences) and coverslipped using ProLong Gold antifade mounting media (Thermo Fisher Scientific). The Vectra Polaris automated quantitative pathology imaging system (Akoya Biosciences) was used to scan the labelled slide. Six fields of view, within the area selected by the pathologist, were scanned at 20x and 40x magnification using appropriate exposure times and loaded into inForm[33] for spectral unmixing and autofluorescence isolation using the spectral libraries.

mIF data analysis

After spectral unmixing and merging of six 20x fields of view, one single-tiff image was extracted for each marker and its intensity was rescaled from 0 to 1 with custom R scripts. The resulting single-tiff images were then used as input for SIMPLI. Images were pre-processed to remove the background and used for single-cell segmentation by applying a global thresholding to the DAPI channel and selecting all objects with a diameter between four and 60 pixels. $CD8^+PD1^+$ cells and $PDL1^+$ cells were identified using mean intensity thresholds of 0.01 for CD8, 0.005 for PD1 and 0.01 for PDL1. All thresholds were inspected by an expert histologist (JS).

The distribution of the minimum distances between $PDL1^+$ cells and $CD8^+PD1^+$ cells were calculated from the coordinates of the centroids of each cell in the image. All $PDL1^+$ cells and $CD8^+PD1^+$ cells at a distance from each other lower than double the maximum cell radius (24 pixel = 12 μm) were considered as proximal. All other cells were classified as distal.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, F.D.C., upon request.

CODE AVAILABILITY

SIMPLI's code, documentation and an example dataset are available at https://github.com/ciccalab/SIMPLI. The software code is protected by copyright. No permission is required from the rights-holder for non-commercial research uses.

Commercial use will require a license from the rights-holder. For further information contact translation@crick.ac.uk.

REFERENCES

1. Parra, E.R., Francisco-Cruz, A. & Wistuba, I.I. State-of-the-art of profiling immune contexture in the era of multiplexed staining and digital analysis to study paraffin tumor tissues. *Cancers* **11**, 247 (2019).
2. Giesen, C. *et al.* Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nature Methods* **11**, 417-422 (2014).
3. Angelo, M. *et al.* Multiplexed ion beam imaging of human breast tumors. *Nature medicine* **20**, 436-442 (2014).
4. Goltsev, Y. *et al.* Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed Imaging. *Cell* **174**, 968-981.e15 (2018).
5. Lin, J.R., Fallahi-Sichani, M., Chen, J.Y. & Sorger, P.K. Cyclic immunofluorescence (CycIF), a highly multiplexed method for single-cell imaging. *Current protocols in chemical biology* **8**, 251-264 (2016).
6. Jackson, H.W. *et al.* The single-cell pathology landscape of breast cancer. *Nature* **578**, 615-620 (2020).
7. Schürch, C.M. *et al.* Coordinated Cellular Neighborhoods Orchestrate Antitumoral Immunity at the Colorectal Cancer Invasive Front. *Cell* **182**, 1341-1359.e19 (2020).
8. Berg, S. *et al.* ilastik: interactive machine learning for (bio)image analysis. *Nature Methods* **16**, 1226-1232 (2019).
9. McQuin, C. *et al.* CellProfiler 3.0: Next-generation image processing for biology. *PLOS Biology* **16**, e2005970 (2018).
10. van Maldegem, F. *et al.* Characterisation of tumour immune microenvironment remodelling following oncogene inhibition in preclinical studies using an optimised imaging mass cytometry workflow. *bioRxiv*, 2021.02.02.429358 (2021).
11. Zanotelli, V.R.T. & Bodenmiller, B. ImcSegmentationPipeline: A pixelclassification based multiplexed image segmentation pipeline. *Zenodo* (2017).
12. Jones, T.R. *et al.* CellProfiler Analyst: data exploration and analysis software for complex image-based screens. *BMC Bioinformatics* **9**, 482 (2008).
13. Eling, N., Damond, N., Hoch, T. & Bodenmiller, B. cytomapper: an R/Bioconductor package for visualisation of highly multiplexed imaging data. *bioRxiv*, 2020.09.08.287516 (2020).
14. Timms, J.A. *et al.* Immunocluster: A Computational Tool to Explore the Immune Profile and Cellular Heterogeneity of Hematological Diseases Using Liquid and Imaging Mass, and Flow Cytometry Data. *Blood* **136**, 9-10 (2020).
15. Stoltzfus, C.R. *et al.* CytoMAP: A Spatial Analysis Toolbox Reveals Features of Myeloid Cell Organization in Lymphoid Tissues. *Cell Reports* **31**, 107523 (2020).
16. Somarakis, A., Unen, V.V., Koning, F., Lelieveldt, B. & Höllt, T. ImaCytE: Visual Exploration of Cellular Micro-Environments for Imaging Mass Cytometry Data. *IEEE Transactions on Visualization and Computer Graphics* **27**, 98-110 (2021).
17. Yang, T. *et al.* SPIAT: An R package for the Spatial Image Analysis of Cells in Tissues. *bioRxiv*, 2020.05.28.122614 (2020).
18. neighbouRhood. https://github.com/BodenmillerGroup/neighbouRhood (2019).
19. Catena, R., Montuenga, L.M. & Bodenmiller, B. Ruthenium counterstaining for imaging mass cytometry. *The Journal of pathology* **244**, 479-484 (2018).

20. Bankhead, P. *et al.* QuPath: Open source software for digital pathology image analysis. *Scientific Reports* **7**, 16878 (2017).
21. Pabst, O. & Slack, E. IgA and the intestinal microbiota: the importance of being specific. *Mucosal Immunology* **13**, 12-21 (2020).
22. Dorn, I., Schlenke, P., Mascher, B., Stange, E.F. & Seyfarth, M. Lamina Propria Plasma Cells in Inflammatory Bowel Disease: Intracellular Detection of Immunoglobulins Using Flow Cytometry. *Immunobiology* **206**, 546-557 (2002).
23. Song, W. & Craft, J. T follicular helper cell heterogeneity: Time, space, and function. *Immunological Reviews* **288**, 85-96 (2019).
24. imctools. https://github.com/BodenmillerGroup/imctools (2017).
25. Pau, G., Fuchs, F., Sklyar, O., Boutros, M. & Huber, W. EBImage—an R package for image processing with applications to cellular phenotypes. *Bioinformatics* **26**, 979-981 (2010).
26. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature Biotechnology* **36**, 411-420 (2018).
27. Blondel, V.D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment* **2008**, P10008 (2008).
28. Leland, M., John, H., Nathaniel, S. & Lukas, G. UMAP: uniform manifold approximation and projection. *Journal of Open Source Software* **3**, 861 (2018).
29. Henning, C. fpc. https://cran.r-project.org/web/packages/fpc/index.html (2020).
30. Di Tommaso, P. *et al.* Nextflow enables reproducible computational workflows. *Nature Biotechnology* **35**, 316-319 (2017).
31. Kurtzer, G.M., Sochat, V. & Bauer, M.W. Singularity: Scientific containers for mobility of compute. *PLOS ONE* **12**, e0177459 (2017).
32. Sochat, V.V., Prybol, C.J. & Kurtzer, G.M. Enhancing reproducibility in scientific computing: Metrics and registry for Singularity containers. *PLOS ONE* **12**, e0188511 (2017).
33. Kramer, A.S. *et al.* InForm software: a semi-automated research tool to identify presumptive human hepatic progenitor cells, and other histological features of pathological significance. *Scientific Reports* **8**, 3418 (2018).

ACKNOWLEDGMENTS

We thank Sharavan Vishaan Venkateswaran for testing SIMPLI. F.C is supported by Cancer Research UK [C43634/A25487], the Cancer Research UK King's Health Partners Centre at King's College London [C604/A25135], the Cancer Research UK City of London Centre [C7893/A26233], innovation programme under the Marie Skłodowska-Curie grant agreement No CONTRA-766030 and the Francis Crick Institute, which receives its core funding from Cancer Research UK (FC001002), the UK Medical Research Council (FC001002), and the Wellcome Trust (FC001002). For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

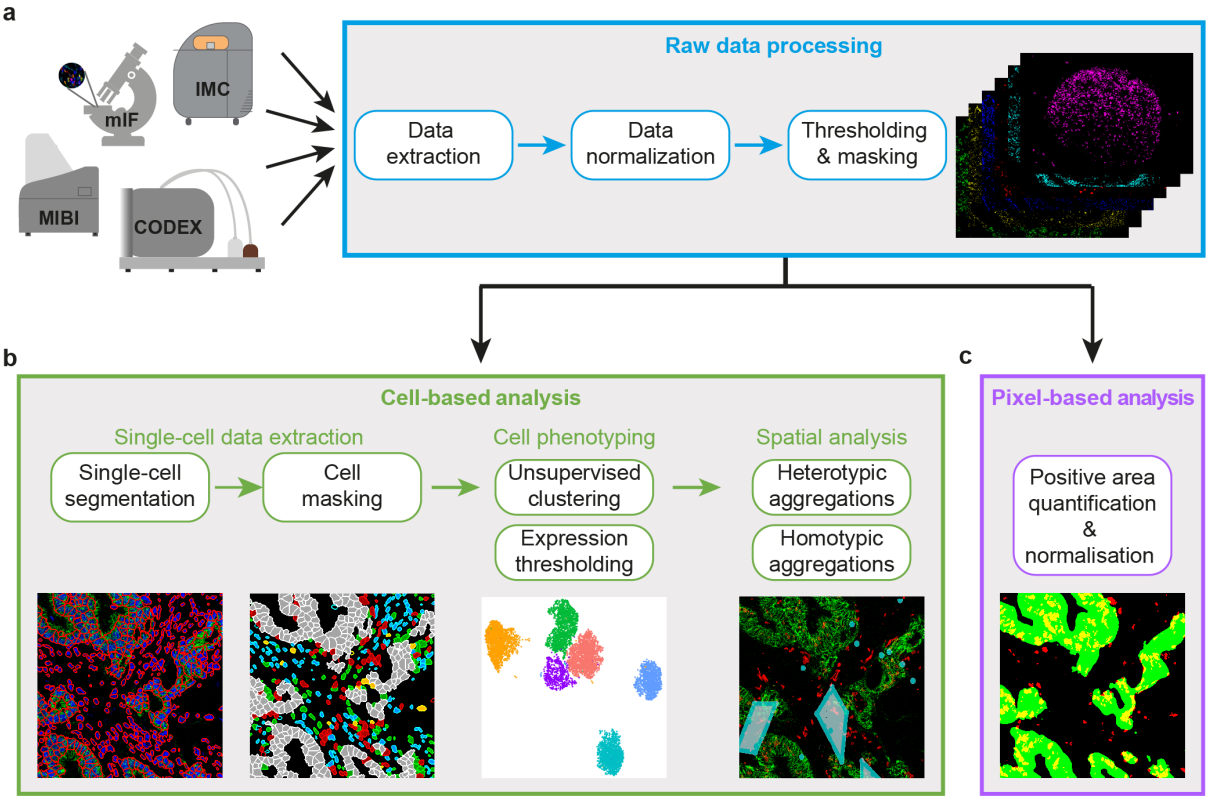
AUTHOR CONTRIBUTIONS

F.D.C. conceived and directed the study with support of J.S. M.B. developed the software with the help of D.T. and M.R.K. L.M. and A.A.S performed the experiments. M.B, L.M, A.A.S., M.J.P., J.S., and F.D.C. analysed the data. M.R.J, G.B. and L.L. identified the samples and provided clinical assessments. M.R.J. performed pathological assessments. M.B. and F.D.C. wrote the manuscript with contribution from A.A.S., L.M. and M.J.P. All authors approved the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

Figure 1. Schematics of the SIMPLI workflow

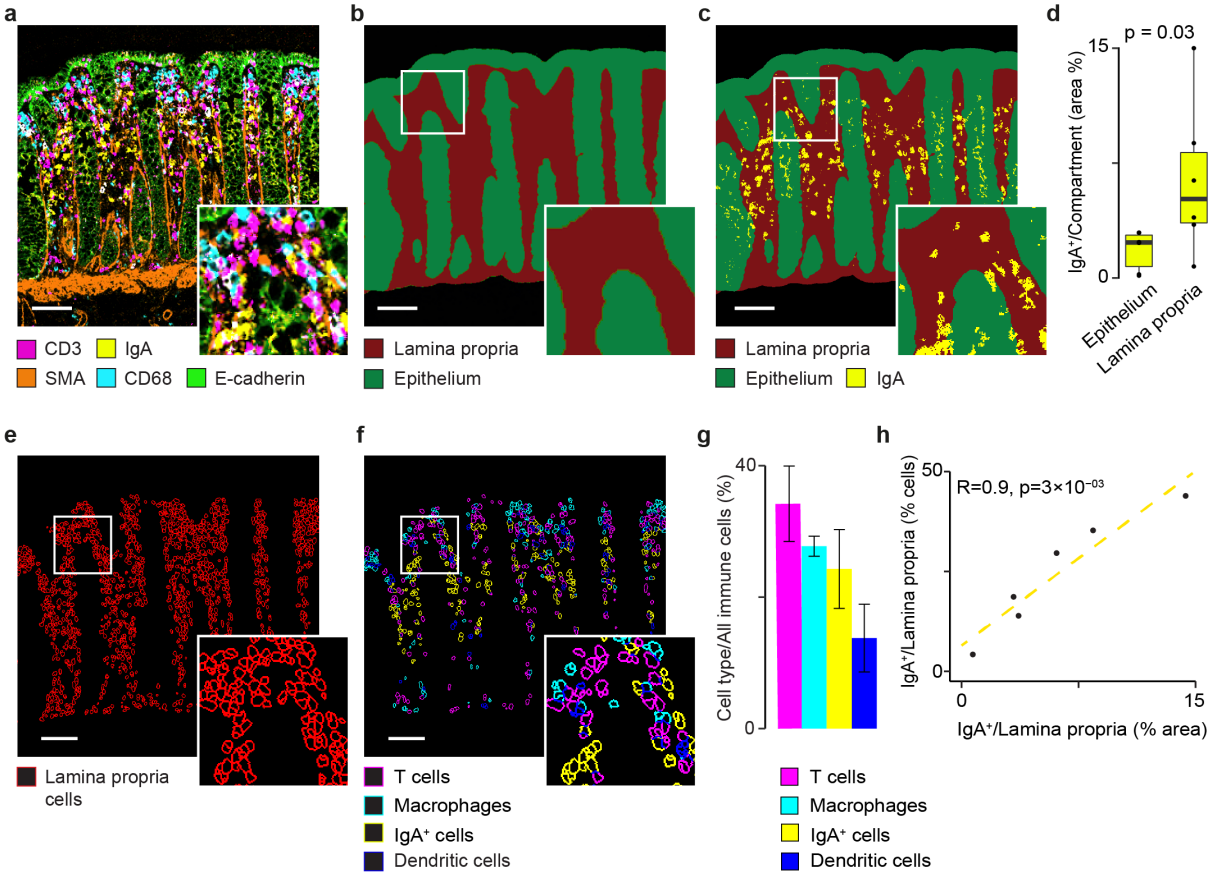


a. Raw data from multiplexed imaging experiments are extracted, normalised and thresholded to remove the background noise and produce tissue compartment or cell population masks. The resulting images can be analysed using a cell-based (**b**) or a pixel-based (**c**) approach.

b. In the cell-based analysis, single cells are segmented and phenotyped using either unsupervised or supervised approaches based on marker expression. The distribution of cells in the tissue can then be investigated through a spatial analysis of homotypic or heterotypic aggregations.

c. In the pixel-based approach, areas positive for user-defined combination of markers are measured and normalised over the area of the whole image or of the masks defining compartments or cell populations.

Figure 2. IgA quantification in human colon mucosa



a. IMC image of a representative sample (CLN6) of normal colon mucosa after extraction and normalisation of raw data.

b. Masks defining the lamina propria and the epithelial compartments obtained by thresholding the vimentin and E-cadherin channels, respectively.

c. Representation of IgA^+ area overlaying the masks of the lamina propria and epithelium compartments.

d. Comparison of normalised IgA^+ areas in the lamina propria and epithelial compartments in CLN1-CLN6. Normalised areas were measured as the proportion of IgA^+ area over the lamina propria and epithelium masks, respectively. Distributions were compared using a two-sided Wilcoxon test.

e. Outlines of the cells in the lamina propria. After single-cell segmentation, all cells overlapping with the lamina propria mask by at least 30% of their area were considered as cells resident in the lamina propria.

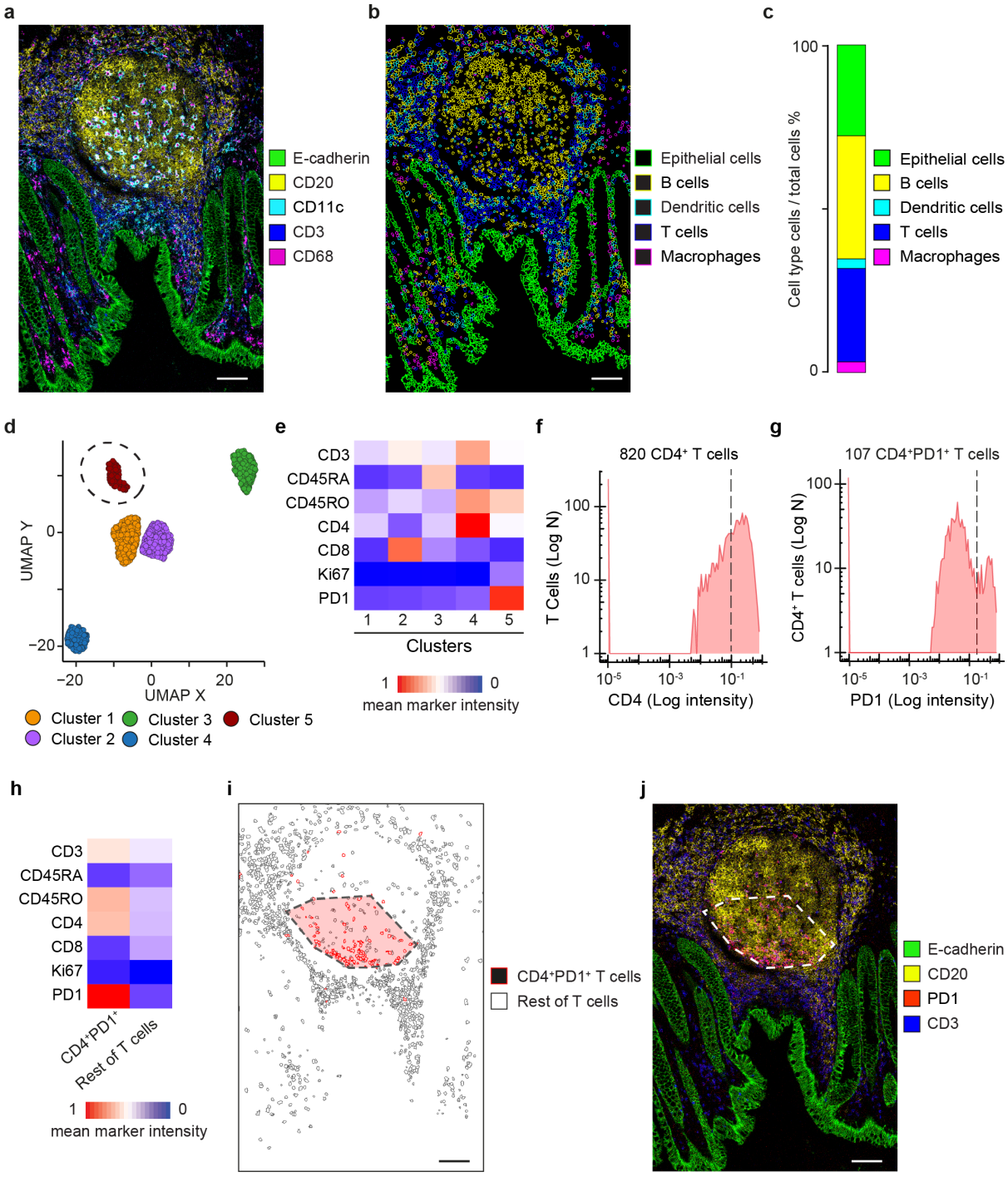
f. Outlines of immune cells resident in the lamina propria identified according to the highest overlap between their area and the masks for IgA^+ cells, T cells, macrophages and dendritic cells.

g. Proportions of T cells, IgA^+ cells, macrophages and dendritic cells over all immune cells in the lamina propria across CLN1-CLN6. Bars represent the standard error.

h. Correlation between normalised IgA^+ area and the proportion of IgA+ cells over the total immune cells in in the lamina propria in CLN1-CLN6. Pearson correlation coefficient R and associated p-value are shown.

Scale bar in all images = 100μm.

Figure 3. Single-cell characterisation of T cells in germinal centres



a. IMC image of a normal appendix (APP1) showing a central germinal centre with the columnar epithelium delimiting the appendiceal lumen.

b. Outlines of T cells, B cells, macrophages, dendritic and epithelial cells identified through the highest overlap with the respective masks.

c. Proportions of T cells, B cells, macrophages, dendritic and epithelial cells over all cells.

d. UMAP plot of 1,466 T cells grouped in five clusters by unsupervised clustering according to the expression of seven markers of T cell function (Supplementary Table 2). Cluster 5 (circled) corresponds to $CD4^+PD1^+$ T cells.

e. Expression profiles of five clusters. The mean expression value of each marker across all cells is reported. The colour scale was normalised independently.

f. Density plot of CD4 expression in T Cells. Cells with ≥0.1 CD4 expression were considered as $CD4^+$ T cells. This threshold was identified through histological inspection of the CD4 channel images.

g. Density plot of PD1 expression in $CD4^+$ T cells. Cells with ≥0.15 $PD1^+$ expression were considered as $CD4^+PD1^+$ T cells. This threshold was identified through histological inspection of the PD1 channel images.

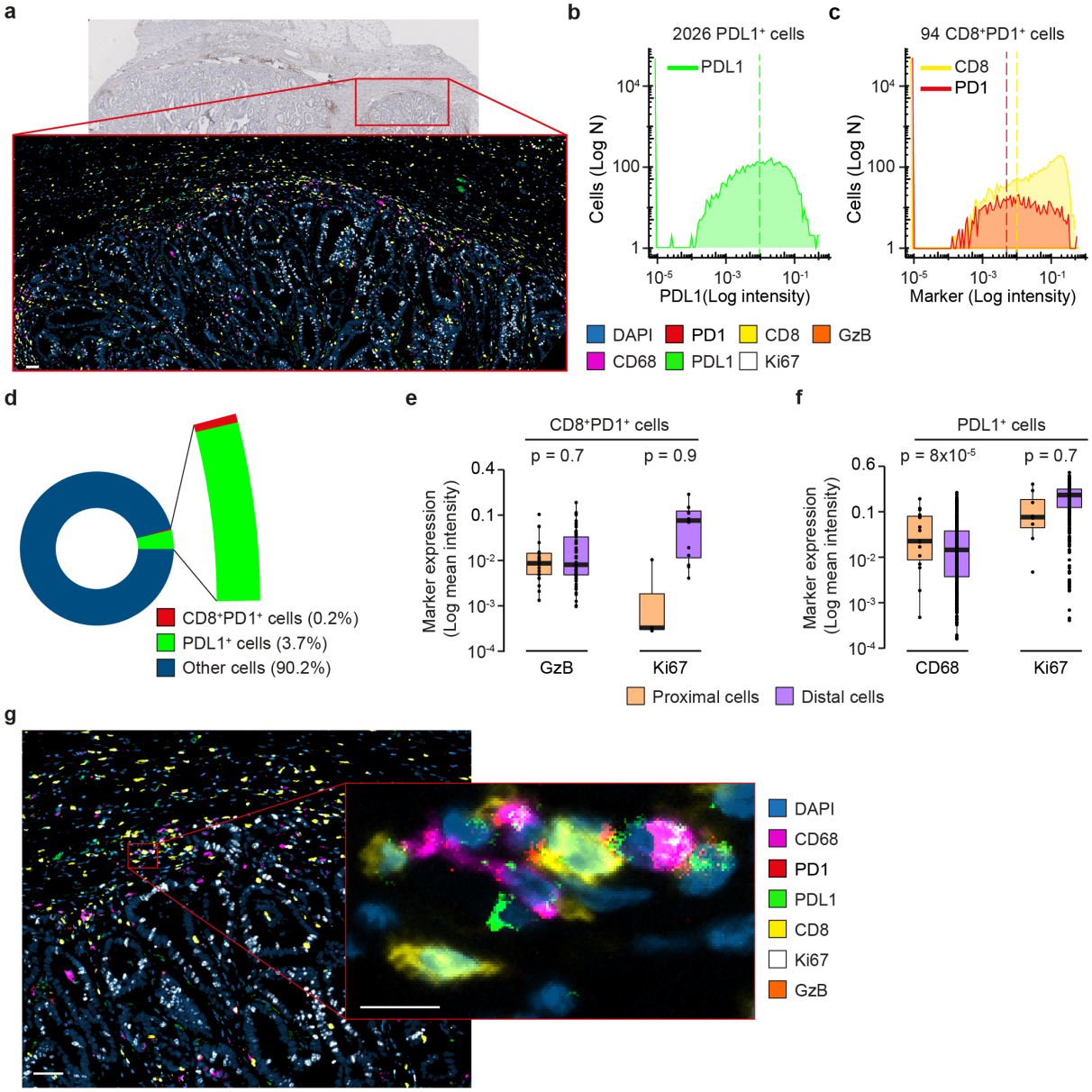
h. Expression profiles of the $CD4^+PD1^+$ T cells and rest of T cells. For both populations, the mean value of each marker across all cells is shown. The colour scale was normalised independently.

i. Position map of T cells within the ROI. The area of a high-density cluster of ≥5 $CD4^+PD1^+$ T cells per 10,000μm^2 (Methods) is highlighted in red.

i. IMC image showing the localisation of the PD1 signal within the ROI.

Scale bar for all images = 100µm.

Figure 4. Characterisation of PDL1⁺ and PD1⁺ T cells in colorectal cancers



a. CD3 immunohistochemistry (main image) and sequential mIF (zoom-in, 20x magnification) of a colorectal cancer sample (CRC1). The mIF image corresponded to a 5mm^2 tissue area at the invasive margins of the tumour and was obtained by combining the pre-processed images of seven markers. Scale bar = 50μm.

b. Density plot of PDL1 expression in CD8$^-$ cells. Cells with ≥0.01 PDL1 expression were considered as PDL1$^+$ cells. This threshold was identified through histological inspection of PDL1 channel images.

c. Density plots of CD8 and PD1 expression in T cells. Cells with ≥0.01 CD8 expression and ≥0.005 PD1 expression were considered as $CD8^+PD1^+$ T cells. These thresholds were identified through histological inspection of CD8 and PD1 channel images.

d. Proportions of $CD8^+PD1^+$ cells, $CD8^+PD1^-$ T cells and $PDL1^+$ cells over total cells.

e. Comparison of the mean intensity of GzB and Ki67 between $CD8^+PD1^+$ T cells proximal and distal to $PDL1^+$cells. Proximal $CD8^+PD1^+$ T cells were defined as those at less than 12μm from a $PDL1^+$ cell.

f. Comparison of the mean intensity of CD68 and ki67 between $PDL1^+$cells proximal and distal to $CD8^+PD1^+$ T cells. Proximal $PDL1^+$ cells were defined as those at less than 12μm from a $CD8^+PD1^+$ T cell.

Distributions in (**e**) and (**f**) were compared using a two-sided Wilcoxon test.

g. High resolution mIF image showing a $CD68^+PDL1^+$ cell in close proximity to a $CD8^+PD1^+$ T cell. Image was scanned at 40x magnification. Scale bar = 10μm.