Supplementary Note 1: Impact of image background subtraction on derived single-cell data.

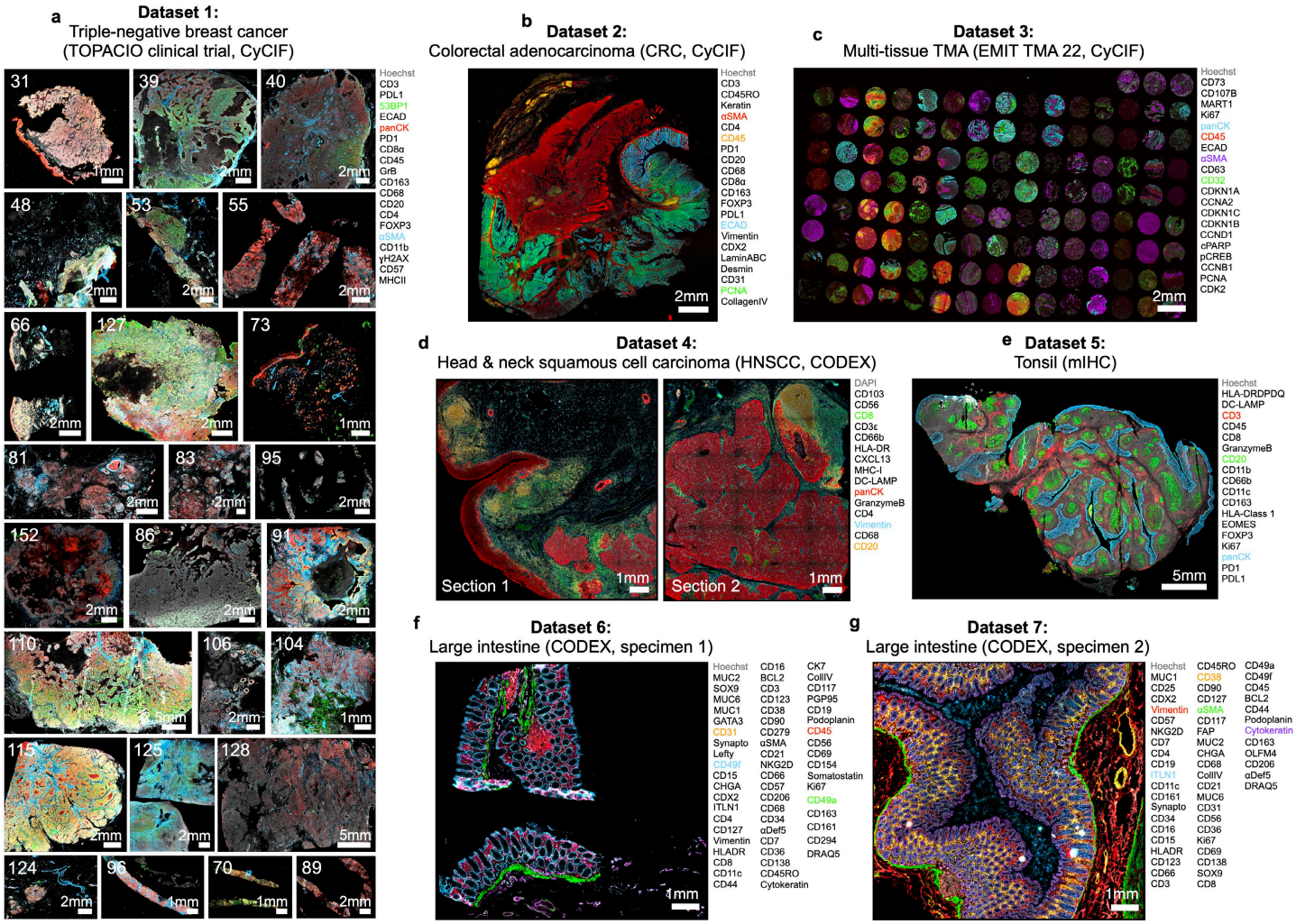
Background subtraction is commonly used with multiplexed imaging to remove autofluorescence and fluorescence arising from non-specific antibody binding to the specimen. However, we identified a number of challenges associated with this approach. For example, plotting histograms of the distribution of per-cell signal intensities channel in the pre-QC TOPACIO dataset revealed small numbers of cells with zero-valued signal intensities in all channels (**Supplementary Fig. 2a**). We reasoned that this effect was due to rolling ball image background subtraction[45] which was used to increase antibody signal-to-noise, but which had the unanticipated consequence of creating cells with signal intensities equal to zero that, after log-transformation, were far lower than values associated with other cells in the image. This effect was readily observed when the UMAP embedding was colored by channel signal intensity, as it revealed small clusters of extremely dim cells among much larger numbers of clusters whose signals were comparatively bright (**Supplementary Fig. 2b,c**). Using the panCK channel to better understand how cells with low signal intensities impacted the TOPACIO clustering result, we found that clusters within meta-cluster B (e.g., cluster 14) were exclusively composed of cells with zero-valued signals, while those in meta-cluster C (e.g., cluster 174) had signals that were all > 0, and those in meta-cluster F (e.g., cluster 197) were comprised of a mixture of cells with zero and non-zero signals (**Supplementary Fig. 2d**). The simple removal of cells with zero-value signal intensities from the pre-QC TOPACIO dataset (with no other quality control measures) eliminated small dark clusters characterized by very low signal intensities and significantly increased the resolution between immunopositive and immunonegative cell populations as seen in both the channel intensity histograms (**Supplementary Fig. 2e**) and UMAP embeddings colored by channel (**Supplementary Fig. 2f**). Resolution between positive and negative cells was further improved in the post-QC TOPACIO clustering after the removal of cells with near-zero signal intensities in addition to other artefacts (**Supplementary Fig. 2g,h**). This was also true of Dataset 6 (CODEX; **Supplementary Fig. 2i,j**). Thus, while background subtraction is useful for improving data quality, especially for low signal-to-noise antibodies, our analysis shows that it can skew the natural distribution of protein signals in an image and have a profound effect on the interpretation of single-cell data due to the spurious formation of irrelevant cell clusters. When using background subtraction, it is important to control for these problems.

Supplementary Note 2: Developing a DL model for automated artefact detection.

Although tools based on visual review are common in microscopy, there are obvious benefits to machine learning approaches[46–49]. To generate initial training data for a DL model to automatically flag arbitrary artefacts in multiplex IF images, three human annotators assembled ground truth artefact masks for 24 CyCIF channels in 11 serial tissue sections of the CRC dataset analyzed in this study (Dataset 2, **Supplementary Fig. 1b**). Single channel images (and their corresponding ground truth artefact masks) were cropped into 2048x2048-pixel image tiles. After class balancing, a total of 3,787 tiles were split 9:1 into training (3,409) and validation (378) sets. Tissue images differed with respect to the channels that were affected by artefacts (**Supplementary Fig. 3a**). The number of tiles containing artefacts also differed between images, ranging from as many as 463 tiles in image 59 to as few as 129 in image 64 (**Supplementary Fig. 3b**). Of the 3,787 total tiles, 1,734 contained pixels annotated as artefacts. Across all tiles, the average percentage of pixels affected by artefacts was ~6.7% (**Supplementary Fig. 1c**).

Our DL model comprised a pretrained ResNet34 encoder[50] coupled to a Feature Pyramid Network (FPN)[51] decoder (ResNet-FPN). The input of the model were image tiles and its output was predicted binary artefact masks. To assess the technical reproducibility of artefact predictions, three independent ResNet-FPN models were trained to convergence starting from FPN network weights initialized using different random seeds. Validation loss (measured via Dice similarity coefficient) ranged from 0.426 to 0.459 (mean = 0.444). To determine the ability of the trained models to generalize across different marker channels, testing was performed on channel 29 of tissue section 54 (**Supplementary Fig. 3d**), which contained artefacts not found in other sections or channels (**Supplementary Fig. 3a**). Performance was assessed by precision-recall (PR) and receiver operating characteristic (ROC) curve analysis. Average precision (AP) ranged from 0.30 to 0.33 for the three models (**Supplementary Fig. 3e**) and area under the ROC curve (AUC) ranged between 0.71 and 0.75 (**Supplementary Fig. 3f**). This demonstrates that the assembly of a DL model for artefact detection in high-plex tissue images is feasible. However, we judge the overall level of performance relative to human reviewers to be inadequate and we strongly suspect that this is due to insufficient training data. CyLinter is nevertheless an ideal way to generate additional training data. Thus, we have established a deposition site at the Synapse data repository (Sage Bionetworks, https://www.synapse.org/#!Synapse:syn24193163/wiki/624232) for collecting CyLinter-curated image artefacts. We anticipate that further training of our ResNet-FPN model on this corpus of collected artefacts will ultimately yield a highly-performant model for integration into future iterations of the CyLinter workflow.

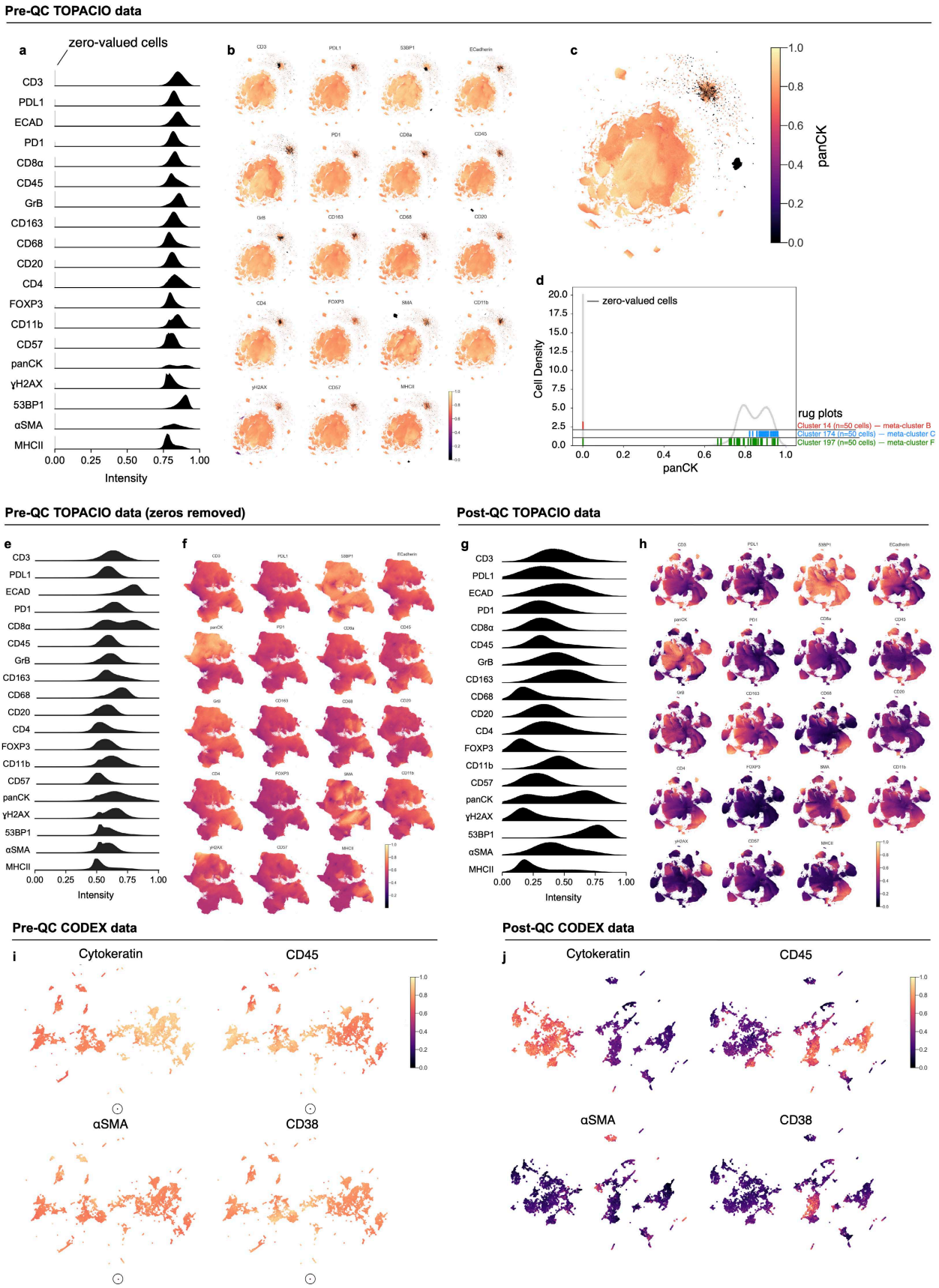
FIGURES/LEGENDS



Supplementary Fig. 1 | Overview of the seven multiplex IF datasets analyzed in this study.

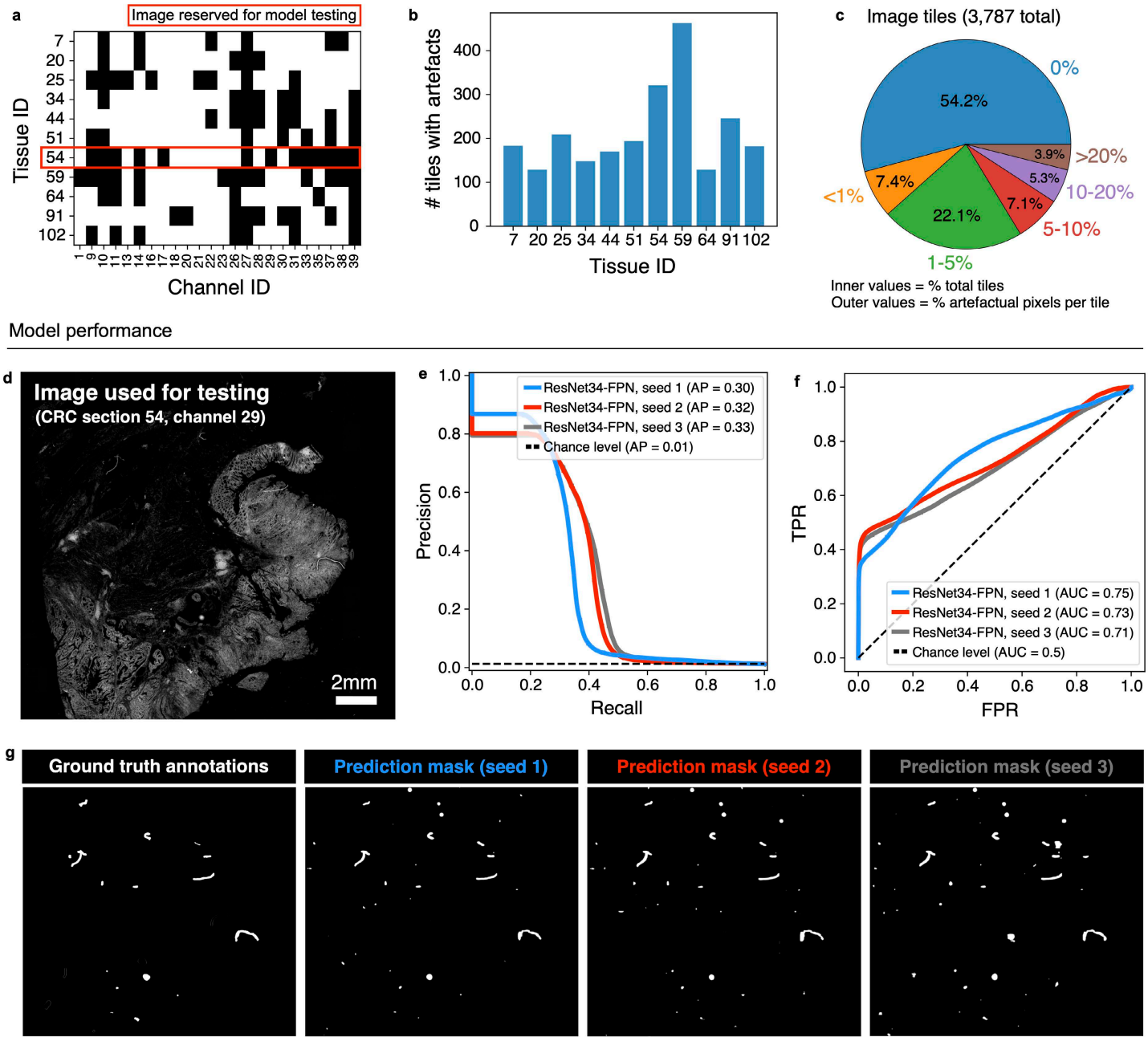
a, Dataset 1 (TOPACIO, CyCIF): 25 human TNBC clinical trial specimens (~6-353 mm^2). Numbers in upper left of each panel indicate specimen number. Channels shown are Hoechst (gray), 53BP1 (green), panCK (red), and αSMA (blue). **b**, Dataset 2 (CRC, CyCIF): an ~172 mm^2 whole-slide section of primary human colorectal adenocarcinoma. Channels shown are Hoechst (gray), αSMA (red), CD45 (orange), ECAD (blue), and PCNA (green). **c**, Dataset 3 (EMIT TMA22, CyCIF): 123 healthy and diseased human tissue cores each ~2 mm^2 arranged on a single microscope slide. Channels shown are Hoechst (gray), panCK (blue), CD45 (red), αSMA (purple), and CD32 (green). **d**, Dataset 4 (HNSCC, CODEX): two ~42 mm^2 whole-slide sections of human HNSCC. Channels shown are DAPI (gray), CD8 (green), panCK (red), vimentin (blue), and CD20 (orange). **e**, Dataset 5 (Tonsil, mIHC): an ~92 mm^2 whole-slide section of normal human tonsil. Channels shown are Hoechst (gray), CD3 (red), CD20 (green), panCK (blue). **f**, Dataset 6 (Large intestine, CODEX, specimen 1): an ~7 mm^2 whole-slide section of normal human large intestine from a 78-year-old African American male. Channels shown are Hoechst (gray), CD31 (orange), CD49f (blue), CD45 (red), CD49a (green). **g**, Dataset 7

(Large intestine, CODEX, specimen 2): an ~12 mm^2 whole-slide section of normal human large intestine from a 24-year-old white male. Channels shown are Hoechst (gray), Vimentin (red), ITLN1 (blue), CD38 (orange), αSMA (green), Cytokeratin (purple). Markers to the right of each dataset indicate the full marker set captured in the corresponding image(s). See **Supplementary Table 1** for specimen identifiers and data access information.



Supplementary Fig. 2 | Impact of image background subtraction on derived single-cell data. a, Ridge plots showing the distribution of cells according to channel signal intensities in the pre-QC TOPACIO dataset

showing the presence of zero-valued cells in each channel (vertical lines at far left). **b**, Channel colormaps applied to cells in the pre-QC TOPACIO embedding showing the presence of small, dark clusters corresponding to cells with at or near-zero signal intensities in the corresponding channel which by contrast makes all other cells appear bright for a given marker. **c**, PanCK channel from panel (b) enlarged to show detail. **d**, Histogram distribution of cells in the pre-QC TOPACIO dataset according to panCK signal. Rugplot plots (vertical ticks at bottom of histogram) show where randomly selected cells from cluster 14 (meta-cluster B, red), cluster 174 (meta-cluster C, blue), and cluster 197 (meta-cluster F, green) reside in the distribution. **e**, Ridge plots showing the distribution of cells according to channel signal intensities in the pre-QC TOPACIO dataset after the removal of zero-valued cells. **f,** Channel colormaps applied to cells in the pre-QC TOPACIO embedding after the removal of zero-valued cells showing that small, dark populations of cells are abrogated by the removal of zero-valued outliers. **g**, Ridge plots showing the distribution of cells according to channel signal intensities in the post-QC TOPACIO dataset allowing the natural distribution of signals to become apparent. **h,** Channel colormaps applied to cells in the post-QC TOPACIO embedding showing high contrast between populations of immunonegative and immunopositive cells for each marker. **i**, Channel colormaps applied to cells in the pre-QC CODEX embedding (Dataset 6) showing scant dim outliers (circles) which, in contrast, make other cells in the embedding appear bright for each marker. See **Online Supplementary Fig. 9** for full set of marker channels. **j**, Channel colormaps applied to cells in the post-QC CODEX embedding showing high contrast between immunopositive and immunonegative cell populations cells after dim outliers have been removed. See **Online Supplementary Fig. 10** for full set of marker channels.



Supplementary Fig. 3 | Developing a DL model for automated artefact detection tissue. a, Binary matrix showing the channels impacted by visual artefacts (e.g., illumination aberrations, slide debris, etc.) in 11 sections of the same CRC specimen. **b**, Barchart showing the number of 2048x2048-pixel image tiles affected by artefacts per tissue section. **c**, Pie chart showing the percentage of image tiles used for model training and validation (inner percentages) containing different percentages of artefactual pixels (outer percentages). **d**, CRC tissue section 54, channel 29 was used for model testing. **e**, Precision-recall plot showing the average precision (AP) for three replicates of the ResNet-FPN model architecture whose FPN network was initialized with different model weights to evaluate technical reproducibility. **f**, Receiver operating characteristic (ROC)

curve showing the area under the curve (AUC) values for the same three replicates of the ResNet-FPN model shown in panel (e). **g**, Ground truth artefact mask (far left) and predicted artefact masks from the three replicate ResNet-FPN models.

Online Supplementary Fig. 1 | Example artefacts in Dataset 1 (TOPACIO) (https://www.synapse.org/#!Synapse:syn53781614). **a**, Twelve (12) examples of tissue folds. **b**, Twelve (12) examples of slide debris. **c**, Four (4) examples of coverslip air bubbles. **d**, Twelve (12) examples of image blur.

Online Supplementary Fig. 2 | Image galleries of clustering cells from pre-QC Dataset 2 (CRC) (https://www.synapse.org/#!Synapse:syn53781627). Twenty (20) examples of cells from each of 22 clusters identified in the pre-QC CRC dataset showing the top three most highly expressed markers (1: green, 2: red, 3: blue) and Hoechst dye (gray). A single white pixel at the center of each image highlights the reference cell. Nuclear segmentation outlines are superimposed to show segmentation quality.

Online Supplementary Fig. 3 | Image galleries of clustering cells from pre-QC Dataset 6 (CODEX) (https://www.synapse.org/#!Synapse:syn53781635). Twenty (20) examples of cells from each of 32 clusters identified in the pre-QC CODEX dataset (normal large intestine, specimen 1) showing the top three highly expressed markers (1: green, 2: red, 3: blue) and Hoechst dye (gray). A single white pixel at the center of each image highlights the reference cell. Nuclear segmentation outlines are superimposed to show segmentation quality.

Online Supplementary Fig. 4 | Image galleries of clustering cells from pre-QC Dataset 1 (TOPACIO) (https://www.synapse.org/#!Synapse:syn53782191). Twenty (20) examples of cells from each of 48 (of 492) clusters identified in the pre-QC TOPACIO dataset showing the top three most highly expressed markers (1: green, 2: red, 3: blue) and Hoechst dye (gray). A single white pixel at the center of each image highlights the reference cell. Nuclear segmentation outlines are superimposed to show segmentation quality.

Online Supplementary Fig. 5 | Image tiles from Dataset 1 (TOPACIO) (https://www.synapse.org/#!Synapse:syn53779745). Down-sampled, single-channel images from the 25 TNBC tissue specimens analyzed in this study used to estimate the number of image tiles impacted by microscopy artefacts. Artefact counts table and patient metadata table are also provided.

Online Supplementary Fig. 6 | Image galleries of clustered cells from post-QC Dataset 2 (CRC) (https://www.synapse.org/#!Synapse:syn53781719). Twenty (20) examples of cells from each of 78 clusters identified in the post-QC CRC dataset showing the top three most highly expressed markers (1: green, 2: red, 3: blue) and Hoechst dye (gray). A single white pixel at the center of each image highlights the reference cell. Nuclear segmentation outlines are superimposed to show segmentation quality.

Online Supplementary Fig. 7 | Image galleries of clustered cells from post-QC Dataset 6 (CODEX) (https://www.synapse.org/#!Synapse:syn53781730). Twenty (20) examples of cells from each of 28 clusters identified in the post-QC CODEX dataset showing the top three most highly expressed markers (1: green, 2: red, 3: blue) and Hoechst dye (gray). A single white pixel at the center of each image highlights the reference cell. Nuclear segmentation outlines are superimposed to show segmentation quality.

Online Supplementary Fig. 8 | Image galleries of clustered cells from post-QC Dataset 1 (TOPACIO) (https://www.synapse.org/#!Synapse:syn53781892). Twenty (20) examples of cells from each of 43 clusters identified in the post-QC TOPACIO dataset showing the top three highly expressed markers (1: green, 2: red, 3: blue) and Hoechst dye (gray). A single white pixel at the center of each image highlights the reference cell. Nuclear segmentation outlines are superimposed to show segmentation quality.

Online Supplementary Fig. 9 | Channel colormaps applied to cells in the pre-QC Dataset 6 (CODEX) embedding (https://www.synapse.org/#!Synapse:syn53781812).

Online Supplementary Fig. 10 | Channel colormaps applied to cells in the post-QC Dataset 6 (CODEX) embedding (https://www.synapse.org/#!Synapse:syn53781819).