Hunter et al, Supplemental Notes

S2: Expanded Methods

S2.1 Tissue Section Preparation and Regions of Interest Selection

Formalin-fixed paraffin-embedded 2mm human tonsil tissue cores were obtained from the Novopath Tissue Biobank (Royal Victoria Infirmary, Newcastle upon Tyne) and embedded into a 3 core Tissue Microarray (TMA). TMA blocks were constructed manually using PFM Medical biopsy punches. Cores were selected using haematoxylin and eosin-stained slides to guide suitable areas in the donor blocks. Cores were places in a paraffin embedding mould, heated to 65°C and embedded in molten wax before cooling to set. 8µm serial sections were using HM 325 Rotary Microtome (Fisher Scientific, USA) and mounted onto SuperFrost Plus[™] Adhesion slides (Epredia, CAT#10149870).

S2.2 Panel Design and Antibody Validation

A 27-plex antibody panel was designed to identify the immune, signalling and stromal components in the surrounding microenvironment of the human tonsil (See Figure S1 and Table S1). CD3 was purchased pre-conjugated to the assigned metal (Table S1). For all other targets of interest, carrier free antibody clones were sourced, and validated using standard single marker immunofluorescence on appropriate positive control tissues. Following verification of staining pattern and performance quality, approved antibodies were subject to lanthanide metal conjugation using Maxpar X8 metal conjugated to platinum isotopes 194 Pt and 198 Pt were conjugated as described in Mei et al. 2015 (1). (AbC[™] Total Antibody Compensation Beads (Thermo Fisher, USA, CAT#A10513) were stained with conjugated antibodies and analysed in suspension using the Helios[™] mass cytometer (Standard Biotools, USA) and successful conjugation was determined via detection of appropriate signal in the correct channel. Positive control tissues were then stained with conjugated antibodies. Suitable

antibody titres were evaluated using Imaging Mass Cytometry[™] via the Hyperion[™] Imaging System (Standard Biotools, USA).

S2.3 IMC Staining

Tonsil sections were baked for 1 hour at 60°C in a dry oven followed by dewaxing in two changes of fresh xylene (Sigma-Aldrich, USA,CAT# 534056-4L), 5 minutes each. Slides were rehydrated in decreasing grades of ethanol (100% 90% 70% 50%), for 5 mins each, and then washed twice for 5 mins in Milli-Q[®] Type 1 ultrapure water. Antigen retrieval was carried out using Tris-EDTA (pH9) buffer 0.5% Tween in the microwave at maximum power for 20mins. Slides were cooled to 70 °C, washed in two changes of Milli-Q[®] water 5 mins each and followed by two washes in PBS (Fisher Scientific, USA, CAT# 10728775). Tissue sections were blocked with 3% BSA (Sigma-Aldrich, CAT# A3059) in PBS for 45 mins at room temperature and then incubated overnight at 4°C with a cocktail of the 28 metal-conjugated antibodies (Table S1). Slides were washed with 0.2% Triton X-100 (Thermofisher, USA, CAT#85111) for 8 mins followed by two 8 min washes in PBS. Cell nuclei were stained with Cell-ID[™] Intercalator-Ir (Standard Biotools, USA, CAT#201192A) at a 1 in 400 dilution, for 30 mins at room temperature. The slides were then washed for 5mins in Milli-Q[®] water and air dried prior to ablation.

S2.4 Hyperion set up, quality control (QC) and data acquisition

Prior to each sample acquisition, the Hyperion Tissue Imager was calibrated and rigorously quality controlled to achieve reproducible sensitivity based on the detection of 175Lutetium. Briefly, a stable plasma was allowed to develop prior to ablation of a single multi-element-coated "tuning slide" (Standard Biotools). During this ablation, performance was standardised to an acceptable range by optimising system parameters using the manufacture's "auto tune" application or by manual optimisation of XY settings whilst monitoring 175Lutetium dual counts. After system tuning, Tonsil sections were loaded onto the Hyperion system in order to create Epi-fluorescence panorama images of the entire tissue surface to guide region of interest selection (ROI). Two ROIs of approx. 500µm² encompassing lymphoid follicles and surrounding structural cells were selected for ablation per batch

run. Small regions of tonsil tissue were first targeted to ensure complete ablation of tissue during the laser shot with ablation energies adjusted to achieve this where required. Finally, ablations were performed at 200Hz laser frequency to create a resultant MCD file containing all data from ROIs. Correction of signal 'spillover' between isotopes was performed as per the protocol described in here https://bodenmillergroup.github.io/IMCDataAnalysis/spillover-correction.html without deviation (2).

S2.5 Basic image QC and segmentation

S2.5.1 Ilastik Model Training

Ilastik (3) is a "random forest" machine learning tool used for classification of pixels or objects from images. It relies on providing training images, partially labelled by the user, in order for the machine learning model to determine what features within the image are being identified by the human user. For the two llastik models trialled in this manuscript, pixels were classified into either nuclear or background. This simplifies the model learning and application processes. The models were both trained on nuclear Iridium 193 datasets, however one was trained on images acquired of tonsil tissue data, while the other was trained on images of "Vero" cells. This comparison served to determine if Ilastik model training material matching the input material for analysis was important for correct macro functioning. Partial labelling was applied to the training images followed by training of the Ilastik model. Following training, the models were applied to the training images and checked visually for regions where the model did not perform as expected. In these regions, additional labelling is then placed, and the model is retrained. Once output probability map images match the expected results sufficiently, the model is saved and applied to any further datasets of unseen data. In the experimental dataset whereby no Ilastik model was applied to the nuclear channel Iridium 193, the raw unprocessed channel was used for the following CellProfiler steps instead.

S2.5.2 CellProfiler Cell Segmentation

CellProfiler (4,5) is an image processing and analysis pipeline tool. In this research, we have developed a CellProfiler pipeline to take input individual channels from the MATLAB processing pipeline, along with Ilastik processed nuclear channels, to output single-cell quantitative data. The CellProfiler pipeline requires the input data folders to be specified within the program, along with any folder naming conventions to be explicitly listed. The intensities of the membrane channel, the nuclear input channel, and the nuclear probability channel are first rescaled to allow for comparable intensity levels between images, regardless of any batch effect that might be present. The nuclear probability maps are then segmented into binary objects. This is achieved through a global "otsu thresholding", splitting the pixels into three classes (combining the middle intensity pixels with the background pixels), leaving just the nuclear pixels as white on a black background. A threshold smoothing scale of 1 and correction factor of 1 is also applied, with shape then used to distinguish between clumped objects. The pipeline was set to automatically calculate the size of the smoothing filter and the minimum distance between maxima for splitting of clumped objects. To facilitate the faster processing of the large datasets provided by multiplexed images, the pipeline was allowed to use lower resolution images to identify the local maxima. The membrane channel is then smoothed (keeping edges) to facilitate the following step of identifying the cell boundaries. A propagation method was used, using the identified nuclear objects as initial seeds, along with global otsu thresholding (classifying middle intensity pixels to foreground). A regularisation factor of 0.05 is applied, meaning the boundary is determined by the intensity gradient in the membrane channel. Holes are filled within each object. The cell boundaries are then expanded by 2 pixels. Visualisation images are generated and saved for checking segmentation accuracy in quick sense-checks. The cell objects are then used to measure the intensity of each of the channels in their area, outputting intensity information along with some morphological data regarding the cells, all on a single-cell basis in .npy file format.

S2.5.3 ImageJ Macro Channel Combination

A Fiji macro was created to potentially aid with cell boundary segmentation by creating a single combined channel to potentially better represent the whole cell membrane for use in CellProfiler. The raw Hyperion images were difficult to interpret visually as a whole due to the quantity of channels (28 metal signals/channels). Channels containing membrane or cytosol markers could be looked at individually to determine overall cell outline, however in isolation many channels had relatively low signal-to-noise giving a poorly defined, speckled appearance, while the specificity of other markers gave a clear indication of cell boundaries of some cell types but not others, meaning the single channel option was not optimal as a visual reference.

The images were processed with a Fiji macro which combined multiple non-nuclear channels into a single reference channel with greater signal-to-noise than its individual parts for a clearer illustration of cell morphology. Following visual interrogation, 14 non-nuclear channels with the best signal-to-noise ratio were selected. These were: channel 14 (CD45RA In115), channel 15 (CD68 Nd142), channel 17 (Ki67 Nd144), channel 19 (CD138 Nd146), channel 20 (CD163 Sm147), channel 23 (MPO Nd150), channel 30 (CD206 Gd157), channel 31 (CD79a Gd158), channel 36 (IFITM3 Dy163), channel 38 (CD57 Ho165), channel 43 (CD3 Er170), channel 45 (CD31 Yb172), channel 47 (CD4 Yb174) and channel 48 (HLADR Lu175). To process the images, the contrast of each channel was increased individually using "Enhance Contrast" with a default saturation of 0.35%. The look up table (LUT) was then applied. "Despeckle" was run on all of the channels to smooth noise. The channels were then stacked and projected into separate maximum intensity projections (MIP). The following outputs from each Hyperion image stack were produced: a composite image showing nuclear (green) and non-nuclear (red) regions for visual validation of segmentation, and a single channel non-nuclear MIP for use in CellProfiler for segmentation.

S2.5.4 Spillover corrections, transformation, normalisation, metadata addition and .FCS file creation.

Output npy files are subsequently processed within the MATLAB script. This portion of the script has three main functions: 1) Correct the data for spillover between the channels; 2) scale the data according to an *arcsinh* transformation; and 3) normalise the data to correct for batch variation (Z correction). Following these steps, the files are converted into FCS format to facilitate use with FCS Express for cluster analysis.

Due to the nature of multiplexed imaging, a degree of spillover between similarly labelled compounds is to be expected. As the Hyperion system, used for this analysis, relies on the distinction between very similar weight metal ions (even different isotopes of the same metal), the spillover could negatively impact interpretations, especially in a variation-rich field such as biology. To remove the contribution of metal isotopes in non-target channels we followed the protocol as outlined by Chevalier et al (2) whereby a slide was made that had each of the 28 metals used in this study spotted individually. The Hyperion system was then set to ablate each spot one at a time and measure across the entire detector range the proportion (%) of the target channel that appeared in non-target channels in order to create a spillover correction matrix that was used to subsequently correct the mean intensity values for each channel derived from our multi-stained tonsil tissue.

A Hyperbolic arcsine (*arcsinh*) scaling factor (6) is applied to the data with the intention of ensuring the high variance of a potentially larger value (positive signal) is given the same significance and weighting as a negative (low) signal that has low variance. This scale normalisation is essential for dimensionality reduction and clustering algorithms to work well.

Z-score normalisation was then applied to the spillover-corrected and *arcsinh*-processed data. The goal of this step is to align the median around 0 and compresses the scale to minimise influence of comparative brightness from batch effects and also inter/intra-marker variance due to biology and batch effects respectively.

S2.6 Data analysis and exploration using FSC Express software

S2.6.1 Introduction

The "OPTIMAL" analysis approach is designed to work with any software, script or package that can read FCS file format data. We have decided to use FCS Express from DeNovo by Dotmatics as it is a widely available software solution for cytometry data analysis, however the "OPTIMAL" approach could be executed using other commercial software as well as open source solutions. Successful implementation of this approach using FCS Express requires some basic and advanced knowledge of the software. We would direct anyone to look at the extensive YouTube tutorials that can be found here <u>https://denovosoftware.com/videos/</u>. However as much as is practical, we will outline the key steps in this supplemental methods section. Note that this is not designed to serve as a user guide for the software, which already exists https://denovosoftware.com/full-access/manual-tutorials/. The result of our segmentation approach is the generation of .FCS files that are composed of a combination of morphometric features (area, eccentricity), spatial features (x and y centroids) sample metadata (batch number, patient number) and metal "intensity" features (mean pixel intensity) that have been "compensated" (corrected) for signal spillover as described previously. Each metal intensity parameter exist in two formats, one set of metal intensity features have been arcsinh transformed using the pre-determined optimal cofactor for IMC data of 1, and the second set have been Z-scored transformed in addition to the arcsinh transformation in order to minimise the influence of technical variation (batch effect) and also natural variation in the intensity of various markers/antibodies that would dominate the data structure (DimRedux visualisation, clustering or heat maps).

S2.6.2 Determining the optimal transformation cofactor for IMC data

In order to determine the optimal *arcsinh* transformation co-factor value for IMC data we performed a "titration" of possible values from 120 to 0.1 using the original (pixel capped and compensated) metal signal parameters and using FCS Express "mathematical transformations" option in FCS Express "Pipelines". We then used the Fisher Discrimination Ratio (Rd for short), also known as "Linear Discriminate Analysis" (LDA) to enumerate the separation and resolution between a gated low (or negative) and high (or positive) signal using simple gating in FSC Express. The Rd is calculated as follows (7):

$$Rd = \frac{(Median \ hi - Median \ lo)}{(rSD \ hi + rSD \ low)}$$

It is then possible to plot for any parameter the *arcsinh* transformation value (x axis) versus the Rd value (y axis) to determine the cofactor that provides the best resolution of low and high signals and thus maximise the data resolution and structure prior to any dimensionality and clustering. However we strongly suggest that the use of these transformations be empirically tested using the OPTIMAL approach.

S2.6.3 Dimensionality reduction, batch effect determination and correction

The fastest and simplest way to visualise multi-dimensional single cell data is usually via some kind of dimensionality reduction approach whereby *n* dimensions are compressed and represented by 2 or sometimes 3 parameters (8). There are numerous options to achieve this, all with various benefits and drawbacks as well as tuneable hyper-parameters that will influence the output. In all cases the choice and implementation of the dimensionality reduction method must remain true to the original underlying data structure. Via the use of FCS Express and the "pipelines" tool we were able to assess multiple dimensionality reduction approaches using both embedded options within the software (tSNE and UMAP) as well as a number of external algorithms (TriMAP, PacMAP and fltSNE) via the python script interface in the FCS Express pipelines module. References for each of these algorithms can be found in the main manuscript. Information and instructions on how to run python scripts as part of the FCS Express pipelines module as well as the necessary software links and scripts that need to be installed with the relevant publications can be found here https://denovosoftware.com/full-access/knowledge-base/python-transformation/.

To begin the analysis open a new incidence of FCS Express and navigate to the "data" menu on the top options ribbon and select the "data list" option. At this point it is possible to select the "+" icon and choose which files to load. For this stage it is important to select the "merge FCS files" option as this will bring all samples in as a single merged entity and allow any analysis to be carried out simultaneously while still preserving individual file identity (a new column will be created called "file identifier" that can be used to de-convolute to individual sample level as necessary). At this stage, it is good practice to open up a plot by dragging the merged file on to the workspace. We tend to select a density plot of Area on the X axis versus eccentricity on the Y axis. By then "right clicking" on this plot and selecting "gate stats", this will show the total number of events loaded from all the merged files and can be cross checked against the expected total number of events. If these numbers are in agreement, one can proceed to set up and run a set of dimensionality reduction and visualisation methods to rapidly assess a range of key issues such as the presence of any batch effects, attempts to correct these as well as selecting the best overall method for presenting the data structure. Navigate to the "tools" option on the top options ribbon and select "transformations". This will open up a new window and clicking on the "+" icon allows you to then select the "pipelines" option in the drop down menu. Details on how to set up and run pipelines in FCS Express can be found here https://denovosoftware.com/flow-cytometry-pipelines/. Briefly selecting the pipeline option will open up a further window where it should be possible to input which parameters you want to have available for the further analysis. We recommend selecting all available parameters in the data set at this stage. There is also an option to "automatically run pipeline" that is ticked by default. This should be unticked so as to avoid the software attempting to run a pipeline before it has been finalised. Creating a new pipeline will activate a second "+" icon to the right of the first one. Selecting this will provide a new drop down menu where the required pre-set analysis options will be present as well as the Python terminal for running scripts that are external to the software (see later section). For the analysis of possible batch effect and to determine whether our chosen method of Z-

score normalisation was able to remove it, we used UMAP. The Z-score normalisation equation is shown below:

$$z \ score = \frac{x - \mu}{\sigma}$$

This was done by selecting "dimensionality reduction" then "UMAP" from the drop down menu. This creates a new step in the pipeline that can be named (for example) "UMAP_arcsinh_only" and select in the option window to only use the arcsinh c.f. 1 transformed parameters that we know have successfully stained our tonsil tissue from our basic image QC step (2.3 and table S1). We used the default hyper-parameter settings for UMAP (number of neighbours = 15, min. Low Dim Distance = 0.1 and iterations = 500) we did however add a parameter suffix of "archsinh only" to be able to identify each set of UMAP parameters in the same pipeline. We created a duplicate of this UMAP step in the same pipeline and renamed it "UMAP_arcsinh_Zscore" and selected the arcsinh c.f.1 and Z-score corrected versions of the same parameters as before. We kept the hyper-parameters of the UMAP identical but changed the parameter suffix to "arcsinh_Zscore" to provide distinction between these UMAP parameters and the other set of UMAP parameters. Once all steps were cross checked, the pipeline was executed by re-ticking the "automatically run pipeline" option and clicking on the icon to "apply to all plots" located to the right of the "+" icon. Successful launch of the pipeline can be determined by the progress bar at the bottom of the FCS Express workspace. Run time is dependent on numerous factors such as computer specifications, number of events/samples as well as the complexity of the pipeline, both in terms of steps and the algorithm itself. To give some context, generating two sets of UMAP parameters from the ~100,000 single cells across 24 samples took approximately 10 minutes on a Microsoft Surface Pro Series 4 with an 4 Intel(R) Core(TM) i7-7660U CPU @ 2.50GHz 2.50 GHz, 8 GB RAM and 64-bit operating system with x64-based processor. We were then able to create a colour dot plot for the arcsinh only UMAP parameters 1 and 2, as well as

one for the *arcsinh* and Z-score normalised set. It was then possible to plot batch "batch number" versus Iridium signal (Z-score version) as a density plot and set individual gates for each of the 12 data batches with suitable colours. Displaying the batch gates on each UMAP plot allowed us to quickly assess if there was any batch effect (*arcsinh* only UMAP parameters) and if it had been corrected/minimised (*arcsinh* plus Z-score normalised UMAP parameters and *arcsinh* plus 0-1 normalised parameters).

S2.6.4 PacMap visualisation via Python script interface

Having shown empirically that Z-score normalisation of the arcsinh transformed parameters eliminated batch effects in the data, we used these for subsequent clustering and optimal visualisations. UMAP visualisation provided a good two-dimensional representation of the 27 antibody parameter data set, however we wanted to assess if there was an option that would improve upon this further. Our criteria for assessing these methods was based on the overall data structure projection; namely the presence or absence of defined "islands" as well as how well key fiducial markers mapped to these structures. For example, structures with concentrated maximal expression of CD3 or CD79a were preferred over apparent diffuse staining patterns. At time of writing, FCS Express currently has tSNE and UMAP directly available in the pipelines options. However it is possible to access and implement additional dimensionality reduction algorithms via the "miscellaneous" option in the "pipelines" sub menu and then selecting "Python Transformation". More details on what supporting software is required as well as the necessary scripts can be found here: https://denovosoftware.com/full-access/knowledge-base/python-transformation/. Briefly, once C++, Anaconda (Python) and the FCS Express Environment (for Anaconda) have been installed it is possible to select the desired script and install the package using Anaconda by selecting the FCS Express environment and opening up a new terminal. To install (for example) PacMap type pip install pacmap and the required dependencies will be installed automatically. It is then possible to copy the required script text from the FCS Express web link and paste into the python script window in the pipeline (clear the existing script first). If successful, on saving the relevant hyper-parameters should

be available as well as a list of input features to select. As mentioned the *arcsinh* plus Z-score normalised parameters were used. For PacMap, the "n neighbours" value was calculated based on the recommended equation (9) and for 109,535 single cells this was "26". For all other hyper-parameters and dimensionality reduction algorithms default settings were employed (Flt-SNE, TriMap, tSNE, UMAP).

S2.6.5 FLOWSOM and Phenograph Clustering

In order to determine the resident phenotypes in our segmented tonsil data set of 109,535 single cells we compared two different clustering algorithms namely FLOWSOM (10) and Phenograph (11); both of which are available in FCS Express directly in the "pipelines" menu. The panel of metal-tagged antibodies selected was designed to find at least 15 known cell phenotypes within the human tonsil (see table S1 and Figure S1) and as such allowed us to benchmark the different clustering approaches as well as the segmentation method. To create a FLOWSOM pipeline with embedded PacMap visualisation we created a new "pipeline" and then selected the "Pre-Defined Algorithms" option in the drop down. This allows access to a complete "FLOWSOM" script with all the required modules. The first step "new scaling" is not required as the data has already been transformed within the FCS files and can be deleted or unticked. The last step "New parameter removal" can also be removed. In the "New Batch Self-Organizing Map" module, we selected the arcsinh transformed Zscore normalised parameters representing the 27 phenotypic markers in the panel. All other parameters were left as default and as such a 10 x 10 SOM grid generated 100 clusters (SOMs). The only other module we altered was the "new consensus clustering" where we asked for 30 consensus clusters to be derived from the original 100 SOMs (twice the expected 15 clusters, so still over clustered). We then created a new pipeline module to run Phenograph clustering. Again this was done by selecting the Pre-Defined Algorithms" option in the drop down menu, but this time selecting "Phenograph". As before, the first part of the module "scaling" can be removed as the data has already been suitably scaled/normalised. Within the "New K-nearest neighbour" module we selected

a value for "number of neighbours" that generated 30 clusters. The value used for the 109,535 single cells was 17-18. Cluster number was assessed by creating a plate heat map from the "insert" tab "heat map" option and selecting "Louvain communities" to display on the x-axis. Once all required cluster parameters and PacMap coordinates were created and basic outputs checked, we exported the data as a single merged .DNS file (Data Stream) using the "export" function in the "data" menu. We also created individual .DNS files by repeating the export step but selecting the "export split on classification" drop down menu and selecting "classification identifier". The reason to create .DNS files is that they are smaller than the source files and will contain all the previous features as well as the new cluster info and PacMap coordinates. All further analysis and exploration were done using the .DNS files.

S2.6.6 Data visualisation, exploration and analysis

A new incidence of FCS Express was launched and the merged .DNS file containing all original metadata and features plus the newly created dimensionality reduction coordinates for UMAP and PacMap as well as all FLOWSOM cluster information (100 SOMs and 30 cSOMs) was loaded into the data view. Any meta-data gates such as those set previously for batches were recreated using a bivariate density plot of batch (x) versus Ir 193 signal (*arcsinh* and Z-score version on y). The "plate heat map" option was selected from the "heat map" menu in the "insert" tab and either the "batch SOM cluster assignment" option was selected for x-axis display on the plot. This created a radial spanning tree of the original 100 SOMs. It was then possible to right click on this plot and select the "format" option then "overlays". This gives access to the ability to select a different radial statistic for a given marker and emphasise any SOMs characterised by high expression of that marker on the plot. We then created a second plate heat map and this time selected "consensus clustering assignments" as the xaxis parameter from the plots drop down menu. This displayed the 30 cSOMs in a grid of 10 x 3. We then used the "well gate" function from the "gating" tab and selected all "wells" and asked to create individual gates for each with the prefix "cSOM" and to select individual colours for each (however this function often works poorly so manual colour selection is required). It was then possible to create a new PacMap colour dot plot, and select the "gates to display" option in the plot format menu (again accessible via a right mouse click on the plot).

In order to explore the outputs from the clustering (FLOWSOM or Phenograph), "parameter heat maps" were selected from the "insert" then "heat maps" sub menu. It is then possible to introduce multiple overlays of the same merged .DNS file on a "per gate" or in this case "per cluster" basis. To do this, right click on the newly created heat map and select "Add Overlay using Advanced Open Data Dialog". This is the quickest way to select the current open file for duplication. Once selected a new window will open and select the "duplicate for gates" option. Select all the necessary cSOM gates and click ok. The heat map requires further editing however to generate an output that can be displayed and interpreted. To modify the heat map further, right click and select "format". Select "specific options" and then choose (as a preference) to display in portrait mode. This now places the markers (parameters) on the y-axis (rows) and the cSOMs (clusters) on the x-axis (columns). At this stage, also select the option to "apply colours based on the values in each column" as this will normalise the heat map values by marker within each cluster (only possible because we have used the Z-score normalised versions). Next select "parameters to display" and ensure that the appropriate *arcsinh* and Z-score normalised metal channel features are selected by ticking the "only the items checked below" box and then selecting the correct parameters. Next go to the "overlays" option and select all overlays in the list. Change the "statistic to show" setting to the "median". It is also good practice to check that all the clusters are shown. It is also possible in this menu to move clusters (columns) left or right in the heat map order. This is useful after annotation to group all T-cell subsets (for example) together. It is also often useful to modify the axis. This can be done in the "axes" sub menu and allows one to change font sizes, colours and also rotate labels 90 degrees; this is useful for the cluster names. Cluster annotation was done using expert analysis and input. In this case it was guided by the panel information shown in table S1. Briefly, gate (cluster) names were edited to reflect broad or specific

annotations. Where two clusters were clearly very similar, they were given a label such as "Memory T cells 1, 2 etc." Where the cluster was deemed to be unique it was given a more definitive annotation. To create merged populations, we used the "well gate" approach on a plate heat map showing all consensus clusters (so 30 in this case from both FLOWSOM and Phenograph). Merging was achieved by multi-selecting the appropriate clusters and the new merged cluster was given a definitive annotation (for example Memory CD4 T cells). A new parameter heat map was then created to only show the final merged consensus clusters.

As a second stage of validation, we also plotted simple colour dot plots with the centroid x (x-axis) versus the centroid y (y-axis) coordinates and coloured by either cSOMs or the final merged clusters. We did this to ensure that the cell types/states we had annotated mapped to the expected structure and anatomical locations in the tonsil (see figure S1). In order to be able to show one ROI/FCS file at a time, we also used the plate heat map function to display file identifier on the x-axis and create well gates for each of the 24 merged tonsil FCS files. We could then show/display each gated file/sample/image on the x/y centroid plot. We found that the centroid plots were visually more pleasing when we set the fill background to black and also removed all axes ticks, labels and titles. This can be done via the "format" menu.

Once all clusters were annotated we created a simple excel file that contained the name and number of the final clusters (so 1-22 for example) and info regarding which cSOMs made up these final clusters. So for example cluster 1 could be named as "memory CD4 T cells" and could be made up from merging cSOM3 and cSOM9 from the original 30 cSOMs. We also created a .CSV using the "export" option from the "data" menu. We only selected the minimal information needed for the advanced spatial mapping and neighbourhood analysis. These were "file identifier" "centroid x", centroid y" and "consensus clustering assignment". This created a matrix of all 109,535 cells across all 24 images, each assigned a cluster membership (1-30). Finally, we also created a batch statistics export for further analysis of the final clusters. To do this, we opened up a new incidence of FCS Express and loaded the individual .DNS files for all 24 tonsil ROIs. As mentioned previously, these contained all the necessary consensus clustering info as well as the original meta-data from the original FCS files. We had to recreate the cluster (cSOM) gates using the plate heat map option and individual well gates with the "cSOM" prefix. We also created a single bi-variate plot and displayed the first file in the data list. Use any two parameters, however we often use centroid x versus centroid y. This is important as the "data source" guide for running the stats report as a batch process (see later). Once these gates were created (could be using any of the .DNS files), we navigated to the "batch" option on the menu ribbon. We then selected "batch actions". In this new window, under the "add report" window, we select the "export to EXCEL (column mode)". Leave the option to "start with an empty file" in the "append option" area and in the "output file options" direct where to save the file and what to call it. Select "ok" and it should now be possible to add in the statistics required in the output. To add these, right click on the "EXCEL (column mode)" in the "Batch Process Action" window and select "add item" and "statistics token". Select in the sub window "data source" and select the single bi-variate plot created previously. Next move to the "Statistic" menu. In the "gate" window select "no gate" and in the "statistic" window select "filename". This means the first column on the export will be the FCS file/ROI name. Press "Ok" to add to the export actions and it should now appear under the "EXCEL (column mode)" in the "Batch Process Action" window. Highlight the bi-variate plot so the borders turn red and drag it on to the "EXCEL (column mode) icon. This should open a new window that says "choose the format to paste". Select "statistics token". This will open up a new window where it is possible to select all possible gates (ungated to cSOM X) by holding down the shift key and left clicking on the last gate at the bottom of the list. In the "statistics" window, again hold down control and multi-select "filename", "# of events" and (optional) "% of all cells". Once the EXCEL stats export routine has been finalised, ensure that the bivariate plot is showing the data from the first file in the data list then select "Run" in the "Batch" menu. FCS Express will then visibly iterate through each sample and create an EXCEL file from

the results. These data can then be further analysed either in EXCEL or PRISM (Graph Pad) and should be arranged as: First column: filename (ROI); Second column: Total #cells (ungated); Third column: #cells in cSOM 1 and so on. The export of frequencies is optional as they can be derived using the total cell counts and the counts per cSOM.

S2.7 Spatial analysis

The MATLAB code developed for spatial analysis works by examining each cell identified within the cell mask file for a single image and is based on the code previously described in HistoCat (12). It uses that cell mask and expands the cell region by the number of pixels specified by the user within the code. The function will then count the number of unique cells found within the expanded cell area, excluding the original cell, keeping a list of each cell identifier within the label image. The cluster identities of each of these neighbour cells is then assessed. This process is repeated for each cell within a ROI. In order to compare the spatial arrangement of the cells for higher-than-expected or lowerthan-expected interactions (where no positive or negative interactions are the nul hypothesis), the cell mask image is then used as a framework with the cell cluster identities randomly mapped onto the cell locations. The number of random images generated is dictated by the "Iteration" value specified by the user within the MATLAB code. If the original image shows a higher level of clustering than the randomly distributed images in more than 1-(significance threshold) proportion of the random images, then the interaction is deemed significantly positive for that image. If the reverse is true, showing a lower level of neighbour correspondence than the random images, then a significantly negative interaction is identified. In all other cases, no significant interaction is identified. The number of ROIs belonging to a group (i.e. pathology) that show a positive, negative, or neutral interaction is then averaged over the group condition, outputting a single heatmap with values ranging from a possible 1 (all images showed significant interaction between the cell types) to -1 (all images showed a significant avoidance between the cell types), with intermediate values indicating that some images showed interactions while others have not. The heatmap generated is asymmetric, allowing

assessment of interactions that occur uni-directionally (i.e. cell type X interacts positively with cell type Y, but cell type Y does not interact in either direction with cell type X).

Tested within this research were three methods of spatial analysis, based on the above description. The first method (Ungated Original) used the pixel expansion method used within the HistoCat code. Another method (Ungated Disk) performs the process above the same, with the exception of using an alternative radial approach to cell area expansion (Disk erosion/dilation), which appears to better preserve the shape of the cells examined and thus more likely to find "true" neighbours. Finally, an additional gating step was incorporated (Gate Disk), which excluded cells based on two factors: area and edge. Only cells with physiologically reasonable areas (20-200 μ m², equating to 5 – 15 μ m dimeter) were included, and cells in contact with the edge of the image (and thus showing inaccurate neighbour reporting on their blank edged) were also excluded.

Visualisation of the cluster spatial information was also performed by applying a high-contrast colour labelling to the original cell masks based on their reported cluster identity, with the possibility to map only selected cell types for ease of visualisation.

S2.8 Code availability

All code and data files used in this manuscript is available here:

https://www.ebi.ac.uk/biostudies/studies/S-BSST1047

Supplemental Figure Legends

Table S1: Key information concerning the panel of 27 antibodies used in this study to stain human tonsil tissue over 12 individual, temporally distinct batches. The final two columns denote the "ground truth" cell type that each marker in isolation or in combination was selected to identify in human tonsil as well as the expected spatial location in order to benchmark the clustering approaches used in this study.

Figure S1: (A) A single IMC image showing Iridium (DNA, red pseudo colour) overlaid with Ki67 (white pseudo colour) and CD31 (green pseudo colour) as per legend key. The major anatomical and structural features are labelled. (B) A gallery of grey scaled IMC images showing staining patterns for each of the 27 antibodies in table S1 plus the two Iridium channels (191 and 193). (C) Images showing the comparative staining coverage of EPCAM (left panel) and a combination of immune cell membrane signals (right panel).

Figure S2: IMC images from all 24 Tonsil ROIs across the 14 temporally distinct staining batches. Images shown are at 1 pixel per micron and equivalent to a 10x optical image and are between 0.25 and 0.5 mm² of total image area. The pseudo-coloured overlays show DNA by virtue of Iridium intercalator (red), CD3 expression (Blue) and CD79a expression (Green) as per the legend. In all cases ROIs were selected to try and capture as much of the diverse tonsil structure as possible such as follicles/germinal centres and epithelium.

Figure S3: Cell segmentation maps for all 24 Tonsil ROIs across the 12 temporally distinct staining batches derived from the "Tonsil EPCAM model" of cell segmentation whereby the tonsil nuclear image was used to train an Ilastik model based on classification of nuclear pixels and background, then single cell boundaries were defined by the EPCAM membrane signal using CellProfiler to generate single cell information from the Ilastik model-derived probability map. The markers are denoted by the pseudo colours as shown in the legend.

Figure S4: Cell segmentation maps for all 24 Tonsil ROIs across the 12 temporally distinct staining batches derived from the "Nucleus only" of cell segmentation whereby the tonsil nuclear image was used directly by CellProfiler to generate single cell information without any Ilastik-derived probability map. In all cases the object boundaries are shown on a merge of the nuclear (iridium) and EPCAM images.

Figure S5: Bi-variate single cell level scatter plots for all 24 Tonsil ROIs across the 12 temporally distinct staining batches derived from the "Tonsil EPCAM model" of cell segmentation showing CD3 intensity levels on the x axis versus CD79a intensity levels on the y axis. Both parameters are the *arcsinh* c.f.1 transformed, Z-score normalised versions. Gates have been set to determine the frequency of CD3 and CD79a single positive events, as well the double positive (DP) "nonsense" cells.

Figure S6: Bi-variate single cell level scatter plots for all 24 Tonsil ROIs across the 12 temporally distinct staining batches derived from the "Nucleus only" of cell segmentation model showing CD3 intensity levels on the x axis versus CD79a intensity levels on the y axis. Both parameters are the *arcsinh* c.f.1 transformed, Z-score normalised versions. Gates have been set to determine the frequency of CD3 and CD79a single positive events, as well the double positive (DP) "nonsense" cells.

Figure S7: The top row shows UMAP plots coloured by batch (as indicated in the legend) created using the *arcsinh* transformed parameters only (first column), the *arcsinh* parameters transformed by 0-1 scaling (middle column) and the *arcsinh* parameters Z-score transformed. The middle row of UMAP plots are density weighted and the bottom row are density weighted by CD3 expression.

Figure S8: Cell cluster maps showing for all 24 Tonsil ROIs across the 12 temporally distinct staining batches derived from the "Tonsil EPCAM model" of cell segmentation as in S2. All 21 final cell type clusters are shown as per the legend.

Figure S9: (A) A graphical representation of how the cluster threshold function works when considering across all images what clusters to measure in the neighbourhood analysis. In the ten mock

images shown, the brown population only appears in one image so setting a 10% threshold means that it will not be considered however at a 1% it would be. The purple population appears in nine of ten images so at a 90% cut off, this population and the brown population will not be included in any neighbourhood analysis. (B) A graph showing the frequency of all 21 final cSOMs in all of the 24 tonsil images. Note that all 24 images contained all of the 21 final cSOMs making the threshold function unnecessary in this study.

Figure S10: (A) Segmentation checker imagines for an expanded set of cell segmentation approaches as indicated. (B) Bi-variate plots for the single cell segmented outputs from the ROIs shown in A with CD3 (x axis) and CD79a (y axis) expressions levels plotted. Gates are set to capture the frequency (%) of CD3 single positive, CD79a single positive and double positive events. (C) PacMap plots for each of the single cell outputs for all ROIs from each of the respective segmentation approaches with expression density set on CD3 levels. (D) The same as in C but with expression density set on CD79a levels. (E) Heatmaps of the FLOWSOM clustering outputs from either the "nuclear only II" (left panel) or CellPose (right panel) segmentation approaches.

This user guide will assist users in using the macro and scripts present within this repository. For further information or queries, please contact George Merces at <u>George.Merces@newcastle.ac.uk</u>

Folder Organisation

1. Arrange your folders according to the naming convention listed below

Name	Date modified	Туре	Size
📙 1_ImageData	22/05/2023 21:51	File folder	
2_image_extraction	11/07/2023 15:10	File folder	
3_ilastik_model	22/05/2023 16:27	File folder	
4_structured_data_for_CP	24/05/2023 11:38	File folder	
5_CP_pipeline	23/05/2023 10:07	File folder	
fcs_files	24/05/2023 22:23	File folder	
5_CP_pipeline_vero.zip	24/05/2023 23:03	Compressed (zipp	757,010 KB
fcs_files.zip	24/05/2023 22:51	Compressed (zipp	92,763 KB
Notes.txt	22/05/2023 16:29	Text Document	1 KB

- 2. Put all raw .tiff images into the folder "1_ImageData"
- 3. Put the Matlab macros, along with any excel files needed for the macros, in folder "2_image_extraction" and organise according to the image below

Name	Date modified	Туре	Size
	22/05/2023 16:07	File folder	
	22/05/2023 15:49	File folder	
😰 Gut_Data_Analysis_Master.xlsx	02/05/2023 14:06	Microsoft Excel W	39 KB
🕼 Gut_Data_Analysis_Master_1stRun.xlsx	20/04/2023 11:35	Microsoft Excel W	37 KB
Number_Name_Conversion.xlsx	22/05/2023 15:45	Microsoft Excel W	10 KB
🖺 OPTIMAL_Pipeline_Part_1.m	24/05/2023 22:03	MATLAB Code	22 KB
🖺 OPTIMAL_Pipeline_Part_2_Heatmaps.m	14/03/2023 16:08	MATLAB Code	14 KB
🖺 OPTIMAL_Pipeline_Part_3_Visualisaton.m	14/03/2023 16:45	MATLAB Code	13 KB
🖬 Saskia_MAY23.xIsx	23/05/2023 15:00	Microsoft Excel W	21 KB
📭 spillover2.1.xlsx	24/05/2023 11:05	Microsoft Excel W	16 KB

4. Structure folder "3_ilastik_model" to have your ilastik model in the folder, with a folder for your output Pmaps (this must match with what is in the Matlab macro later, so ensure the code matches the information)

Name	Date modified	Туре	Size
OUTPUT_PMAPS	24/05/2023 12:18	File folder	
👔 20210906_Tonsil_pixel_classification_v1.0	22/05/2023 16:27	ilastik project	35,035 KB
👔 20230220_JD_Nuclear_Model.ilp	22/02/2023 09:51	ilastik project	5,694 KB
🚪 20230328_JD_Nuclear_Model.ilp	22/05/2023 16:36	ilastik project	7,542 KB

5. The folder "4_structured_data_for_CP" should be empty when you initially run the Matlab macro for the first time, the script will populate the folder with other folders for the individual channels in your image and folders for the nuclear, membrane, and probability

map)S
-----	----

Name	Date modified	Туре	Size	
Channel_001	24/05/2023 12:18	File folder		
Channel_002	24/05/2023 12:18	File folder		
Channel_003	24/05/2023 12:18	File folder		
Channel_004	24/05/2023 12:18	File folder		
Channel_005	24/05/2023 12:18	File folder		
Channel_006	24/05/2023 12:18	File folder		
Channel_007	24/05/2023 12:18	File folder		
Channel_008	24/05/2023 12:18	File folder		
Channel_009	24/05/2023 12:18	File folder		
Channel_010	24/05/2023 12:18	File folder		
Channel_011	24/05/2023 12:18	File folder		
Channel_012	24/05/2023 12:18	File folder		
Channel_013	24/05/2023 12:18	File folder		
Channel_014	24/05/2023 12:18	File folder		
Channel_015	24/05/2023 12:18	File folder		
Channel_016	24/05/2023 12:18	File folder		
Channel_017	24/05/2023 12:18	File folder		
Channel_018	24/05/2023 12:18	File folder		
Channel_019	24/05/2023 12:18	File folder		
Channel_020	24/05/2023 12:18	File folder		
Channel_021	24/05/2023 12:18	File folder		
Channel_022	24/05/2023 12:18	File folder		
Channel_023	24/05/2023 12:18	File folder		
Channel_024	24/05/2023 12:18	File folder		
Channel_025	24/05/2023 12:18	File folder		
Channel_026	24/05/2023 12:18	File folder		
Channel_027	24/05/2023 12:18	File folder		
Channel_028	24/05/2023 12:18	File folder		

6. Structure the folder "5_CP_pipeline" with two folders, "OUTPUTS" and "PIPELINE". The "PIPELINE" folder should only contain the CellProfiler pipeline file to be used with the Matlab macro. The "OUTPUTS" folder should be organised according to the image below. Note, the excel files and the contents of these subfolders will not exist until after the CellProfiler pipeline has completed running

Name	Date modified	Туре	Size
CellMasks	24/05/2023 19:00	File folder	
Segmentation_Checker	24/05/2023 19:00	File folder	
Tissue	24/05/2023 19:00	File folder	
🖬 MyExpt_Cells.csv	24/05/2023 19:17	Microsoft Excel C	4,144,522 KB
MyExpt_Experiment.csv	24/05/2023 19:17	Microsoft Excel C	222 KB
🖬 MyExpt_Image.csv	24/05/2023 19:17	Microsoft Excel C	1 KB
🖬 MyExpt_Nuclei.csv	24/05/2023 19:24	Microsoft Excel C	1,654,134 KB
MyExpt_secondary_objects.csv	24/05/2023 19:30	Microsoft Excel C	1,654,134 KB

Metadata Incorporation

1. Format your metadata excel file according to the example below. Any data that is planned on being incorporated into the fcs files at the end of the macro must be in numerical format

AutoSave 💽 🖽 🎾 Y C 🔍 👻			,♀ Search										George Me	erces GM			×
File Home Insert Page Layout Form	ulas Data Review View Automate	Help													Comments	s 🖻 Shi	are ~
Calibri ~ 11	• A* A* ≡ ≡ ₩• • 8₽ Wrap Text	General	-		Normal	Bad	Good	1 ·	*		∑ AutoSu I Fill ~	m č <mark>A</mark> V	0				
v SFormat Painter B I U v ⊞ v 🖉	💁 🗸 🖌 📄 🚍 🚍 🖽 🖽 Merge & (ienter - 🖽 - % 🤊	500 +0 Condit	ional Formatas ting ~ Table ~	Neutral	Calculation	Chec	k Cell	insert Delet	E Format	🞸 Clear 🗸	Sort & Filter ~	Select ~	Data	Sensitivity		
Clipboard 5 Font	5 Alignment	Fa Numbe	r G	-		Styles			Cells			Editing		Analysis	Sensitivity		~
F56 * 1 × √ <i>f</i> x																	~
	-									1			-				
1 Sample id	Patch no	Richark	Tirruo	Timo			,	ĸ	L M	IN	0	P	ų	ĸ	5		
2 BATCH001 BOI003 TONSIL START ome tiff	batch_no	1 NA	Tonsil	Time	1												
3 BATCHOOL ROUDS TONSIL END ome tiff		1 NA	Tonsil		2												
4 BATCH002 BOI003 TONSIL START ome tiff		2 NA	Tonsil		1												
5 BATCH002 ROI006 TONSIL END.ome.tiff		2 NA	Tonsil		2												
6 BATCH003 ROI003 TONSIL START.ome.tiff		3 NA	Tonsil		1												
7 BATCH003 ROI006 TONSIL END.ome.tiff		3 NA	Tonsil		2												
8 BATCH004 ROI003 TONSIL START.ome.tiff		4 NA	Tonsil		1												
9 BATCH004_ROI006_TONSIL_END.ome.tiff		4 NA	Tonsil		2												
10 BATCH005_ROI003_TONSIL_START.ome.tiff		5 NA	Tonsil		1												
11 BATCH005_ROI004_TONSIL_END.ome.tiff		5 NA	Tonsil		2												
12 BATCH006_R0I003_TONSIL_START.ome.tiff		6 NA	Tonsil		1												
13 BATCH006_ROI006_TONSIL_END.ome.tiff		6 NA	Tonsil		2												
14 BATCH007_ROI003_TONSIL_START.ome.tiff		7 NA	Tonsil		1												
15 BATCH007_ROI006_TONSIL_END.ome.tiff		7 NA	Tonsil		2												
16 BATCH008_ROI003_TONSIL_START.ome.tiff		8 NA	Tonsil		1												
17 BATCH008_ROI006_TONSIL_END.ome.tiff		8 NA	Tonsil		2												
18 BATCH009_ROI003_TONSIL_START.ome.tiff		9 NA	Tonsil		1												
19 BATCH009_ROI006_TONSIL_END.ome.tiff		9 NA	Tonsil		2												
20 BATCH010_ROI003_TONSIL_START.ome.tiff		10 NA	Tonsil		1												
21 BATCH010_ROI006_TONSIL_END.ome.tiff		10 NA	Tonsil		2												
22 BATCH011_R0I003_TONSIL_START.ome.tiff		11 NA	Tonsil		1												_
23 BATCH011_R0I006_TONSIL_END.ome.tiff		11 NA	Tonsil		2												_
24 BATCH012_R0I003_TONSIL_START.ome.tiff		12 NA	Tonsil		1												_
25 BATCH012_ROI006_TONSIL_END.ome.tiff		12 NA	Tonsil		2												_
26																	_
27																	
28																	-
29																	_
30																	-

Training An Ilastik Model

Note, if you are using the llastik model provided, this step may be skipped, and you can just follow the instructions for "OPTIMAL_Pipeline_Part_1.m".

- 🜠 ilastik No Project Loaded × Project Settings Help Create New Project xel Class Autocontext (2-stage) Pixel Classification + Object Classification Object Classification [Inputs: Raw Data, Pixel Prediction Map] Object Classification [Inputs: Raw Data, Segmentation] Manual Tracking Workflow [Inputs: Raw Data, Pixel Prediction Map] Tracking [Inputs: Raw Data, Binary Image] Tracking [Inputs: Raw Data, Pixel Prediction Map] Animal Tracking [Inputs: Raw Data, Binary Image] Animal Tracking [Inputs: Raw Data, Pixel Prediction Map] Tracking with Learning [Inputs: Raw Data, Binary Image] Tracking with Learning [Inputs: Raw Data, Pixel Prediction Map] 🔶 Carving Boundary-based Segmentation with Multicut Cell Density Counting Data Conversion Open Project ... Browse Files Active Requests: 0 Cached Data: 0.0 MB ...
- 1. Open Ilastik and create a Pixel Classification Model

2. Drag a selection of your images (in our case, we use rescaled nuclear channel images) into the images window. Note, for ease of distribution of your model, we recommend setting the image "Location" to "Internal", which increases the file size for your model but ensures it

3. Select you Features for use. We recommend selecting a wide range of options



4. Train your model: Rename the label options to be relevant for your sample (e.g. nuclear vs non-nuclear) and label pixels sporadically around your images. Do not label too much, as this can lead to a lack of general ability of the model, and can train it too heavily to the training

should be executable by anyone who downloads it.



5. Select "Live Update" to train your model. If the output probability is not good enough, try adding additional labelling information to your images, focussing on regions where the probability mapping has made errors



6. Once you're happy with your probability mapping model, choose the export settings. In the Matlab script given as part of this manuscript, we only export the nuclear option, and we export this to a location close to the Ilastik model file. The exports are then detected and resaved in the appropriate location for CellProfiler with a different name within the Matlab script. Note the name used to save the image under, as this needs to match with the Matlab

script for proper function.

1 Janual Darks	Dataset	Export Location	Action			•
	R0004 207502-4Ch 051 MS 117	Y:/Bioimaging_Staff/George/	Evenet			
2. Feature Selection	R01001 200105-1C5-051-MS-001	20230220_ID_Hyperion/20230221_Full_Datas Y:/Bioimaging_Staff/George/	Propert			
3. Training		20230220_ID_Hyperion/20230221_Full_Datas Y:/Bioimaging_Staff/George/	Count .			
 4. Prediction Export 	K01001_204410-101_001_M3_010	20230220_JD_Hyperion/20230221_Full_Datas Y//Bioimaging Staff/George/	Export			
Export Settings	R0I001_206601-1Ch_061_MS_015	20230220_ID_Hyperion/20230221_Full_Datas	Export			
Source: Probabilities •	R01001 307303-32% 061 MS 022	M Image Export Options	Eurort	?	×	
Choose Export Image Settings						
Actions		Source Image Description	tute Orden une	Data Turan Bast 77	St. 1	
🕭 Export All 👌 Delete All		Shaper (753, 663, 1)	ANS CIGET YAC	Deta type: mataz		
1. Both Processing Used 1: Provider 0=25.0% Kerr bits 0=100.0%		Tradivative ⊂ An → → → → → → → → → → → → → → → → → →	Interpret di Stati - Interpret	25 Data Typer until UTUT / NARS (robuster) / Frank		
				× 4		

7. Save your finished model to an appropriate location, and make sure this is the location called within the Matlab script to ensure proper macro function

Matlab Code I

1. Open llastik on your device, select an appropriate folder to save your output probability maps to. Make sure the settings are as listed in the image to ensure correct use of the

images in macro running

📶 Image Export Options			? ×
Source Image Description			
Shape: (500, 500, 1)	Axis Order: yxc	Data Type: float32	
Cutout Subregion			
y ✓ All ↓ x ✓ All ↓ c All 0 ↓ 1			
Transformations Convert to Data Type: integer 8-bit Renormalize [min,max] from: 0.00 Transpose to Axis Order: Output Image Description	1.00 to: 0 255]	
Shape: (464, 465, 1)	Axis Order: yxc	Data Type: uint8	
Output File Info Format: tiff v File: Y:/Bioimaging_Staff/George/20230315_Auto	DECRA/3_ilastik_model/OUTPUT_PMAPS/	{nickname}_pmap.tiff	Select Reset filepath
		ОК	Cancel

2. Open the Matlab script "OPTIMAL_Pipeline_Part_1.m". Rename the folders at the start according to their locations on your current device



 In the Matlab script, select the channels for your membrane/cytoplasmic stain to use for cell boundary determination (EPCAM_channel) and also the DNA stain for nucleus probability mapping (nuclei_channel). Please note, channel number 1 refers to the first channel in your image (not 0 as in some software)

💋 E	ditor		Bioimaging_Staff\George\20230315_AutoDECRA\2_image_extraction\OPTIMAL_Pipeline_Part_1.m	6	•	×
	ΟΡΤΙ	MAL	_Pipeline_Part_1.m 🗙 OPTIMAL_Pipeline_Part_1.m 💥 OPTIMAL_Pipeline_Part_1.m 💥 🕂			
3	5	L	%%this is the same folder you are saving your mask npy files to	4	•	A
3	6		<pre>Mask_working_directory='Y:/Bioimaging_Staff/George/20230315_AutoDECRA/5_CP_pipeline\OUTPUTS\CellMasks\';</pre>			
3	7					
3	8					
3	9				۰.	_
4	0		%%Sets the display rescaled data option to True			-
4	1		display_rescaled_data = true;			=
4	2					
4	3	Ę	%%Defines the channel number relating to the membrane stain and the nuclear			
4	4	L	%%stain, respectively			
4	5		EPCAM_channel = 29;			
4	6		nuclei_channel = 51;			_
4	7					_
4	8	P	%%Takes the Tiff file names from the raw data directory and transposes them			-
4	9	-	%%into an array			-
5	0		<pre>raw_data_info = dir ([raw_data_directory,'*.tiff']) ;</pre>			
5	1		<pre>image_names = {raw_data_info.name}';</pre>			-
5	2	_				
5	3	닏	%%Runs a loop on each of the images in your raw data directory			
5	4		%%Change this to a single image for testing purposes if necessary			
5	5	_	*************************			
5	6	上	for Loop = 1:Length(image_names)			
5	2	닌	Aberines the image to be processed based on the loop progress, finding			
5	8		ATTE COPPECT Image name based on this			-
5	9		**************************************			
6	0		image_number = 100p			Ξ
6	1		fileName = Char(Image_names(Image_number))			=
0	2		i TTEAQUE ≥ I TTEAQUE (1:ELOL.2)			
6	4				Ŧ	
0	-		4	•		

4. Rename the locations for Ilastik and Cell Profiler, using the **exact** format shown in the template code. The spacing and additional features are essential for proper function



 Incorporate the names of the excel files to be used for metadata incorporation within the fcs files at the end of the code, including the spillover matrix for compensation of signal bleedthrough.

Z Edit	101 - 1	:\bioimaging_start\seorge\20230315_AutoDECKA\2_image_extraction\OP1IIVIAL_Pipeline_Part_1.m		0 ×
∫ OI	PTIMA	AL_Pipeline_Part_1.m × OPTIMAL_Pipeline_Part_1.m × OPTIMAL_Pipeline_Part_1.m × +		
224		%% Simple error check based on number of ROI	4	- 🔺
225		if Number_of_masks~=Number_of_ROI		
226		error('Number of masks and ROI dont match');		
227		end		
228			_	
229		%% Load compensation matrix and extract important information		
230				=
231		comp_matrix=readtable(splitover2.1.Xisx); comp_comp_comp_matrix_readtable(splitover2.1.Xisx);		
232		comp_channels-comp_mocra.roper(is.verlablevames;		
233		compensation matrix-table2creation matrix(: 2:end)):		
235		compensation_metric-tepterseterseterseterseters		
236				=
237				
238		%% Load the ROI spreadsheet to get the PATIENT metadata	_	1-1
239		XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX		
240		<pre>ROI_data = readcell('Tonsil_Data_Analysis_Master.xlsx','sheet',4);</pre>		1-
241		<pre>Patient_data = readcell('Tonsil_Data_Analysis_Master.xlsx','sheet',5);</pre>		
242		Patient_identifier=Patient_data(2:end,2);		-
243		ROI_list=ROI_data(2:end,1);		-
244		Patient_name_list=ROI_data(2:end,3);		
245	F	<pre>for loop=1:numel(Patient_name_list)</pre>		
246	F	for loop2=1:numel(Patient_identifier)		
247		<pre>if contains(Patient_name_list(loop),Patient_identifier(loop2)) ==1</pre>		-
248		Patient_number_list(loop)=loop2;		
249		end		
250	F	end		=
251	L	end		-
252		***************************************		
200			•	
📝 Edito	or - Y:	\Bioimaging Staff\George\20230315 AutoDECRA\2 image extraction\OPTIMAL Pipeline Part 1.m	(×
Edito	or - Y: TIMAI	\Bioimaging_Staff\George\20230315_AutoDECRA\2_image_extraction\OPTIMAL_Pipeline_Part_1.m	0	×
CP CP	or - Y: TIMA	\Bioimaging_Staff\George\20230315_AutoDECRA\2_image_extraction\OPTIMAL_Pipeline_Part_1.m L_Pipeline_Part_1.m × OPTIMAL_Pipeline_Part_1.m × OPTIMAL_Pipeline_Part_1.m × +	6	×
Z Edito	or - Y: TIMAI	\Bioimaging_Staff\George\20230315_AutoDECRA\2_image_extraction\OPTIMAL_Pipeline_Part_1.m L_Pipeline_Part_1.m \CopTIMAL_Pipeline_Part_1.m %%Load the spreadsheet containing essential metadata (in this example,) ×
255 256	or - Y: TIMAI	<pre>\Bioimaging_Staff\George\20230315_AutoDECRA\2_image_extraction\OPTIMAL_Pipeline_Part_1.m L_Pipeline_Part_1.m X OPTIMAL_Pipeline_Part_1.m X + %%Load the spreadsheet containing essential metadata (in this example, %%Patient, Batch_no, Time, and Tissue)</pre>	•	×
255 256 257		<pre>\Bioimaging_Staff\George\20230315_AutoDECRA\2_image_extraction\OPTIMAL_Pipeline_Part_1.m L_Pipeline_Part_1.m X OPTIMAL_Pipeline_Part_1.m X + %%Load the spreadsheet containing essential metadata (in this example, %%Patient, Batch_no, Time, and Tissue) %***********************************</pre>	•	×
255 256 257 258	Dr - Y: TIMAI	<pre>\Bioimaging_Staff\George\20230315_AutoDECRA\2_image_extraction\OPTIMAL_Pipeline_Part_1.m L_Pipeline_Part_1.m X OPTIMAL_Pipeline_Part_1.m X + %%Load the spreadsheet containing essential metadata (in this example, %%Patient, Batch_no, Time, and Tissue) %%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%</pre>	•) ×
255 256 257 258 259	pr - Y: TIMAI	<pre>\Bioimaging_Staff\George\2023015_AutoDECRA\2_image_extraction\OPTIMAL_Pipeline_Part_1.m L_Pipeline_Part_1.m X OPTIMAL_Pipeline_Part_1.m X + %\Load the spreadsheet containing essential metadata (in this example, %\Patient, Batch_no, Time, and Tissue) %************************************</pre>	•) ×
Edito OP 255 256 257 258 259 260	pr - Y: TIMAI	<pre>\Bioimaging_Staff\George\20230315_AutoDECRA\2_image_extraction\OPTIMAL_Pipeline_Part_1.m L_Pipeline_Part_1.m X OPTIMAL_Pipeline_Part_1.m X + %\Load the spreadsheet containing essential metadata (in this example, %\Patient, Batch_no, Time, and Tissue) %\XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX</pre>	•	• ×
Edito OP 255 256 257 258 259 260 261		<pre>\Bioimaging_Staff\George\20230315_AutoDECRA\2_image_extraction\OPTIMAL_Pipeline_Part_1.m L_Pipeline_Part_1.m X OPTIMAL_Pipeline_Part_1.m X + %\Load the spreadsheet containing essential metadata (in this example, %\Patient, Batch_no, Time, and Tissue) %\X************************************</pre>	•	• A
Edito 255 256 257 258 259 260 261 262	pr - Y: TIMAI	<pre>\Bioimaging_Staff\George\20230315_AutoDECRA\2_image_extraction\OPTIMAL_Pipeline_Part_1.m L_Pipeline_Part_1.m X OPTIMAL_Pipeline_Part_1.m X + %%Load the spreadsheet containing essential metadata (in this example, %%Patient, Batch_no, Time, and Tissue) %%Patient_Batch_no, Time, and Tissue) %%Patient_name_list=Spreadsheet.Patient; Batch_number_list=Spreadsheet.Patient; for loop=1:numel(Patient_name_list); for loop=1:numel(Patient_inetidiation) </pre>	G	• ×
Copy Editor 255 256 257 258 259 260 261 262 263 264		<pre>\Bioimaging_Staff\George\2023015_AutoDECRA\2_image_extraction\OPTIMAL_Pipeline_Part_1.m L_Pipeline_Part_1.m X OPTIMAL_Pipeline_Part_1.m X + %\Load the spreadsheet containing essential metadata (in this example, %\Patient, Batch_no, Time, and Tissue) %************************************</pre>	C) ×
Contemporation Contemporatio Contemporation Contemporation Contemporation Contemp		<pre>\Bioimaging_Staff\George\2023015_AutoDECRA\2_image_extraction\OPTIMAL_Pipeline_Part_1.m L_Pipeline_Part_1.m X OPTIMAL_Pipeline_Part_1.m X + %\Load the spreadsheet containing essential metadata (in this example, %\Patient, Batch_no, Time, and Tissue) %************************************</pre>	((A
Copy Editor 255 256 257 258 259 260 261 262 263 264 263 264		<pre>\Bioimaging_Staff\George\20230315_AutoDECRA\2_image_extraction\OPTIMAL_Pipeline_Part_1.m L_Pipeline_Part_1.m X OPTIMAL_Pipeline_Part_1.m X + %\Load the spreadsheet containing essential metadata (in this example, %\Patient, Batch_no, Time, and Tissue) %\V************************************</pre>		• ×
Contemporation of the second s	Pr - Y: TIMAI	<pre>\Bioimaging_Staff\George\2023015_AutoDECRA\2_image_extraction\OPTIMAL_Pipeline_Part_1.m L_Pipeline_Part_1.m X OPTIMAL_Pipeline_Part_1.m X + %\Load the spreadsheet containing essential metadata (in this example, %\Patient, Batch_no, Time, and Tissue) % % % Spreadsheet=readtable('Tonsil_MAY22.xlsx'); Patient_name_list=Spreadsheet.Patient; Batch_number_list=Spreadsheet.Patient; for loop=1:numel(Patient_name_list); for loop=1:numel(Patient_name_list) for loop=1:numel(Patient_identifier) if contains(Patient_iname_list(loop),Patient_identifier(loop2)) ==1 Patient_number_list(loop)=loop2; end end</pre>		
Contemporation of the second s	Pr - Y: TIMAI	<pre>\Bioimaging_Staff\George\2023015_AutoDECRA\2_image_extraction\OPTIMAL_Pipeline_Part_1.m L_Pipeline_Part_1.m X OPTIMAL_Pipeline_Part_1.m X + %\Load the spreadsheet containing essential metadata (in this example, %\Patient, Batch_no, Time, and Tissue) %************************************</pre>		
 Edita OP 254 256 257 258 259 260 261 262 263 264 265 266 267 268 269 		<pre>\Bioimaging_Staff\George\2023015_AutoDECRA\2_image_extraction\OPTIMAL_Pipeline_Part_1.m L_Pipeline_Part_1.m X OPTIMAL_Pipeline_Part_1.m X + %\Load the spreadsheet containing essential metadata (in this example, %\Patient, Batch_no, Time, and Tissue) %************************************</pre>		
Z56 257 258 259 260 261 262 263 264 265 266 267 268 266 267 268 269 270	Pr - Y: TIMA	<pre>\Bioimaging_Staff\George\20230315_AutoDECRA\2_image_extraction\OPTIMAL_Pipeline_Part_1.m L_Pipeline_Part_1.m x OPTIMAL_Pipeline_Part_1.m x + %\Load the spreadsheet containing essential metadata (in this example, %\Patient, Batch_no, Time, and Tissue) %************************************</pre>		
 Edita OP 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 	Pr - Y: TIMAA P	<pre>\Bioimaging_Staff\George\2023015_AutoDECRA\2_image_extraction\OPTIMAL_Pipeline_Part_1.m L_Pipeline_Part_1.m X OPTIMAL_Pipeline_Part_1.m X + %\Load the spreadsheet containing essential metadata (in this example, %\Patient, Batch_no, Time, and Tissue) % % % % % % % % % % % % % % % % % % %</pre>		
 Editt OP 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 		<pre>\Bioimaging_Staff\George\20230315_AutoDECRA\2_image_extraction\OPTIMAL_Pipeline_Part_1.m L_Pipeline_Part_1.m X OPTIMAL_Pipeline_Part_1.m X + %\Load the spreadsheet containing essential metadata (in this example, %\Patient, Batch_no, Time, and Tissue) % %**********************************</pre>		
 Editt OP 255 256 257 258 260 261 262 263 264 265 266 267 268 269 270 271 273 	Pr - Y: TIMAI	<pre>\Bioimaging_Staff\George\20230315_AutoDECRA\2_image_extraction\OPTIMAL_Pipeline_Part_1.m L_Pipeline_Part_1.m x OPTIMAL_Pipeline_Part_1.m x + %\Load the spreadsheet containing essential metadata (in this example,</pre>		
 Editt OP 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 		<pre>\Bioimaging_StafkGeorge\20230315_AutoDECRA\2_image_extraction\OPTIMAL_Pipeline_Part_1.m X</pre>	C	
 Editt OP 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 		<pre>\Bioimaging_Staff\George\20230315_AutoDECRA\2_image_extraction\OPTIMAL_Pipeline_Part_1.m X</pre>	6	
 Editt OP 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 273 274 275 276 		<pre>\Bioimaging_Staff\George\20230315_AutoDECRA\2_image_extraction\OPTIMAL_Pipeline_Part_1.m L_Pipeline_Part_1.m x OPTIMAL_Pipeline_Part_1.m x + %\Load the spreadsheet containing essential metadata (in this example, %\Patient, Batch_no, Time, and Tissue) %***********************************</pre>	•	
 Editt. COP 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 		<pre>\Bioimaging_Staff\George\20230315_AutoDECRA\2_image_extraction\OPTIMAL_Pipeline_Part_1.m L_Pipeline_Part_1.m x OPTIMAL_Pipeline_Part_1.m x + %\Load the spreadsheet containing essential metadata (in this example, %\Patient, Batch_no, Tine, and Tissue) % %**********************************</pre>	•	
 Editt OP 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 		<pre>\Bioimaging_Staff\George\2023315_AutoDECRA\2_image_extraction\OPTIMAL_Pipeline_Part_1.m X</pre>	6	
 Editt OP 255 256 257 258 259 260 261 262 263 266 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 		<pre>\Bioimaging_Staft/George\20230315_AutoDECRA\2_image_extraction\OPTIMAL_Pipeline_Part_1.m L_Pipeline_Part_1m X OPTIMAL_Pipeline_Part_1.m X + %\Load the spreadsheet containing essential metadata (in this example,</pre>	•	
 Editt OP 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 273 274 275 276 277 278 279 280 		<pre>\Bioimaging_Stafk/George\20230315_AutoDECRA\2_image_extraction\OPTIMAL_Pipeline_Part_1.m X</pre>	6	
 Editt. OP 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 		<pre>\Bioimaging_Staft\George\20230315_AutoDECRA\2_image_extraction\OPTIMAL_Pipeline_Part_1.m X L_Pipeline_Part_1.m X OPTIMAL_Pipeline_Part_1.m X + Xkload the spreadsheet containing essential metadata (in this example, XkPatient, Batch_no, Time, and Tissue) Xkwowset = adtable('Tonsil_MAV22.xlsx'); Patient_name_list=Spreadsheet.Patient; Batch_number_list=Spreadsheet.Batch_no; Patient_identifier=unique(Patient_name_list); for loop=1:numel(Patient_identifier)</pre>	6	
 Editt OP 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 273 274 275 276 277 278 279 280 281 282 		<pre>\Bioimaging_Staff.George\20230315_AutoDECRA\2_image_extraction\OPTIMAL_Pipeline_Part_1.m LPipeline_Part_1.m × OPTIMAL_Pipeline_Part_1.m × + %Koad the spreadsheet containing essential metadata (in this example, %%Patient, Batch_no, Time, and Tissue) %%Some spreadsheet-norsil_MAY22.xlsx'); Patient_name_list=Spreadsheet.Patient; Batch_number_list=Spreadsheet.Patient; Batch_number_list=Spreadsheet.Batch_no; Patient_identifier=unique(Patient_name_list); for loop2=1:numel(Patient_name_list); for loop2=1:numel(Patient_name_list); for loop2=1:numel(Patient_name_list); for loop2=1:numel(Patient_name_list); for loop2=1:numel[Patient_name_list); for loop2=1:numel[Patient_name_list(loop),Patient_identifier(loop2)) ==1 Patient_number_list(loop)=loop2; end end condition_list=Spreadsheet.Tissue; %%Xincorporates the tissue type into the stored data (in this case Tonsil %%Should represent all of the images, with no Syn images %************************************</pre>	6	
 Editt OP 255 256 257 258 259 260 261 262 263 266 267 268 269 270 273 274 275 276 277 278 279 280 281 282 283 		<pre>\Bioimaging_Staft\George\20230315_AutoDECRA\2 image_extraction\OPTIMAL_Pipeline_Part_I.m LPipeline_Part_I.m \vee OPTIMAL_Pipeline_Part_I.m \vee + \vee Value the spreadsheet containing essential metadata (in this example, \vee XPatient, Batch_no, Time, and Tissue) \vee XVXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX</pre>		

6. Incorporate the specific metadata necessary for inclusion within the fcs files later. Make sure this information is purely numeric within your files, otherwise issues with fcs file generation

may occur

Z Editor - Y	(\Bioimaging_Staff\George\20230315_AutoDECRA\2_image_extraction\OPTIMAL_Pipeline_Part_1.m		×
OPTIMA	AL_Pipeline_Part_1.m × OPTIMAL_Pipeline_Part_1.m × OPTIMAL_Pipeline_Part_1.m × +		
282	<pre>test=find(Tissue_index==0); %should be 0 i.e. all ROI allocated a</pre>		A
283	if test==0		
284	error('Not all ROIs given a tissue type')		
285	end		
286			
287			-
288			=
289 📮	%%Incorporates the time type into the stored data (in this case half should		
290	%%be START and half should be END)		
291 ^L			
292	time_list <mark>=</mark> Spreadsheet.Time		
293	<pre>Start_ROI_Index = find(contains(time_list,'START'));</pre>		
294	<pre>End_ROI_Index = find(contains(time_list,'END'));</pre>		
295	<pre>check_number=numel(Start_ROI_Index)+numel(End_ROI_Index);</pre>		-
296	Time_index=zeros(numel(time_list),1);		-
297	Time_index(Start_ROI_Index)=1;		
298	Time_index(End_ROI_Index)=2;		-
299	<pre>test=find(Time_index==0); %should be 0 i.e. all ROI allocated a</pre>		
300	if test==0		-
301	error('Not all ROIs given a time point')		-
302	end		
303			
304			
305			-
306 🖵	%%Gets the list of channels from the images based on the metadata inherent		
307	%%within each Tiff file		=
308 -	- *************************************		
309	current_ROI_directory=ROI_tolder{1};		-
310	<pre>current_R0I=R0I_tilenames{1};</pre>	-	
311	current ROI filename and directorv=[current ROI directorv.'\'.current ROI]:		
		,	

7. Select the channels of interest, i.e. the channels that have been specifically stained and imaged during your experiment (note here that again the first channel is channel 1 etc.)

💋 Editor - Y	\Bioimaging_Staff\George\20230315_AutoDECRA\2_image_extraction\OPTIMAL_Pipeline_Part_1.m	C) ×	
OPTIMA	L_Pipeline_Part_1.m 🗙 OPTIMAL_Pipeline_Part_1.m 🗙 OPTIMAL_Pipeline_Part_1.m 🗶 🕇			
310	current_NOL_OITectory=NOL_TOIDer{1};		· 🔒	1
310	current_KOI=KUI_TILenames{1};			
311	current_KOI_TILEname_and_cirectory=[current_KOI_cirectory, \ ,current_KOI];			
212	foritef()			
214				l
215	principarine (), space listerat channel press (support DOI directory support DOI print on pat);			1
316	check_chamiei_hames_list-get_chamies(current_koi_urrento), current_koi, print_or, print_or,			•
317				
318 🗐	Wallows the user to then define which channels actually contain a label of			
319	Stinteret modify this list has don your our junctures			
320	required channel numbers=[7,9,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,			
321	31.32.33.34.35.36.37.38.39.40.41.42.43.44.45.46.47.48.49.53.54.57]:			ł
322				
323 -	%%Prints out the channels selected along with their name according to the			١
324	%%metadata			1
325 L	XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX			
326	choosen channel names=check channel names list(required channel numbers);			
327	<pre>fprintf('CHOOSEN CHANNEL CHECKLIST\r\r')</pre>			
328 📮	<pre>for loop=1:numel(required_channel_numbers)</pre>			
329	<pre>fprintf('Channel Number = %i : Label = %s \r',loop,choosen_channel_hames{loop})</pre>			
330 L	end			
331	<pre>fprintf('\r')</pre>			
332			-	
333				
334				
335	8%		ΙΞ	i
336	<pre>fprintf('BUILDING CELL LIST FOR ANALYSIS\r')</pre>			ļ
337	fprintf('\r\r')			
338	count=0;	*		
	4			

8. Adapt the fcs conversion portion of the code to include the metadata specified earlier, making sure to adjust the headers as well

💋 Editor -	Y:\Bioimaging_Staff\George\20230315_AutoDECRA\2_image_extraction\OPTIMAL_Pipeline_Part_1.m	🖻 🗙
OPTIM	IAL_Pipeline_Part_1.m 🗙 OPTIMAL_Pipeline_Part_1.m 🗙 OPTIMAL_Pipeline_Part_1.m 🛪 🕇	
411	%%Add patient/ROI metadata to single cell file list	A
412	- *************************************	
413	number_of_cells=size(ONE_ROI_DATA,1);	
414	<pre>all_ROI_list(all_ROI_loop).number_of_cells=number_of_cells;</pre>	
415	<pre>Single_cell_ROI_number=repelem(all_ROI_loop,number_of_cells)';</pre>	
416	<pre>Single_cell_Patient_number=repelem(Patient_number_list(all_ROI_loop),number_of_cells)';</pre>	
417	<pre>Single_cell_File=repelem(Patient_name_list(all_ROI_loop),number_of_cells)';</pre>	
418	<pre>Single_cell_Pathology_index=repelem(Tissue_index(all_ROI_loop),number_off_cells)';</pre>	
419	<pre>Single_cell_batch_number=repelem(batch_number,number_of_cells)';</pre>	
420	FCS_ROI_DATA=[AS_DATA, AS_NORM_DATA, EXTRA_DATA, Single_cell_ROI_number, Single_cell_Patient_number, Single_cell_Patholo	
421	if count==1	
422	COMPLETE_COMP_DATA_SET=[AS_DATA, AS_NORM_DATA, EXTRA_DATA, Single_cell_ROI_number, Single_cell_Patient_number, Single	
423	else	
424	COMPLETE_COMP_DATA_SET=[COMPLETE_COMP_DATA_SET; [AS_DATA, AS_NORM_DATA, EXTRA_DATA, Single_cell_ROI_number, Single_ce	
425	end	
426		
427		
428		
429 -	%%Creates the fcs file data, and saves as fcs to the appropriate folder	
430	%%using the appropriate file name	
431	- *************************************	
432	<pre>fcs_name_all=all_ROI_list(all_ROI_loop).channels_filename;</pre>	
433	[filepath,filename,ext] = fileparts(fcs_name_all);	
434	<pre>fcs_name=[fcs_directory,filename,'.fcs'];</pre>	
435	fcs_channel_list=[marker_channels,norm_marker_channels,extra_markers , {'ROI_number'},{'Patient_number'},{'Pathology_ind	
436	fprintf('\r')	
437	fprintf('Writing %s \r',filename)	
438	fca_writefcs(fcs_name,FCS_ROI_DATA,fcs_channel_list, fcs_channel_list)	
439		
440		

9. Repeat this process for the "collated fcs" file at the very end of the macro

💋 Eo	litor - `	f:\Bioimaging_Staff\George\20230315_AutoDECRA\2_image_extraction\OPTIMAL_Pipeline_Part_1.m	×
	OPTIM	AL_Pipeline_Part_1.m 🛪 OPTIMAL_Pipeline_Part_1.m 🛪 OPTIMAL_Pipeline_Part_1.m 🛪 🕂	
43	1	***************************************	A
43	2	<pre>fcs_name_all=all_ROI_list(all_ROI_loop).channels_filename;</pre>	
43	3	[filepath,filename,ext] = fileparts(fcs_name_all);	
434	4	<pre>fcs_name=[fcs_directory,filename,'.fcs'];</pre>	
43	5	fcs_channel_list=[marker_channels,norm_marker_channels,extra_markers , {'ROI_number'},{'Patient_number'},{'Pathology_ind	
43	5	fprintf('\r')	-
43	7	<pre>fprintf('Writing %s \r',filename)</pre>	_
43	в	<pre>fca_writefcs(fcs_name,FCS_ROI_DATA,fcs_channel_list, fcs_channel_list)</pre>	
43	э		
44	3		
44	1 ¹	- end	
44	2		
44	3		-
44	4 Ę	%%Creates one final fcs file containing all the data processed within this	-
44	5	%%pipeline	-
44	5 l		
44	7	<pre>fprintf('\r')</pre>	-
44	В	<pre>filename='Combined_dataset';</pre>	
44	Э	<pre>fcs_name=[fcs_directory,filename,'.fcs'];</pre>	-
45	3	<pre>fprintf('Writing %s \r',filename)</pre>	-
45	1	<pre>fcs_channel_list=[marker_channels,norm_marker_channels,extra_markers , {'R0I_number'},{'Patient_number'},'Pathology_index'}</pre>	
45	2	<pre>fca_writefcs(fcs_name,COMPLETE_COMP_DATA_SET,fcs_channel_list, fcs_channel_list)</pre>	
45	3		
45	4		-
45	5	<pre>fprintf('\r\r')</pre>	
45	5		=
45	7	%%End of Macro	=
45	8		
45	9		
46	3		

10. Run the script. There is a manual input stage following CellProfiler calling to check that the script has been effective before attempting to continue. If CellProfiler has failed to run, it is possible to run CellProfiler manually according to the instructions in the following section and to then press "Enter" on the keypad back in Matlab to continue the script.

CellProfiler Manual Use

If CellProfiler does not run, you may manually open the CellProfiler software from outside Matlab and run according to the instructions below

 Drag all image folders from the folder "4_structured_data_for_CP" into the images window. Note, this system is set up for 60 channel images, and modifications will need to be made if a different number of channels are actually present.

le Edit Test Windows Help		
con ret wholes rep (c) test wholes rep (c) test (c) t	Picesing or project, use the Image: module to compile a list of files and or folders that you want to analyze. You can also specify a set of rules to include only the desired files in your selected folders. Picesing or project, use the Image: module to compile a list of files and or folders that you want to analyze. You can also specify a set of rules to include only the desired files in the project of the pr	
Output Settings View Workspace ? Adjust modules: II Pause Stop Analysis		
	Processing: 1 of 22 image sets completed Running 12 workers.	1/801

2. Modify the save locations for all "SaveImages" modules, along with the "ExportToSpreadsheet" module if you plan on using that data.

	×
G Images G Metadata G Metadata G Metadata G RescaleIntensity G RescaleIntensity G RescaleIntensity G RescaleIntensity G RescaleIntensity G Isomethy G	
Image: Constraining of the Recalibilitation of the Recalibility o	^
	2
US GrayloLolor Gravelandutines Enter file arefit Exempetation checker	?
Destry Objects Number of dials 3	?
O Svetmags Svetmags Svetmags Svetmags Svetmags Svetmags Svetmags Svetmags	?
Converdbjectsformage Converdbjectsfor	
O MessureObjectSteStAppe Overwrite existing files without warning? O MessureGranularity O MessureTexture Uters to any	1
O Messure/Operationed/Database Page 1 (Dery You ▼) Constrained and (Dery You ▼) Deproforgreadheet Record the file and path information to the saved image? Ver ⊛ No	?
Outrust Settion: View Worksoner#	
Adjust modules: • • • • • •	

3. When ready to run the pipeline, simply press "Analyse Images" and wait for the pipeline to finish before returning to Matlab

Clustering Data Excel

 Structure your clustering data according to the image, with the original cluster assignment in column "cSOM_s" and the new cluster assignment in column "Tier2ClusterAssignment" or "Tier1ClusterAssignment" depending on the clustering tier. Put the name of the cluster in column "Name".

AutoSarve Corr 🗄 🏷 - C Cluster_Assignments_and_Merges_T	DNSILadar - Excel	earch								George Mere	ಜ 🚳 🖽		
File Home Insert Page Layout Formulas Data Review View A	utomate Help										🖓 Com	ments 🖻	Share ~
Cut Calibri 11 → A* A* = = =	Vrap Text General ~		Not	rmal	Bad	Good) 📼 🗰	∑ AutoSum ~	AN O			
Paste Copy -		Conditional Fo	rmat as Nor	utral	Calculation	Check Cell	- Insert	Delete Format	😨 Fill 🗸	Sort & Find &	inahze Sensiti		
✓ Ø Format Painter B 1 0 * □ * ○ * ▲ * = = = □ * □ * □	Merge & Center ~ 100 ~ 70 7 100 -20	Formatting ~ T	able ~	0001		chieda cen		· ·	🞸 Clear 🐃	ilter ~ Select ~	Data		
Clipboard IS Font IS Alignment	rs Number rs			Style	5			Cells	Editir	ig j	nalysis Sensit	vity	~
D61 ▼ I × √ fr													v
A 8	c	D	E F	G	н		к	L M	N	ΟΙΡΙ	0	a I s	1.19
Tier2ClusterAssignment Name	cSOM s											_	_
1 Epithelium	1												
2 Memory CD8 T cells	2												
2 Memory CD8 T cells	19												
3 M2 Macrophages	3												
3 M2 Macrophages	29												
4 B cells	5												
4 B cells	10												
5 Memory CD4 T cells	6												
0 5 Memory CD4 T cells	9												
1 6 Epithelium (proliferating)	7												
2 7 Folicular B cells (proliferating)	8												
3 7 Folicular 8 cells (proliferating)	23												
4 8 Mature Macrophages	11												
5 9 Folicular T cells	12												
6 10 Plasma cells	13												
7 11 Endothelium	14												
8 11 Endothelium	16												
9 11 Endothelium	24												
0 12 Unclassified	15												
1 12 Unclassified	4												
2 13 Effector CD4 T cells	17												
3 14 Epithelium and Immune cells	18												
4 15 Naïve CD8 T cells	20												
5 16 anti-Inflammatory Macrophages	21												_
6 16 anti-Inflammatory Macrophages	25												_
7 17 CD4 T cells	22												
8 18 STING+ cells	26												
9 19 Apoptotic cells	27												
20 Germinal center Macrophages	28												
1 21 Standard Macrophages	30												
2													
3													
4													

Matlab Macro II

1. Following cluster analysis, it's time to run the Matlab Macro

"OPTIMAL_Pipeline_Part_2_Heatmaps.m". Update the name for the csv file containing details about each cell identified in the process, along with the folder path names for the Mask outputs from CellProfiler, along with where you would like output heatmaps to be saved.

HOME PLOTS APPS EDITOR	AVRISH VEW	83	4 🖁 9 6 🕞 🛙	💿 Search Documentation 🛛 👂 🌲 Sign In
	96 🐲 💬 📩 💌 🖾 Section Break 📄 🙆			
New Open Store - Compare - Go To Q Find - Pater	E the Let During the second strange and the second strange			
Bookmark	A water and a section of the section			
FILE NAVIGATE	CODE ANALYZE SECTION RUN			
= 🔶 💽 🛃 💭 📙 + Y: + Bioimaging_Staff + George + 20230315	AutoDECRA > 2_image_extraction >			م •
Jurrent Folder	Editor - H1:Downloads/20230306_Tonsil_Matlab_Codes_V2\2_image_extraction\OPTIMAL_Pipeline_Part_2_Heatmaps.m	⊙×	Workspace	•
Name +	OPTIMAL_Pipeline_Part_1.m × OPTIMAL_Pipeline_Part_2_Heatmaps.m × +		Name A	Value
- Functions	1 😳 %%This macro serves as part 2 of the image processing pipeline for the	^ A	all_ROL_list	tx27 struct
Outputs	2 SXpublication named "OPTIMAL: An OPTImised Imaging Hass cytometry Analysis		all_ROL_loop	27
-SAUDODECKA_EPCAW_COUNTER_V3X8X	A Strategier to Segmentation and back exponential, Please review the		ans	1
AutoDECRA_Data_Analysis_Master.xlsx	5 XXmacro as necessary	-	AS NORM DATA	19073x31 double
AutoDECRA_EPCAM_Clusters_V2.csv	6	_	batch_number	1
AutoDECRA_EPCAM_Clusters_V3.csv	7 Stelears the command window and resets the Workspace to ensure previous		Batch_number_list	27x1 double
AUTODECRA_EPCAM_CLUSTERS_V4.CSV AutoDECRA_EPCAM_CLUSTERS_V4.CSV	a Astuns do not interfere with the macro running clean		channel_counter	'060'
AutoDECRA_EPCAM_Clusters_V5.csv	10 close all		channel_list	63
AutoDECRA_EPCAM_Clusters_V5.xlsx	11 clc		O check channel na	1x60 cell
AutoDECRA_Graphing_Outputs.zip	12		🔠 check_number	0
AutoDECRA_March23.xlsx	13 - Wedde the "Functions" folder to the path, allowing for use of the	-	choosen_channel	tx31 cell
Cluster Assignments and Merges AutoDECRA.xlsx	15 addpath('./Functions/')		co_tactor	1/21.007
Cluster_Assignments_and_Merges_AutoDECRA_V2x.xlsx	16		COMP DATA	19073x31 double
Final_lung_clustered_output.csv	17 🖸 3RUser-Defined Directories and Files		comp_matrix	31x32 table
OPTIMAL_Pipeline_Part_1_V2.m	18 SXPlease adapt these folder names as necessary to work with your folder 20 SXPLease for under device. The ambiention of each folder folder is inter- ation.		compensated_sin	31x1 double
OPTIMAL Pipeline Part 2 Heatmaps.asy	20 Skinedistely prior to the folder being named		COMPLETE COM	31x31 double 20000-72 double
OPTIMAL_Pipeline_Part_2_Heatmaps.m	21	_	Complete Communication	219090479 00001E
OPTIMAL_Pipeline_Part_2_Heatmaps_MultiFactor_Plotting.m	22 📮 %%Cluster Output File for the single cell information including consensus		eount count	27
OPTIMAL_Pipeline_Part_3_Visualisation.m	23 Sicluster identification, ROI number, pathology index etc.		counter	'001'
spilover2.1.xisx	24 Tiename-'onsil_Prod_Custers_v2.csv'; 25 Finaltoni/Loutersdoutnut = imperfile(filename);		input	-i C:/Users/george/MATLAB_Macros/20230315_Auto
- spinorenzi - copyxiax	26 3XDefines the mask working directory		cp_output	o C:/Users/george/MAILAB_Macros/20230315_Auto
	27 Mask_working_directory='C:\Users\george\MTLAB_Macros\20230306_Tonsil_Matlab_Codes_V2\5_CP_pipeline\OUTPUTS\CellMasks\';		cp project	
	28 Tablefines the output folder for where to save the heatmaps generated in		cp_settings	'reir'
	29 * Skitis code autoriticaldes = 'Cultineer/assessitif' AB Marses/20120206 Tangil Hatlah Codes (1)2 (asses autoritical)(atouts)'		current_mask	'CellMasks027.npy'
	31		current_mask_dire	"Y/\Bioimaging_Staff\George\20230315_AutoDECRA\5
	32	_	un current BOI	"SPI Tonsil 28062021 ROI004 ROI 004 ome tiff"
	33		current_ROI_direct	Y/\Bioimaging_Staff\George\20230315_AutoDECRA\1
	34 - XExtracts the cluster numbers and e.g. pathology numbers from the file	Ψ =	current_ROI_filena	'Y:\Bioimaging_Staff\George\20230315_AutoDECRA\1
	23 4	F	display_rescaled_d	1
	Command Window	۲	EPCAM channel	29
	New to MATLAB? See resources for Getting Started.	×	EPCAM_max	4
	TARGET CHANNEL NUMBER = 29 : LABEL = CD03(Lu175D1)	^	EPCAM_min	0
	READING DATA FROM		ext ext	1.6fF
	ORIGINAL CHANNEL NUMBER = 49 : LABEL = CD45(Yb176D1)		EXTRA_DATA	150/3i0 double
	TARGET CHANNEL NUMBER = 30 : LABEL = CD45(Yb176D1)		1 fcs_channel_list	1x73 cell
utoDECRA_MAY22.xlsx (Microsoft Excel Worksheet)	READING DATA FROM ORIGINAL (FROMPEL NUMBER = 53 : LEREL = DNL2(IY)93D()		fcs_directory	'Y:/Bioimaging_Staff/George/20230315_AutoDECRA/fi
	TARGET CHANNEL NUMBER = 31 : LABEL = DNA2(II193D1)		fcs_name	'Y:/Bioimaging_Staff/George/20230315_AutoDECRA/fe
			ECS ROL DATA	SPJ_ronsil_zoubzuz1_ikuruu4_K01_004.ome.bff* 19073v73.double
	Writing SPJ_Tonsil_28062021_ROIO04_ROI_004.cmm		filename	'Combined_dataset'
No details available			dieName	'CMW_Baseline_28062021_ROI001_ROI_001'
	writing compined_dataset		diepath filepath	-
		- 1	ilastic_data	' Y:/Bioimaging_Statt/George/20230315_AutoDECRA/4
	h >>	~	ilastik_model	'=Y:/Bioimaging_Staff/George/20230315_AutoDECRA/ Y
		>	<	
1.	Zoom: 10%. UTF-8	0	RI F scrint	In 11 Col 4

 Update the name for the clustering information file that you have for your data, including the names of relevant sheets for specific clustering tiers if necessary. Make sure to adapt the "TIER_TO_USE" parameter to the relevant tier for your data

 Montphili Neocea - academic use 					5 A
HOME PLOTS APPS EDITOR			6 8	h 🛍 Ə Ə 🖨 🤅) 💿 Search Documentation 🛛 🔎 🐥 Sign In 🛛
	94 94 25	Section Break			
记 🛄 🖾 Compare 🗸 🔛 🏹 📜 🖓	Profiler				
New Open Save Aprint - Go To Contract Refacts	or E 12 Analyze	Run Tag kun and Advance Run Step Stop			
• • • Bookmark • •	<u>-</u>	Section 📓 Run to End 🔹			
FILE NAVIGATE	CODE ANALYZE	SECTION RUN			Ā
🔹 🔶 💽 🔂 🛜 📒 🔸 Y: 🔸 Bioimaging_Staff 🔸 George 🔸 20230315_	AutoDECRA + 2_image_extract	ion F			- ₽.
urrent Folder 💿	Editor - H:\Downloads\2	230306_Tonsil_Matlab_Codes_V2\2_image_extraction\0PTIMAL_Pipeline_Part_2_Heatmaps.m	• ×	Workspace	۲
□ Name ↔	OPTIMAL Pipeline Part	1.m × OPTIMAL Pipeline Part 1.m × OPTIMAL Pipeline Part 2 Heatmaps.m × +		Name e	Value
Functions	54 - XXXXXXXX		· A	all POI Gat	to 27 stored
Outputs	55 TIER TO	JSE=2;		all ROLloop	27
-SAutoDECRA_EPCAM_Clusters_V5.xlsx	56 if TIER	'0_USE==1		ans	1
AutoDECRA_MAY22.xisx	57 % Re	id tier 1 info		AS_DATA	19073x31 double
AutoDECRA_Data_Analysis_Master.xlsx	58 new_	<pre>:luster_table= readtable('Cluster_assignments_and_Merges_TONSIL.xlsk','Sheet','TIER 1');</pre>	_	AS_NORM_DATA	19073x31 double
AutoDECRA_EPCAM_Clusters_V2.csv	59 Tier	1 name from sheet-new_cluster_table.Vars;		batch_number	1
AutoDECKA_EPCAM_Clusters_V3.csv	61 Tier	1_cson=new_closter_table.vars; 1_assignments=new_cluster_table.Vard;		Batch_number_list	27x1 double
AutoDECRA EPCAM Clusters V5 COPY PRE ALTERATION.csv	62 E for	cop=1:numel(Tier 1 assignments)		channel_counter	1900 million
AutoDECRA_EPCAM_Clusters_V5.csv	63	/printf('Tier 1 assignment = %i : Original cSOM = %i \r', Tier 1 assignments(loop), Tier 1 cSOM(/	loop))	channel loon	63
AutoDECRA_EPCAM_Clusters_V5.xisx	64	<pre>/printf('%s \r',Tier_1_name_from_sheet(loop));</pre>		theck channel na.	1x60 cell
AutoDECRA_Graphing_Outputs.zip	65 end			deck_number	0
AutoDECRA_March23.xlsx	66 % Ge	ierate new cluster list	_	Choosen_channel	1/31 cell
AutoDECRA_MAY22xisx	67 numb	<pre>.r_of_Tier_1_cSOM=numel(unique(Tier_1_CSOM));</pre>		co_factor	1
Cluster_Assignments_and_Merges_AutoDECRA.xisx	68 [lier	1_new_cluster_number=zeros(numer(cluster_number),1);		comp_channels	1x31 cell
Rinal lung clustered output csv	70	Juster indicesting (cluster number=Tier 1 cSOM(loon));		COMP_DATA	1907sid1 double
OPTIMAL_Pipeline_Part_1_V2.m	71	fier 1 new cluster number(cluster indices)=Tier 1 assignments(loop);		companyated sin	31x32 table
OPTIMAL_Pipeline_Part_1.m	72 end			compensation m	31x31 double
OPTIMAL_Pipeline_Part_2_Heatmaps.asv	73 Tier	1_new_cluster_number(Tier_1_new_cluster_number==0)=NaN;	_	COMPLETE_COM	219593x73 double
OPTIMAL_Pipeline_Part_2_Heatmaps.m	74 Tier	1_assignments_indices=unique(Tier_1_assignments);		condition_list	27x1 cell
OPTIMAL_Pipeline_Part_2_Heatmaps_MultiFactor_Plotting.m	75 Tier	l_names=unique(Tier_l_name_from_sheet,'stable');		di count	27
CPTIMAL_Pipeline_Part_3_visualisation.m	76 3 50	.p Justen numbervillen 1 neu clusten numberv		counter	'001'
spillover2.1 - Conv.visv	78 064	luster_number=lier_l_new_cluster_number;		cp_input	-i C:/Users/george/MATLAB_Macros/20230315_Auto
40 4 ·····	79 else			cp_output	Colleast/george/MATLAB_Macros/20250515_Addt
	80 % Re	ud tier 2 info		in co project	- in Cr/lisers/neoroe/MATLAB_Macros/20230315_Autr
	81 new_	:luster_table= readtable('Cluster_assignments_and_Merges_TONSIL.xlsk','Sheet','TIER 2');		cp_settings	1-c-r
	82 Tier	_2_name_from_sheet=new_cluster_table.Name;		current_mask	'CellMasks027.npy'
	83 Tier	<pre>2_cSOM=new_cluster_table.cSOM_s;</pre>		current_mask_dire	'Y:\Bioimaging_Staff\George\20230315_AutoDECRA\5
	as D for	<pre>assignments=new_cluster_table.ller2clusterAssignment;</pre>	-	current_mask_file	'Y:\Bioimaging_Staff\George\20230315_AutoDECRA\5
	86	opinif('Tier 2 assignment = %i : Original cSOM = %i \r'.Tier 2 assignments(loon).Tier 2 cSOM('	1000))	current_KUI	SPJ_Tonsil_28062021_R01004_R01_004.ome.bff
	87	printf('%s \r',Tier_2_name_from_sheet{loop});		current ROI filena.	Y/Bioimaging_staff/George/20230315_AutoDECRA(1 V/Bioimaging_Staff/George/20230315_AutoDECRA(1
	88 L			display rescaled d	1
	Common different and			End_ROI_Index	0
	Command Window		•	EPCAM_channel	29
	New to MATLAB? See resou	ces for <u>Getting Started</u> ,	*	EPCAM_max	4
	TARGET CHANNEL	NUMBER = 29 : LABEL = CD03(Lu175D1)	^	EPCAM_min	0
	READING DATA FROM	(EXTRA DATA	1977b5 double
	ORIGINAL CHANNEL	NUHBER = 49 : LABEL = CD45(Yb176D1)		il extra markers	16 cell
	TARGET CHANNEL	NUMBER = 30 : LABEL = CD45(Yb176D1)		11 fcs_channel_list	1x73 cell
utoDECRA_MAY22.xlsx (Microsoft Excel Worksheet) V	ODICING DATA PRO			fcs_directory	'Y:/Bioimaging_Staff/George/20230315_AutoDECRA/fi
	TABORT CHAMPEL	NUMBER - 03 : LABEL - DWAR(LIIISDL)		fcs_name	'Y:/Bioimaging_Staff/George/20230315_AutoDECRA/fi
	TARGET CHARTER	NONDER - ST. LEDEL - DERELITISDES		fcs_name_all	'SPJ_Tonsil_28062021_ROI004_ROI_004.ome.tiff'
	Writing SPJ Tons	1 28062021 ROI004 ROI 004.cme		FCS_ROLDATA	190/3x/3 double
				filehame	Comuneo_ustaset (CMW Receipe 20052021 POI001 POI 001)
No details available	Writing Combined	dataset		incivanie incivanie	
				ilastik data	'Y:/Bioimaging Staff/George/20230315 AutoDECRA/4
				ilastik instruction	'headlessproject'
	<i>f</i> ₹ >>		~	ilastik_model	'=Y:/Bioimaging_Staff/George/20230315_AutoDECRA/ *
	<		>	<	>
II.a.		Zoom 10	0% UTF-8 C	RLF scrint	In 11 Col 4

3. Adapt the settings for the heatmap generation in the section shown below. The key parameters are "pVal_sig" which sets the threshold for what proportion of the simulation images need to show positive, neutral, or negative interactions relative to the original dataset to be determined as a significant positive, neutral, or negative interaction.

"cut_off_percent" determines the proportion of images a cell type needs to be present in to be counted within the analysis. "permutations" refers to the number of simulation images generated. This is important in conjunction with the "pVal_sig" value, as too few permutations with a high significance value could lead to aberrant results. Finally "number_of_pixels" refers to the number of pixels a cell mask is expanded by to determine neighbouring cells, the smaller this number the smaller the region the macro will assess to identify neighbours. This number should be adjusted based on cell density and cell size.



4. Create heatmaps based on clinical/metadata information by using "for" loops: If you have metadata information that you wish to group your data by for heatmap generation, e.g. for comparison between pathologies, you can use this data as long as it is present within your single cell csv file.



5. Set a size gate on the cells to be analysed by adapting the lower and higher area gate parameters

📣 MATLAB R2022a - academic use											- a ×
HOME PLOTS APPS EDITOR	PUBLISH VIEW								16	4 🛍 N & 🛱 🕄	🗴 🛞 Search Documentation 🛛 🔎 🚑 Sign In
👍 🐚 🗐 🗠 ann a 🔄 🌳 🖉	£ % % ?	🔯 🔚 Section Break		6							
New Onen Save the Line Go To Q Find - Refu		Run and Advance	Run	Sten	Ston						
New open save print	• 🖌 •	Section 📸 Run to End	-	Step	Stop						
FILE NAVIGATE	CODE ANALYZE	SECTION		RUN							Ā
💠 🔶 💽 😹 🔀 📙 🕨 Y: 🔸 Bioimaging_Staff 🔸 George 🔸 2023031	15_AutoDECRA + 2_image_extra	ction +									م •
Current Folder		20230306_Tonsil_Matlab_Codes_V2\'	2_image_e	axtraction\C	ортима	L_Pipeline_Part_2_Heatmaps.r	n	(×	Workspace	۲
Name -	OPTIMAL_Pipeline_Pat	rt_1.m × OPTIMAL_Pipeline_Pa	rt_1.m ×	OPTIME	AL_Pipe	aline_Part_2_Heatmaps.m 🛛	+			Name +	Value
Functions	224	"crusters] to_prot-new_crus	Cer_mans	der (cause		marces_co_proc/,			-	all ROI list	bi27 struct
IE _ Outputs	226	and	CONTRACT.	CREASES AND C.	6333					all_ROLloop	27
-SAutoDECRA_EPCAM_Clusters_V5.xlsx	227									ans ans	1
AutoDECRA_MAY22XISX AutoDECRA_Data Analysis MasterxIsx	228	%%Applies distinguishable	: colors	function	in to '	the clusters to ensure	2		-	AS_DATA	19073x31 double
AutoDECRA_EPCAM_Clusters_V2.csv	229	Silvese of visualisation	colors/		af 11	(tent)				batch number	1
AutoDECRA_EPCAM_Clusters_V3.csv	230	colors = distrigoramente_	CO10+ 8(+	Jumper _or	T_C10	sters);				Batch_number_list	27x1 double
AutoDECRA_EPCAM_Clusters_V4.csv	232									channel_counter	1060'
AutoDECRA_EPCAM_Clusters_V5_COPY_PRE_ALTERATION/Cov	233	%%Finds all edge cells in	the in-	mage, for	reno	val				Channel_list	fx36 cell
AutoDECRA_EPCAM_Consets_v3.csv	234	[cells_on_the_edge_of_FOV	<pre>[] = Find</pre>	.d_Edge_Cr	.ells(IMAGE_MASK);				channel_loop	63 1.40 mil
AutoDECRA_Graphing_Outputs.zip	235	WARRIES upper and lower				11 anan dafinas hu				check_channel_na	0
AutoDECRA_March23.xlsx	230	Sithe user, selecting cel	gate	attings -	to ce. o smal	11 area derines by 1 or too big for				Choosen_channel	hd1 cell
AutoDECRA_MAY22.xlsx	238	%%removal		010 111	-	1 01 000 010				a co_factor	1
Cluster_Assignments_and_Merges_AutoDECRA_Visx	239 -	*****	ATTAX		aax					comp_channels	b31 cell
Final lung clustered output csv	240	lower_area_gate=20;								COMP_DATA	19073x31 double
OPTIMAL_Pipeline_Part_1_V2.m	241	upper_area_gate=200;	- keen	1011 00	- 14	3				compensated sin	37x32 toble Stat double
OPTIMAL Pipeline_Part_1.m	242	[cells_to_remove, cells_on_gated	o_keep,	COLL_MAN	ea_iii 5K. lov	stj =	a sate):			compensation_m	31x31 double
OPTIMAL_Pipeline_Part_2_Heatmaps.asv	244	munnunnunnun	anner.	anna.	6333	ci_o.co_80cc,			-	COMPLETE_COM	219593x73 double
OPTIMAL_Pipeline_Part_2_Heatmaps.m OPTIMAL_Pipeline_Part_2_Heatmaps.m	245									Condition_list	27x1 cell
OPTIMAL_Pipeline_Part_2_Heatmaps_Multi-actor_moungan OPTIMAL_Pipeline_Part_3_Visualization m	246 🛱	%DMerges the list of cell	.s on edg	ige and th	the wro	ong size and removes				count	27
spilover2.1.xixx	247	%%these from analysis for	heatmap	pping						counter	1001
spillover2.1 - Copy.xisx	248	magned calls to canouasts	alls on	the edd	AXA of	ECV cells to serve'	1.	1		cp output	- o C:/Users/george/MATLAB Macros/20230315 Auto
	250	fmatr.adileneighlist cree	ation ga	ated(INA0	AR MAS	K.number of pixels,mer	read calls to remove);			cp_path	'C:/Users/george/MATLA8_Macros/CellProfiler/CellPr
	251	in manual manual second	ALLETT.	ATTAXAT	an					cp_project	-p C:/Users/george/MATLAB_Macros/20230315_Auto
	252									cp_settings	inc. rf
	253	NV Assertance de la seconda de		-		of 1000 1073				current_mask dire	CellMasks027.npy Wilkinimaning Staffi Ganroal 20230315 AutoDECRAIS
	254	33 Assigns to a row or a r	aatrix x	the neigr	ABOTS	of each cell				current_mask_file	Y:\Bioimaging_staff\George\20230315_AutoDECRA\5
	256	ag neigh=horzcat(matr);			Alex					current_ROI	SPJ_Tonsil_28062021_R01004_R0I_004.ome.tiff
	257	clusteri=zeros(length(ag	neigh),	,max(cell'	fun('	<pre>size',ag_neigh,1)));</pre>				current_ROI_direct	Y:\Bioimaging_Staff\George\20230315_AutoDECRA\1
	258 🛱	<pre>for i=1:length(ag neigh)</pre>							-	current_ROI_filena	Y:\Bioimaging_Staff\George\20230315_AutoDECRA\1
								,		End ROL Index	
	Command Window								۲	EPCAM_channel	29
	New to MATLAB? See reso	urces for Getting Started.							×	EPCAM_max	4
	TARGET CHANNE!	NUMBER = 29 : LABEL = -	CD83 (Lr	a175D1)					^	EPCAM_min	0
	READING DATA FRO	M								EXTRA DATA	160775-5-4
	ORIGINAL CHANNEL	, NUMBER = 49 : LABEL = 4	CD45 (Yb	J176D1)						EXTRA_DATA	1907alo double
	TARGET CHANNEL	. NUMBER = 30 : LABEL = (CD45 (Yb	J176D1)						1 fcs_channel_list	1x73 cell
AutoDECRA_MAY22.xlsx (Microsoft Excel Worksheet)	OPIGINAL CRANNE	AL NUMBER = 53 · LAREL =	DNA2 (Tr	r193D()						fcs_directory	Y:/Bioimaging_Staff/George/20230315_AutoDECRA/6
	TARGET CHANNE'	L NUMBER = 31 : LABEL =	DNA2 (Ir	r193Di)						fcs_name	Y:/Bioimaging_Staff/George/20230315_AutoDECRA/fi
										fcs_name_all	SPJ_Tonsil_28062021_ROI004_ROI_004.ome.titt
	Writing SPJ_Tonr	#11_28062021_R0I004_R0I_	004.om	e						FCS_ROLDATA	Combined dataset
Manufacture and the later										fileName	'CMW_Baseline_28062021_ROI001_ROI_001'
NO détais avanable	Writing Combined	dataset								1 filepath	
										📑 ilastik_data	'Y:/Bioimaging_Staff/George/20230315_AutoDECRA/4
	6									ilastik_instruction	'headlessproject'
	14.55								, [*]	<pre>ilastik_model </pre>	*Y/Bioimaging_Staff/George/20230515_AutobicLIAV *

6. Adjust the save name for your heatmaps to reflect any specific alterations you are trialling or testing. For example, as default we have the heatmaps saved specifying the pathology and the pixel number for cell expansion, but you could incorporate any of the other parameters used within your heatmap generation.

📣 MATLAB R2022a - academic use			- a ×
HOME PLOTS APPS EDITOR P	ISH VIEW	🖥 / h 🛍 h ơ 🖻	🤉 💿 Search Documentation 🛛 🔎 🐥 Sign
	% % 🐉 🗔 🛌 🕼 Section Break 🕒 🕒		
New Onen Store Control - Go To Q Find - Refactor	2 28 Lis Gui Fromer Bun 28 Run and Advance Run Sten Ston		
• • • • • • • • • • • • • • • • • • •	Analyze Section 🔀 Run to End 👻		
FILE NAVIGATE	DE ANALYZE SECTION RUN		
💠 🔶 🔞 🔀 🎏 📙 🕨 Y: 🕨 Bioimaging_Staff 🕨 George 🖡 20230315_A	DECRA + 2_image_extraction +		•
Current Folder 📀	Editor - H1\Downloads/20230306_Tonsil_Matlab_Codes_V212_image_extraction\OPTIMAL_Pipeline_Part_2_Heatmaps.m	Workspace	
Name 🔶	OPTIMAL_Pipeline_Part_1.m × OPTIMAL_Pipeline_Part_1.m × OPTIMAL_Pipeline_Part_2.Heatmaps.m × +	Name +	Value
Functions	279 % Extract higher and lower logic ofter pVal_sig for each image	* 🔺 📑 all_ROL_list	1x27 struct
Outputs SuiteDECRA EPCAM Clusters VS view	<pre>280 parfor_gates_high(ROI_for_path) = {[combos_all,Higher_logis]};</pre>	all_ROI_loop	27
-SAutoDECRA_MAY22.xisx	<pre>281 paring gates_low(ROI_for_path) = ([combos_all,Lower_logic]); 282 to:</pre>	AS DATA	19773v31 double
AutoDECRA_Data_Analysis_Master.xlsx	283 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX	AS_NORM_DATA	19073x31 double
AutoDECRA_EPCAM_Clusters_V2.csv	284 - end	batch_number	1
AutoDECRA_EPCAM_Clusters_V3.csv	285	Batch_number_le	st 27x1 double
AutoDECRA_EPCAM_Clusters_V5_COPY_PRE_ALTERATION.csv	280 Xa Heatmaps for each images individual and special cluster complications [Natrix Delta.Matrix low, Unique all, Unique pour all.Matrix high, Unique high all	channel_counter	1x36 cell
AutoDECRA_EPCAM_Clusters_V5.csv	<pre>288 = Heatmap individual images(parfor gates high,parfor gates low);</pre>	channel_loop	63
AutoDECRA_EPCAM_Clusters_V5.xlsx	289	d check_channel_n	1x60 cell
AutoDECRA March23.xlsx	290	theck_number	u tolt cell
AutoDECRA_MAY22.xlsx	292 XXGenerate an asymmetric heatmap	co factor	1
Cluster_Assignments_and_Merges_AutoDECRA.xlsx	293 Asymmetric_heatmap(parfor_gates_high,parfor_gates_low, Matrix_high,	comp_channels	h31 cell
Cluster_Assignments_and_Merges_AutoDECRA_V2x.xlsx	294 Matrix_low,Unique_high_all,Unique_low_all,new_cluster_names,cut_off_percent)	COMP_DATA	19073x31 double
OPTIMAL Pipeline Part 1 V2.m	295	comp_matrix	3 IX32 TODIE
OPTIMAL_Pipeline_Part_1.m	297	compensation_m	n 31x31 double
OPTIMAL_Pipeline_Part_2_Heatmaps.asv	298 🗄 %%Saves the output heatmap to the output folder based on the pathology	COMPLETE_COM	A., 219593x73 double
OPTIMAL_Pipeline_Part_2_neatmaps.m OPTIMAL_Pipeline_Part_2_Heatmaps_MultiFactor_Plotting.m	299 %%(in this case)	condition_list	27x1 cell
OPTIMAL_Pipeline_Part_3_Visualisaton.m		counter	2/
spillover2.1.xlsx	<pre>302 fig = get(groot, 'CurrentFigure');</pre>	cp_input	-i C:/Users/george/MATLAB_Macros/20230315_Auto
📦 spillover2.1 - Copy.xlsx	303 fig.Position = [10, 10, 1200, 900]	cp_output	-o C:/Users/george/MATLAB_Macros/20230315_Auto
	384 saveName = strcat(outputFolder, "Pathology_", int2str(pathology_to_heatmap), "_Pixel_Radius_", int2str(number_of_pixels)	cp_path	C:/Users/george/MATLAB_Macros/CellProfiler/CellPr
	subs saveas(gcr, savename)	in co settinos	-p C/user/george/marickb_macros/20230515_Add
	307	current_mask	'CellMasks027.npy'
	308 end	current_mask_dir	re 'Y:\Bioimaging_Staff\George\20230315_AutoDECRA\5
	309 114	 current_mask_file current_ROI 	 Y:\bioimaging_Staff\George\20230315_AutobELRA\3 'SRI Toorii 20052031 ROI004 ROI 004 cons tiff'
	311 XXEnd of Harro	current ROI dires	ct., 'Y/Bioimaging Staff/George/20230315 AutoDECRA/1
	312	- = 💽 current_ROL_filen	a 'Y:\Bioimaging_Staff\George\20230315_AutoDECRA\1
	4	display_rescaled_	.d., 1
	ommand Window	EPCAM chappel	29
	New to MATLAB? See resources for Getting Started.	× EPCAM_max	4
	TARGET CHANNEL NUMBER = 29 : LABEL = CD03(Lu175D1)	A EPCAM_min	0
	READING DATA FROM	EXTRA DATA	10072-5 deviate
	ORIGINAL CHANNEL NUMBER = 49 : LABEL = CD45(Yb176D1)	extra markers	h5 cell
	TARKET (MANNEL NUMBER = 30 1 LABEL = CD-S(TDI/6D1) BEADING DATA FROM	1 fcs_channel_list	1x73 cell
AutoDECKA_MAY22.xlsx (Microsoft Excel Worksheet)	ORIGINAL CHANNEL NUMBER = 53 : LABEL = DNA2(Ir193D1)	fcs_directory	'Y:/Bioimaging_Staff/George/20230315_AutoDECRA/f
	TARGET CHANNEL NUMBER = 31 : LABEL = DNA2(II193D1)	tcs_name	'Y:/bioimaging_Staft/George/20230315_AutoDECKA/1 'SRI Tenril 28052031_ROI004_ROI 004 emertial?
		FCS_ROI_DATA	19073x73 double
	Writing SPJ_Tonsil_28062021_R0I004_R0I_004.cmm	ilename	'Combined_dataset'
No details available	Writing Combined dataset	fileName	'CMW_Baseline_28062021_RO1001_RO1_001'
	strain construction	ilastik data	"Yv/Rigimaging Staff/George/20230315 AutoDECRA/
		ilastik_instruction	 'headlessproject'
	£>>	 ilastik_model 	'=Y:/Bioimaging_Staff/George/20230315_AutoDECRA/
	< compared by the second secon	> <	>

7. Run the script

Matlab Script III

 To create colour-coded cell masks of your data based on cell type information, use the Matlab script "OPTIMAL_Pipeline_Part_3_Visualisaton.m". First, adapt the folder paths for the output figure location, the csv file containing every cell with a specific cluster assignment, the directory containing the cell mask npy files from CellProfiler, and the raw image folder (for save name creation)

HOME	PLOTS	APPS	EDITOR	PUBLISH	wew									(6 %	a 🛍 ta et 🖨 (🕐 🛞 Search Documentation 🛛 🔎 🌻 Sign
A	E	201	44	6 % 3	(\$)		Section Break		e.								
~ -	Compare ·		Q Find -		C Profile		Run and Advance		~								
New Open	Save 🖶 Print 👻	Golo	Bookmark •	Ketactor	ar Analys	e Kun Section	Run to End	Kun	Step	Stop							
	FILE		NAVIGATE	CODE	ANALYZE		SECTION		RUN								
4 🔶 🖬 🔓	🕽 💭 📙 🕨 Y: 🕨 Bio	imaging_St	iff → George → 20	230315_AutoDECR	A + 2_image_extr	action >											•
Current Folder				💿 📝 Edi	or - Ht\Downloads	\20230306_Ton	sil_Matlab_Codes_V2	12_image_e	xtraction)	OPTIMAL	Pipeline_Part_3_Visualisato	n.m			Θ×	Workspace	
Name -	A			0	TIMAL_Pipeline_P	nt_t.m.≍]i	OPTIMAL_Pipeline_P	art_1.m ×	OPTIN	(AL_Pipel	ine_Part_2_Heatmaps.m 🛛 🛛	OPTIMAL_Pipeline_	Part_3_Visualisation.m :	× +		Name +	Value
🗉 📙 Function	ns			1	🖓 🕱This	macro servi	es as part 3 of	the ima	ge proc	essing	pipeline for the				^ A	E all ROI list	bi27 struct
Outputs				2	\$\$publ:	ication nam	ed "OPTIMAL: An	OPTimis	ed Imag	ing Mas	s cytometry Analysis					all_ROI_loop	27
A SAUTO	DECRA_EPCAM_Clusters DECRA_MAY22 view	s_volkex			Stinste	nuctions fi	le for detailed	i baca ex	tions o	n how t	case review the					ans ans	1
AutoDEC	CRA_Data_Analysis_Mas	teraise		5	\$\$macro	as necess	ary									AS NORM DATA	19073x31 double
👰 AutoDEG	CRA_EPCAM_Clusters_V	2.csv		6												batch_number	1
AutoDEC	CRA_EPCAM_Clusters_V	3.csv		7	S%Clear	's the comm	and window and	resets t	he Nork	space t	o ensure previous				-	🔒 Batch_number_list	27x1 double
AutoDEC	CRA_EPCAM_Clusters_V CRA_EPCAM_Clusters_V	4.CSV IS CODV PR	E ALTERATION CITY	9	clear	do not into	errere with the	macro n	unnang							channel_counter	060
AutoDE0	CRA_EPCAM_Clusters_V	5.csv		10	close	11										channel loop	63
👰 AutoDEG	CRA_EPCAM_Clusters_V	5.xbx		11	clc											() check_channel_na.	1x60 cell
AutoDEC	CRA_Graphing_Outputs.	zip		12	C Wadde	also Thread	inne" felden te	the set			o use of the					theck_number	0
AutoDEC	CRA MAY22xisx			14	1 XXfunct	tions neces	sary for this m	acro to	run	area in	a use of the					choosen_channel	h31 cell
Cluster_	Assignments_and_Merg	es_AutoDE0	RAxiss	15	addpati	n("./Functi	ons/")								=	Comp channels	b31 cell
👰 Cluster_	Assignments_and_Merg	es_AutoDE	RA_V2xxlsx	16												COMP_DATA	19073x31 double
Final_lur	ng_clustered_output.csv V_Bioeline_Part_1_V2 m			1/	Stolen Stolen	e edent th	rectories and P	11es		to work	with your folder				-	comp_matrix	31x32 table
OPTIMA	AL Pipeline Part 1.m			19	XXname	for your	device. The exp	lanation	of eac	h folde	r/file is given					compensated_sin	31x1 double 31x21 double
🗍 ΟΡΤΙΜΑ	L Pipeline Part 2 Heats	maps.asv		20	Kinner	diately prio	or to the folde	r being	named							COMPLETE COM	219593x73 double
Ο ΟΡΠΜΑ	AL_Pipeline_Part_2_Heatr	maps.m		21												 condition_list 	27x1 cell
	AL_Pipeline_Part_2_Heatr M. Bioeline_Part_2_Vicus	maps_Multi licaton m	Factor_Plotting.m	22	L Sketi	code	put toider for	where to	save t	ne imag	es generated in					eount count	27
spillover	r2.1.xlsx			2.4	output	Folder - 'C	:\Users\george\	MATLAB N	acros\2	0230306	Tonsil Matlab Codes	V2\2 image extra	action\20230208 Ou	tput Figures Tor		counter counter	UUT
spillover	r2.1 - Copy.xlsx			25	%%Defi	ies the clu	ster Output Fil	e for th	e image	s of in	terest	1				cp output	- o C:/Users/george/MATLAB Macros/20230315 Auto
				26	filena	e='Tonsil_	EPCAM_Clusters_	V2.csv';								cp_path	'C:/Users/george/MATLAB_Macros/CellProfiler/CellPr
				27	Storfin	nsliciuste	redoutput = imp	tory	Tilenam	e);						cp_project	-p C:/Users/george/MATLA8_Macros/20230315_Autr
				29	Mask_w	orking_dire	tory='C:\Users	\george\	MATLAB_	Hacros	20230306_Tonsi1_Mat1	ab_Codes_V2\5_CP	pipeline\OUTPUTS\	CellMasks\';		cp_settings	'-c-r' 'CellMacks027.nev/
				30	%%Defi	ies the fil	e Location for	Tonsil O	riginal	Files	to get Mask Filename	for Saving				current_mask_dire.	Y\\Bioimaging_Staff\George\20230315_AutoDECRA\5
				31	origin	al_files 👳	'C:\Users\georg	e (MATLAB	Macros	202204	24_Tonsil_Raw\1_Imag	eData\'				current_mask_file	'Y:\Bioimaging_Staff\George\20230315_AutoDECRA\5
				32												current_ROI	'SPJ_Tonsil_28062021_R0I004_R0I_004.ome.tiff'
				34	- SSExtra	acts the cl	uster numbers a	nd e.g.	patholo	gy numb	ers from the file				-	current_ROL filena	Yillioimaging_staffiGeorge/20230315_AutoDECRA(1) Yillioimaging_staffiGeorge/20230315_AutoDECRA(1)
				35	4									•		display_rescaled_d.	. 1
				Come	and Window										۲	End_ROI_Index	0
				New 1	MATLAB? See res	surces for Getti	ng Started.								×	EPCAM_channel	29
					ARGET CHANNE	T. NUMBER -	29 : LAREL =	CD83 (La	17501)						^	EPCAM_min	0
				RE	DING DATA FR	OM										di ext	'siff'
				OR	GINAL CHANNE	L NUMBER -	49 : LABEL =	CD45 (YE	176Di)							EXTRA_DATA	19073x5 double
					TARGET CHANNE	L NUMBER -	30 : LABEL -	CD45 (Yr	176D1)							extra_markers fcs_channel_list	to cell ty73 cell
AutoDECRA_M	AY22.xlsx (Microsoft Ex	ccel Worksh	eet)	~ RE.	DING DATA FF	IOM										fcs_directory	Y:/Bioimaging_Staff/George/20230315_AutoDECRA/f
				OR OR	CHANNE CHANNE	L NUMBER .	31 : LABEL =	DNA2 (11 DNA2 (1)	193D1)							fcs_name	'Y:/Bioimaging_Staff/George/20230315_AutoDECRA/f
									,							fcs_name_all	'SPJ_Tonsil_28062021_ROI004_ROI_004.ome.tiff'
				Wr	ting SPJ_Ton	sil_280620	21_R0I004_R0I	_004.ome								filename	'Combined dataset'
	No detail	le available														fileName	'CMW_Baseline_28062021_R0I001_R0I_001'
	NO GELSI	o o contrative		Wr	ting Combine	d_dataset										ilepath 👔	
																ilastik_data	'Y:/Bioimaging_Statt/George/20230315_AutoDECRA/-
				fx >>											~	ilastik_model	'=Y:/Bioimaging_Staff/George/20230315_AutoDECRA
				<											>	<	,
1111.4													Zoom: 100%	UTF-8	CRL	F script	Ln 19 Col 56

 Update the name for the clustering information file that you have for your data, including the names of relevant sheets for specific clustering tiers if necessary. Make sure to adapt the "TIER_TO_USE" parameter to the relevant tier for your data

Automatical Strategy Strateg	HOME PLOTS APPS EDITOR	PUBLISH VIEW	64	h 🛍 Ə Ə 💆 🕻) 💿 Search Documentation 🛛 🔎 🦊 Sign In
Notes Notes <th< td=""><td>💠 🐚 🖬 🕅 Compare - 🔿 🗇 🗇</td><td>🐼 % 🐲 💱 🔊 Profiler 📔 🗈 Section Break 📦 🍋</td><td></td><td></td><td></td></th<>	💠 🐚 🖬 🕅 Compare - 🔿 🗇 🗇	🐼 % 🐲 💱 🔊 Profiler 📔 🗈 Section Break 📦 🍋			
• • • • • • • • • • • • • • • • • • •	New Open Save Print - Go To Find - R	fector D et Le Analyze Run 23 Run and Advance Run Step Stop			
Note Note Note Note Note No	• • • Bookmark •	✓ Kal → Lar Annaya Section Kal Run to End →			_
(a) (b) (b) (b) (b) (b) (b) (b) (b) (b) (b	FILE NAVIGATE	CODE ANALYZE SECTION RUN			Ā
Current for O Current for O Current for O Marge: O Note: O Comparison: O Comparison: O Note: O Note: O Comparison: O Comparison: O Note: Note: </td <td>💠 🜩 🛅 🚰 📴 🕨 Y: 🕨 Bioimaging_Staff 🕨 George 🕨 2023</td> <td>015_AutoDECRA + 2_image_extraction +</td> <td></td> <td></td> <td>م •</td>	💠 🜩 🛅 🚰 📴 🕨 Y: 🕨 Bioimaging_Staff 🕨 George 🕨 2023	015_AutoDECRA + 2_image_extraction +			م •
Nort Office: Office: Office: Nort Nort 0 Add/CAL_Mark_Since: 0 0 0 0 0 0 0 0 0 0 0 0	Current Folder	Z Editor - H\Downloads\20230306_Tonsil_Matlab_Codes_V2\2_image_extraction\OPTIMAL_Pipeline_Part_3_Visualisaton.m	⊛ ×	Workspace	•
Internal of the second provide provide second provide provide provide provide provide provide	🗋 Name 🗠	OPTIMAL_Pipeline_Part_1.m × OPTIMAL_Pipeline_Part_1.m × OPTIMAL_Pipeline_Part_2.Heatmaps.m × OPTIMAL_Pipeline_Part_3_Visualisaton.m × +		Name 🗠	Value
 Bit State Control (State Control (Stat	Functions	48	÷ 🔒	all_ROI_list	1x27 struct ^
<pre>d - decide, March Market Market</pre>	Outputs State DECRA EDCAM Charters VE also	49 50 Staffings the cluster assignments based on user input of using either tier		all_ROI_loop	27
add/CCLA_int_Add/CCLA_int_Add/CCLALABLE Add/CCLA_int_Add/CCLALABLE Add/CCLALABLE Add/CCLA	-SAutoDECRA MAY22.xbx	51 3%1 or tier 2. Change the Tier to use depending on what tier you would like		ans AS DATA	1977-21 double
<pre> Action (Co. Mark Come, Com</pre>	AutoDECRA_Data_Analysis_Master.xlsx	52 %Xto use		AS_NORM_DATA	19073x31 double
Add/CDL VLAQ Loss 1/10* Control (Control (Con	AutoDECRA_EPCAM_Clusters_V2.csv			batch_number	1
<pre>b AccCCLAPCALCUMUN_SCOPE_PLATEADORCH b ACCCCLAPCALCUMUN_SCOPE_PL</pre>	AutoDECRA_EPCAM_Clusters_V3.csv	54 ITEX_IO_USE=2; 55 IF THE TO USE==1		Batch_number_list	27x1 double
<pre> Accord Coll Coll Coll Coll Coll Coll Coll Col</pre>	AutoDECRA EPCAM Clusters V5 COPY PRE ALTERATION.csv	56 % Reed tier 1 info		Channel_counter	1/06 cell
AddCCM_MCUS_TSATe AddCCM_MCUS_TSATE<	AutoDECRA_EPCAM_Clusters_V5.csv	57 new_cluster_table= readtable('Cluster_assignments_and_Herges_TONSIL.xlsk', 'Sheet', 'TIER 1');		channel_loop	63
Add/CCM_AMA22is Add/CCM_AMA22is Concert_subject_static_field_interi	AutoDECRA_EPCAM_Clusters_V5.xlsx	58 Tier_1 name_from_sheet=new_cluster_table.Var3;		theck_channel_na	1x60 cell
AudoCCM_MAV22ac 0	AutoDECRA_Graphing_Outputs.zp	59 11er_l_COVINEW_CLUSTer_table.Var6; 60 Tirr1 asignertable.var6;		check_number	0
B. Chere, Jagiment, and, Meg, Jackob, Cachas <i>if print("life": listigness": sli: original cent sli: vli: life:_listigness: sli: original cent sli: vli: vli: vli: vli: vli: vli: vli: v</i>	AutoDECRA MAY22.visx	61 G for loop=1:numel(Tier_1 assignments)		choosen_channel	131 000
0 Constrainment,and,Margan,AndOCA,Value 0 if for inf, and explainment,and,Margan,AndOCA,Value 0 Constrainment,and,Margan,AndOCA,Value 0 if for inf, and explainment, and, Margan, AndOCA,Value 0 if for inf, and explainment, and, Margan, AndOCA,Value 0 if and inf, and explainment, and, Margan, AndOCA,Value 0 if and inf, and explainment, and, Margan, AndOCA, Value 0 if and inf, and explainment, a	Cluster_Assignments_and_Merges_AutoDECRA.xlsx	<pre>62 fprintf('Tier 1 assignment = %i : Original cSOM = %i \r',Tier_1_assignments(loop),Tier_1_cSOM(loop))</pre>	-	Comp_channels	1x31 cell
<pre>Big Description and States and Like 1 Big Description and Big Description and Like 1 Big Description and Big Description and Like 1 Big Description and Big Descri</pre>	Cluster_Assignments_and_Merges_AutoDECRA_V2x.xlsx	<pre>63 fprintf('%s \r',Tier_1_nome_from_sheet{loop});</pre>		COMP_DATA	19073x31 double
<pre> OrBAL_pedic_Met_1</pre>	Final_lung_clustered_output.csv	64 - eno 65 % Generate new cluster list	_	comp_matrix	31x32 table
No deals would be added to send the send the send to se	OPTIMAL_Pipeline_Part_1.m	<pre>66 number_of_Tier_1_cSOM=numel(unique(Tier_1_cSOM));</pre>		compensated_sin	31x1 double
<pre>B GFML peice_bet2.Manuagement for ison index of iter ison index o</pre>	OPTIMAL_Pipeline_Part_2_Heatmaps.asv	<pre>67 Tier_1_new_cluster_number=zeros(numel(cluster_number),1);</pre>		COMPLETE_COM	219593x73 double
 	OPTIMAL_Pipeline_Part_2_Heatmaps.m	68 for loop=1:number of Tier_1_cSOM		1 condition_list	27x1 cell
AutoCCCA_MAV22.bs [Monoth Excl Workshed] v ref	OPTIMAL_Pipeline_Part_2_Heatmaps_MultiFactor_Piotting.m OPTIMAL_Pipeline_Part_3_Visualisaton.m	70 Tier 1 new cluster number (cluster indices)=Tier 1 assignments(loop);		count	27
Pipeboch1-Copysite 72 Tisc_1,res_(List_r_under(intint_s),res_(List_r_under(intint_s)),res) 10 <	spillover2.1.xbx	71 end		counter	ui Crilisers/nenroe/MATLAB_Marros/20230315_Auto
AudotCGA_AMV22.bit (Monosh Burd Workham) V No datis emiliate V V V	💀 spillover2.1 - Copy.xlsx	72 Tier_1_new_cluster_number(Tier_1_new_cluster_number==0)=NeN;		cp_output	-o C:/Users/george/MATLAB_Macros/20230315_Auto
AddOCEA,MAYZJake (Moceash Each Workhed)		73 Tier 1 assignments indices-unique (Tier 1 assignments);		cp_path	'C:/Users/george/MATLAB_Macros/CellProfiler/CellPr
AuddEGGA_MAY22.bit (Moouth burd Weshled) v image: selection of the s		74 TIEF_L_NAMESHOTIQUE(TIEF_L_NAME_FOR_STOLE); 75 % such 25		<pre>cp_project</pre>	-p C:/Users/george/MATLAB_Macros/20230315_Auto
AutoCED_MAV22.lss (Mocosth Locd Worksheet) *		<pre>76 new_cluster_number=Tier_1_new_cluster_number;</pre>	_	cp_settings	'-c-r' 'CellMasks027.nm/
AudoCCA_MAY22.4s (Moouth bud Weshed) * File *		77 new_cluster_names=Tier_1_names;		current_mask_dire	'Y:\Bioimaging_Staff\George\20230315_AutoDECRA\5
AutoCED_MAV22.bit (Mocodf Excf Worksheft) *		78 else	-	current_mask_file	'Y:\Bioimaging_Staff\George\20230315_AutoDECRA\5
No deals enabled 11 11/2 * 2.000 (month, base) for the second for deals belows; 1 1		<pre>// a new cuter 2 into</pre>		current_ROI	'SPJ_Tonsil_28062021_RO1004_ROI_004.ome.tiff'
Image: Provide and the set of th		81 Tie_2_name_from_sheet=new_cluster_table.Name;	-	current_ROL_direct	'Y:\Bioimaging_Staff\George\20230315_AutoDECRA\1 'Y:\Bioimaging_Staff\George\20230315_AutoDECRA\1
Comment Works Comment		82 4		display_rescaled_d	1
AutoOCCRA_MAV22.bis (Moseneth Excit Workheet) * * 2002, Moneth 20 2002, Moneth 20 AutoOCCRA_MAV22.bis (Moseneth Excit Workheet) * 1002, Moneth 20 0 0 AutoOCCRA_MAV22.bis (Moseneth Excit Workheet) * 1002, Moneth 20 0 0 AutoOCCRA_MAV22.bis (Moseneth Excit Workheet) * 1.2002 0 0 0 Moneth Excit Workheet) * 1.2002 Contrast 0002, Moneth 20 0 0 Moneth Excit Workheet) * 1.2002 Contrast 0.2002, Moneth 20 0 0 0 Moneth Excit Workheet) * 1.2002 Contrast 0.2002, Moneth 20 0 0 0 Moneth Excit Workheet) * 1.2002 Contrast, Totol 1.2002, Moneth 20 0		Command Window		End_ROI_Index	0
Andol(CAA_MAV22.htm (Monored) Excel Workshed) > 1 JABET = CAASUET (COASUET (NOTER + 2) = 1 JABET = COAS(1:1750)) > 0 eff > 1 off Andol(CAA_MAV22.htm (Monored) Excel Workshed) > 1 JABET = CAASUET (COASUET (NOTER + 2) = 1 JABET = COAS(1:1750)) > 0 eff > 1 off No deals evaluate Y (Monored) Excel Workshed) > 1 JABET = CAASUET (COASUET (NOTER + 2) = 1 JABET = COAS(1:1750)) > 0 eff > 1 off No deals evaluate Y (Monored) Excel Workshed) > 1 JABET = CAASUET (COASUET (NOTER + 2) = 1 JABET = COAS(1:1500)) > 1 off		New to MATLAR? See resources for Getting Started	×	EPCAM_channel	29
AutoOCCRA_MAV22.bis (Monosoft Excl Workheet) if et imm Victor if et imm AutoOCCRA_MAV22.bis (Monosoft Excl Workheet) if et imm Victor if et imm No details workheet if et imm Victor if et imm No details workheet if et imm Victor if et imm No details workheet if et imm Victor if et imm Victor if et imm No details workheet if et imm Victor if et i		TADOT CHANNEL WINDER = 26 , TADET = CD02/Ful7ED()		EPCAM min	8
AutoSCEA_MAV22.bit OMECONAL_CRAMEEL_INDERER. + 0: 1: LABLE - CEVE(10:17101) Image: Clamber Linderer, 4: Lable - CEVE(10:1701) Image: Clamber Linderer, 4: Lable - CEVE(10:1701) AutoSCEA_MAV22.bit (Moread) # Lable - CEVE(10:1701) Image: Clamber Linderer, 4: Lable - CEVE(10:1701) Image: Clamber Linderer, 4: Lable - CEVE(10:1701) AutoSCEA_MAV22.bit (Moread) # Lable - CEVE(10:1701) Image: Clamber Linderer, 4: Lable - CEVE(10:1701) Image: Clamber Linderer, 4: Lable - CEVE(10:1701) No details available V		REDING DATA FROM		👍 ext	'.tiff'
And/OECRA_MAV22.hts (Mossed) TABLET TABLET Colds Cold And/OECRA_MAV22.hts (Mossed) V TableT Cold (Colds, Molecular) Distribution Distribution And/OECRA_MAV22.hts (Mossed) V TableT Cold (Colds, Molecular) Distribution		ORIGINAL CHANNEL NUMBER = 49 : LABEL = CD45(Yb176Di)		EXTRA_DATA	19073x5 double
AudoECEQ_MAV22.bis (Mosole face) Workheet) V REALTSIE DATA FROM V Signature Signature <td< td=""><td></td><td>TARGET CHANNEL NUMBER = 30 : LABEL = CD45(Yb176D1)</td><td></td><td>extra_markers</td><td>Iso cell Is72 cell</td></td<>		TARGET CHANNEL NUMBER = 30 : LABEL = CD45(Yb176D1)		extra_markers	Iso cell Is72 cell
Object Lookale, Notest,	AutoDECRA_MAY22.xlsx (Microsoft Excel Worksheet)	V READING DATA FROM		fcs directory	'Y:/Bioimaging Staff/George/20230315 AutoDECRA/Fi
No details realiable First 2004-2012 (ROCING 4, 2004-2014) State 1, 2004-2014 (ROCING 4, 2004-2014) State 1, 2004-20		OKIGINAL CHANNEL NUMBER = 53 : LABEL = DNAZ (17193D)		fcs_name	'Y:/Bioimaging_Staff/George/20230315_AutoDECRA/f-
Writing SRy_Tossil_B06/301_B01004_B01_001-ome Inclusion Inclusion Inclusion No deals available Writing Combined_dataset Inclusion Inclusion Inclusion Vision Writing Combined_dataset Inclusion Inclusion Inclusion Inclusion At >> Inclusion Inclusion Inclusion Inclusion Inclusion Inclusion Inclusion Inclusion		TROOL CHARGED HOLDER OF I ABOUT DARK (III)ODAY		fcs_name_all	SPJ_Tonsil_28062021_RO1004_ROI_004.ome.tiff
No details realiable Efficiency Control (Backage Contro))))))))))))		Writing SPJ Tonsil_28062021_R0I004_R0I_004.come		FCS_ROI_DATA	'Combined dataset'
No States shallow Utiling Cublished_dataset Of Repair Visionsed_dataset State_data State_data Area State_data State_data Visionsed_dataset State_data State_data Area State_data State_data Visionsed_dataset State_data State_data	M 11 2 1			fileName	'CMW Baseline 28062021 ROI001 ROI 001'
A lank, deta **/ViBiomaging, Staff Group 2020015, AuchOECAW A lank, interface */ViBiomaging, Staff Group 2020015, AuchOECAW A lank, interface * A lank, interface *	ivo details available	Writing Combined_dataset		ilepath filepath	
/f_ >> heartiesproject heartiesproject > heartiesproject				ilastik_data	Y:/Bioimaging_Staff/George/20230315_AutoDECRA/4
C S S S S S S S S S S S S S S S S S S S		(fr >>		astik_instruction	headlessproject V-/Bioimaging Staff/Genroe/20230315 AutoDECRA/ Y
			>	<	>

3. If you would like to plot only selected cell type clusters, adapt the list of clusters in the array "selected_clusters_to_plot" to only include clusters of interest



4. To incorporate additional information into the save name of the files, update the parameter "saveName" in the macro, by default we include the file name along with the clustering tier specified by the user

📣 MATLAB R2022a - academic use				- 0 ×
HOME PLOTS APPS EDITOR	URUSH NEW		- h 🛍 'S C 🖨 🤇	🗴 🛞 Search Documentation 🛛 🔎 📮 Sign I
🔶 🗀 🔚 🛛 Company 🗸 🏟 🔶 🗸	% 🐄 🖓 📶 Profiler 📔 🗄 Section Break 📄 🧠			
New Open Save Brint - Go To Find - Refa	De les wanter Run Step Stop			
• • • Bookmark • •	🔄 • Section 📓 Run to End •			
FILE NAVIGATE	CODE ANALYZE SECTION RUN			
George + 2023031	JRUELKA - ZJIMBE EKTATION -		Wedness	• 2
D Name e	OPTIMAL Province Part I m × OPTIMAL Province Part I m × OPTIMAL Province Part 2 Hostmark m × OPTIMAL Province Part 3 Visualisation m × +		Nonspice	N.L.:
Functions	229 NEW MASK g(b(border_pixel_loop,1),b(border_pixel_loop,2)) =NEW_MASK_b(b(border_pixel_loop,1),b(border_pixel_loop,2)) = NEW_MASK_b(b(border_pixel_loop,2),b(border_pixel_loop,2)) = NEW_MASK_b(b(border_pixel_loop,2)) = NEW_MASK_b(b(b(border_pixel_loop,2)) = NEW_MASK_b(b(b(b(b(b(b(b(b(b(b(b(b(b(b(b(b(b(b(1_ ^ A		value Iv27 struct
Outputs	230 NEw_MASK_b(b(border_pixel_loop,1),b(border_pixel_loop,2)) = NEw_MASK_b(b(border_pixel_loop,1),b(border_pixel_loop,2)) = NEW_MASK_b(b(border_pixel_loop,2),b(border_pixel_loop,2)) = NEW_MASK_b(b(border_pixel_loop,2)) = NEW_MASK_b(b(b(border_pixel_loop,2)) = NEW_MASK_b(b(b(border_pixel_loop,2)) = NEW_MASK_b(b(b(border_pixel_loop,2)) = NEW_MASK_b(b(b(b(b(border_pixel_loop,2))) = NEW_MASK_b(b(b(b(b(b(b(b(b(b(b(b(b(b(b(b(b(b(b(1_	all_ROL loop	27
SAutoDECRA_EPCAM_Clusters_V5.xlsx	231 end 232 end		ans	1
AutoDECRA_Data_Analysis_Master.xlsx	233 mask_cell_loop		AS_NORM_DATA	19073x31 double
AutoDECRA_EPCAM_Clusters_V2.csv	234 - end		batch_number	1
AutoDECRA_EPCAM_Clusters_V4.csv	235 NPU_MASK(:,;,1)-WEI_MASK;		Batch_number_list	27x1 double
AutoDECRA_EPCAM_Clusters_V5_COPY_PRE_ALTERATION.csv	237 NEW_MASK(:,:,2)=NEW_MASK_8;		() channel_list	1x36 cell
AutoDECRA_EPCAM_Clusters_V5.csv	238 NEW_MASK(:,:,3)=NEW_MASK_D; 239		channel_loop	63
AutoDECRA_Graphing_Outputs.zip	240		check_channel_na.	0
AutoDECRA_March23.xlsx	241 C XXShows the cluster image, and saves to the output folder based on the		choosen_channel	tx31 cell
Cluster Assignments and Merges AutoDECRA.xlsx	242 AAUSET INDUS 243 - XUXXIXIXIXIXIXIXIXIXIXIXIXIXIXIXIXI		co_factor	1 1x21.cel/
Cluster_Assignments_and_Merges_AutoDECRA_V2x.xlsx	244 figure(1)		COMP_DATA	19073x31 double
Final_lung_clustered_output.csv OPTIMAL_Directions_Part 1_V2 m	245 imshow(NEW_MASK,[]) 246 ims name = ims filenemes(BOI to generate image):	-	comp_matrix	31x32 table
OPTIMAL_Pipeline_Part_1.m	247 inlegend(colors(selected_clusters_to_plot,:), labelsArr);		compensated_sin	31x1 double 31x31 double
OPTIMAL_Pipeline_Part_2_Heatmaps.asv	248 title('Clustered mask')		COMPLETE_COM	219593x73 double
OPTIMAL_Pipeline_Part_2_Heatmaps.m OPTIMAL_Pipeline_Part_2_Heatmaps.MultiFactor_Plotting.m	249 savename strat(output/older, ile_, intistr(ilek_id_ose), _image_, img_name, .png) 250 savens(ecf.saveName)		() condition_list	27x1 cell
OPTIMAL_Pipeline_Part_3_Visualisaton.m	251 3383113941139411394113941139411394113941		counter	10011
spillover2.1.xlsx	252 - end		cp_input	-i C:/Users/george/MATLAB_Macros/20230315_Auto
- shunders - colliner	254 333		cp_output	 - o C/Users/george/MATLA8_Macros/20230315_Auto 'C:/Users/george/MATLA8_Macros/CellProfiler/CellPr
	255		cp_project	-p C:/Users/george/MATLAB_Macros/20230315_Auto
	255		cp_settings	'-c-r' 'CaliMada 037 and
	258		current_mask_dire	'Y:\Bioimaging_Staff\George\20230315_AutoDECRA\5
	259 E function inlegend(colorArr, labelsArr) 269 E for instance if the largend apties are peeded:		current_mask_file	'Y:\Bioimaging_Staff\George\20230315_AutoDECRA\5
	261 X colorArr =		current_KOI	SPJ_tonsil_28062021_ROI004_ROI_004.ome.bff V\Bioimaning_Staff_George\20230315_AutoDECR4\1
	262 % Nx3 array of doubles. Each entry should be an RGB percentage value between 0 and 1	-	current_ROL filena	'Y:\Bioimaging_Staff\George\20230315_AutoDECRA\1
	263 4	- F	display_rescaled_d.	1
	Command Window		EPCAM_channel	29
	New to MATLAB? See resources for <u>Getting Started</u> .	×	EPCAM_max	4
	TARGET CHAINEL NUMBER = 29 : LABEL = CD03(Lu175D1)	^	EPCAM_min	u 'tiff
	CRIGING DATA FROM ORIGINAL CHANNEL NUMBER = 49 : LABEL = CD45(Yb176Di)		EXTRA_DATA	19073x5 double
	TARGET CHANNEL NUMBER = 30 : LABEL = CD45 (Yb176D1)		extra_markers	1x5 cell
AutoDECRA_MAY22.xlsx (Microsoft Excel Worksheet)	READING DATA FROM		fcs_directory	'Y:/Bioimaging_Staff/George/20230315_AutoDECRA/fi
	TARGET CHANNEL HUMBER = 53 : LABEL = INA2(119301) TARGET CHANNEL NUMBER = 31 : LABEL = INA2(119301)		fcs_name	'Y:/Bioimaging_Staff/George/20230315_AutoDECRA/fi
			FCS_name_all	SPJ_Tonsil_28062021_ROI004_ROI_004.ome.biff' 19073x73 double
	Writing SPJ_Tonsil_28062021_R0I004_R0I_004.come			'Combined_dataset'
No details available	Writing Combined dataset		fileName	'CMW_Baseline_28062021_R01001_R01_001'
			ilastik_data	'Y:/Bioimaging_Staff/George/20230315_AutoDECRA/4
			ilastik_instruction	'headlessproject'
	丙→>	>	ilastik_model	'=Y:/Bioimaging_Staff/George/20230315_AutoDECRA\ '
	70000 1076 1177-8		RLF script	Ln 255 Col 1
P Type here to search			_	8°C Cloudy 스 및 🙀 4× 13:16 🔲
				31/07/2023

5. Run the script

References within supplemental notes

- 1. Mei HE, Leipold MD, Maecker HT. Platinum-conjugated antibodies for application in mass cytometry. Cytometry A 2016;89:292-300.
- 2. Chevrier S, Crowell HL, Zanotelli VRT, Engler S, Robinson MD, Bodenmiller B. Compensation of Signal Spillover in Suspension and Imaging Mass Cytometry. Cell Syst 2018;6:612-620 e5.
- 3. Berg S, Kutra D, Kroeger T, Straehle CN, Kausler BX, Haubold C, Schiegg M, Ales J, Beier T, Rudy M and others. ilastik: interactive machine learning for (bio)image analysis. Nat Methods 2019;16:1226-1232.
- 4. Carpenter AE, Jones TR, Lamprecht MR, Clarke C, Kang IH, Friman O, Guertin DA, Chang JH, Lindquist RA, Moffat J and others. CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biol 2006;7:R100.
- 5. Stirling DR, Swain-Bowden MJ, Lucas AM, Carpenter AE, Cimini BA, Goodman A. CellProfiler 4: improvements in speed, utility and usability. BMC Bioinformatics 2021;22:433.
- 6. Burbidge JB, Magee L, Robb AL. Alternative Transformations to Handle Extreme Values of the Dependent Variable. Journal of the American Statistical Association 1988;83:123-127.
- 7. Lin TH, Li HT, Tsai KC. Implementing the Fisher's discriminant ratio in a k-means clustering algorithm for feature selection and data set trimming. J Chem Inf Comput Sci 2004;44:76-87.
- Bruce Bagwell C. High-Dimensional Modeling for Cytometry: Building Rock Solid Models Using GemStone and Verity Cen-se' High-Definition t-SNE Mapping. Methods Mol Biol 2018;1678:11-36.
- 9. Wang Y. Understanding How Dimension Reduction Tools Work: An Empirical Approach to Deciphering t-SNE, UMAP, TriMap, and PaCMAP for Data Visualization Journal of Machine Learning Research 2021.
- Van Gassen S, Callebaut B, Van Helden MJ, Lambrecht BN, Demeester P, Dhaene T, Saeys Y. FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data. Cytometry A 2015;87:636-45.
- 11. Levine JH, Simonds EF, Bendall SC, Davis KL, Amir EAD, Tadmor MD, Litvin O, Fienberg HG, Jager A, Zunder ER and others. Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. Cell 2015;162:184-197.
- 12. Schapiro D, Jackson HW, Raghuraman S, Fischer JR, Zanotelli VRT, Schulz D, Giesen C, Catena R, Varga Z, Bodenmiller B. histoCAT: analysis of cell phenotypes and interactions in multiplex image cytometry data. Nat Methods 2017;14:873-876.

Table S1

Antibody	Clone	Vendor	Cat. #	Metal Tag	Working Conc. (ug/mL)	Major target cell type/population	Main spatial location (Protein Atlas)
CD45RO	UCHL1	Thermofisher	CAT#14-0457-82	115In	4	Activated/memory immune cells	Non-follicular, mantle zone
CD45RA	HI100	Thermofisher	CAT#14-0458-82	141Pr	4	Naïve immune cells	Widespread locations
CD68	KP1	Biolegend	CAT#372902	142Nd	5	Macrophages/monocytes	Widespread locations
CD8a	C8/144B	Biolegend	CAT#372902	143Nd	6	CD8 T cells	Non-follicular, mantle zone
KI67	Polyclonal	Novus	CAT#NB500-170	144Nd	3	Proliferating/mitotic cells	Follicles/GCs
Collagen I	3D5E8	Protein Tech	CAT#66761-1-lg	145Nd	2.5	Fibroblasts	Epithelium
CD138	4F3A8	Protein Tech	CAT#67155-1-lg	146Nd	3.5	Plasma B cells	Non-follicular, mantle zone
CD163	EDHu-1	BioRad	CAT#MCA1853	147Sm	10	Perifollicular macrophages	Epithelium and mantle zone
MPO	4C11F6	Protein Tech	CAT#66177-1-lg	150Nd	5	Neutrophils	Non-follicular, mantle zone
CD56	E7X9M	CST	CAT#99746BF	153Eu	12	NK/NK T cells	Non-follicular, mantle zone
CD69	15B5G2	Novus	CAT#NBP2-25236	155Gd	0.25	Memory T cells	Epithelium
EPCAM	Polyclonal	Abcam	CAT#ab71916	156Gd	1	Epithelial cells	Widespread locations
CD206	2A6A10	Abcam	60143-1-lg	157Gd	0.5	Tissue Macrophages and DCs	Epithelium and mantle zone
CD79a	EP3618	Abcam	CAT#ab239891	158Gd	7.5	B cells (all)	Widespread locations
STING	D2P2F	CST	CAT#13647	159Tb	5	T and NK cell subsets	Widespread locations
CD1c	2A7C11	Novus	CAT#NBP2-61726	162Dy	2.5	Dendritic cells (DCs)	Epithelium
IFITM3	Polyclonal	Protein Tech	CAT#11714-1-AP	163Dy	5	Immune cells	Epithelium and mantle zone
CD57	HNK-1	Biolegend	CAT#359602	165Ho	10	GC-resident T/B cells	Follicles/GCs
IL6-R	Polyclonal	Thermofisher	CAT#PA5-100836	167Er	10	Immune cells	Widespread locations
Cleaved Caspase-3	Asp175	CST	CAT#9579S	168Er	5	Apoptotic cells	Follicles/GCs
CD3	Polyclonal	Fluidigm	CAT#3170019D	170Er	7.5	T/NKT Cells	Widespread locations
CD31	EPR3094	Abcam	CAT#ab207090	172Yb	2	Endothelium	Epithelium and mantle zone
CD4	EPR6855	Abcam	CAT#ab181724	174Yb	6	CD4 T cells/Monocytes	Widespread locations
HLA-DR	LN3	Thermofisher	CAT#14-9956-82	175Lu	5	Antigen presenting cells (MHCII)	Widespread locations
CD169	SP213	Abcam	CAT#ab245735	176Yb	2.5	Monocytes	Widespread locations
CD147	E1S1V	CST	CAT#13287BF	194Pt	4	Various cells	Widespread locations
Beta-2-M	EPR21752-214	Abcam	CAT#ab237032	198Pt	1.5	All cells (MHC-I)	Widespread locations

CST = Cell Signalling Technologies; GC – Germinal centres

С



EPCAM



Multi-signal

В















Arcsinh c.f.1 then 0_1 normalisation



-5.48 0.53 6.54 UMAP_1 Arcsinh_0_1

11.97

Arcsinh c.f.1 then Z score normalisation



-4.836 -0.225 4.38 UMAP_1 Arcsinh_Zscore 4.386 8.997











11.02-



















CellPose

