A Platform-Independent Framework for Phenotyping of Multiplex Tissue Imaging Data

Mansooreh Ahmadian[1], Christian Rickert[2], Angela Minic[2], Julia Wrobel[1], Benjamin G Bitler[3], Fuyong Xing[1], Michael Angelo[4], Elena W Hsieh[5], Debashis Ghosh[1], Kimberly R Jordan[*2]

[1]Department of Biostatistics and Informatics, Colorado School of Public Health, University of Colorado-Denver Anschutz Medical Campus, Aurora, CO
[2]Department of Immunology and Microbiology, University of Colorado-Denver Anschutz Medical Campus, Aurora, CO
[3]Division of Reproductive Sciences, Department of OB/GYN, University of Colorado-Denver Anschutz Medical Campus, Aurora, CO
[4]Department of Pathology, Stanford University, Stanford CA,
[5]Pediatrics, Section of Allergy and Immunology, University of Colorado-Denver Anschutz Medical Campus, Aurora, CO
[*]Correspondence: Kimberly Jordan (KIMBERLY.JORDAN@CUANSCHUTZ.EDU)

Abstract

Multiplex imaging is a powerful tool to analyze the structural and functional states of cells in their morphological and pathological contexts. However, hypothesis testing with multiplex imaging data is a challenging task due to the extent and complexity of the information obtained. Various computational pipelines have been developed and validated to extract knowledge from specific imaging platforms. A common problem with customized pipelines is their reduced applicability across different imaging platforms: Every multiplex imaging technique exhibits platform-specific characteristics in terms of signal-to-noise ratio and acquisition artifacts that need to be accounted for to yield reliable and reproducible results. We propose a pixel classifier-based image preprocessing step that aims to minimize platform-dependency for all multiplex image analysis pipelines. Signal detection and noise reduction as well as artifact removal can be posed as a pixel classification problem in which all pixels in multiplex images can be assigned to two general classes of either I) signal of interest or II) artifacts and noise. The resulting feature representation maps contain pixel-accurate representations of the input data, but exhibit significantly increased signal-to-noise ratios with normalized pixel values as output data. We demonstrate the validity of our proposed image preprocessing approach by comparing the results of two well-accepted and widely-used image analysis pipelines.

Introduction

Multiplex tissue imaging technologies such as Multiplexed Ion Beam Imaging (MIBI) [1, 2], Imaging Mass Cytometry (IMC) [3], CO-Detection by indEXing (CODEX)[4, 5], Multiplexed Immunofluorescence (MxIF) [6, 7], and cyclic Immunofluorescence (cycIF) [8, 9] provide researchers with a wealth of information on the single cell level that illustrates the complexity and heterogeneity of tissue samples. By preserving the spatial context of >40 markers measured simultaneously on a tissue, these advanced technologies have opened new avenues for biological discoveries in healthy and diseased microenvironments [10–18]. A substantial challenge remains, however, in developing accurate, robust, and automated computational pipelines for the analyses and interpretation of these complex high-dimensional imaging data.

A high-dimensional tissue image consists of a set of antibody-based visualizations of multiple parameters (markers) measured with fluorescence or mass spectrometry readouts. The signal intensity in each image channel is directly proportional to the expression level of the corresponding marker bound to its target. Since

expression of specific marker combinations is the key determinant of cellular phenotypes, a primary step in the identification of cellular phenotypes is to accurately quantify signal intensities across channels. Accurate quantification of these signal intensities can be hindered by instrumental noise, acquisition artifacts, or experimental variability that are often incorporated in the imaging data during sample preparation and data acquisition. Therefore, most platforms require denoising and artifact removal techniques prior to accurate quantification of signal intensities [2, 19–21]. Remaining artifacts may have far-reaching consequences in downstream analyses, potentially leading to inaccurate cellular identification and false conclusions in statistical comparisons or spatial analyses. Particularly for high-parameter imaging techniques that rely on unsupervised clustering algorithms for cell-type identification, high-quality images are required for robust and accurate quantitative outputs. The data must also be appropriately normalized to limit tissue and batch variability. Without a proper normalization, sample-to-sample intensity variations can cause cells to cluster by individual samples rather than by cell types [22, 23]. However, choosing an appropriate normalization method is a critical task, as the accuracy of cell-type identification has been shown to depend more on the choice of normalization approach than on the clustering algorithm [24]. Therefore, high-quality imaging data should ideally satisfy the following criteria: 1) the data should exhibit high signal-to-noise ratios (SNR) and be as free as possible from artifacts; and 2) be appropriately normalized to remove non-biological signal variability within and across acquisition batches and tissues.

Various image processing and filtering methods can be used for denoising and artifact removal [2, 25–29]. For instance, variants of Gaussian filters are used for noise reduction [30–32]. Binary masks can be generated to specifically remove artifacts that differ from real signal by size and pixel distribution. Recently, a five-step computational pipeline was proposed to prepare MIBI data for downstream analysis [2]. Steps in the pipeline include background subtraction, necrosis removal, batch normalization, denoising, and aggregate removal. In the first step, non-specific background is removed by subtracting counts from pixels where background noise is present. To specify those pixels with background noise, a binary mask (which is 1 where background exists and 0 elsewhere) is generated by filtering and thresholding a blank channel that is specifically included in the data acquisition process for this purpose. Both the thresholding parameter and the number of counts subtracted are determined manually based on visual inspection by the user. In the second step, artifacts such as areas of necrosis are removed by estimating another binary mask for the necrotic region using morphological opening and closing [2]. The number of counts subtracted and the thresholding parameters are decided by the user. Batch effects are then removed in the third step using quantile normalization. In the fourth step for denoising, a k-nearest-neighbor approach is used to generate a binary mask that separates the noisy pixels based on density. The parameter k and the thresholding parameter that separates the high-density pixels (signal) from the low-density pixels (noise) as well as the subtracting pixel counts are determined by the user. In the last step, the images are first smoothed using a Gaussian kernel and then binarized using Otsu's method [33]. This five-step pipeline was recently wrapped in a graphical user interface called MAUI (Mass based imaging Analysis User Interface) [25].

While MAUI facilitates MIBI image preprocessing, it relies on accurate estimation of binary masks with user-defined thresholding parameters that are prone to bias. Thresholding methods separate two distinct features using a single value as the thresholding parameter. However, given the variabilities in complex multiplex tissue imaging data, a single threshold is unlikely to obtain a parameter configuration that perform well across all the images in a dataset. Additionally, estimating binary masks for each artifact in several sequential steps is labor-intensive and not scalable. Finally, new artifacts may appear as data are acquired in new tissue types or with new technologies, requiring additional steps to be added to the aforementioned five-step pipeline.

Here, we aim to create a simple automated framework that provides a single-step unified solution that works across imaging platforms. Our approach combines denoising and removal of various artifacts into a *single* pixel classification step, outputs a feature representation map (FR map) that eliminates the need for any further normalization process, does not rely on estimation of any thresholding parameters and therefore is more robust, and provides a unified solution across multiplex imaging platforms (both mass spectrometry and fluorescence-based imaging). We validate our proposed approach by comparing our results with two well-accepted and widely-used baselines: a) inForm software, a commercial software package for
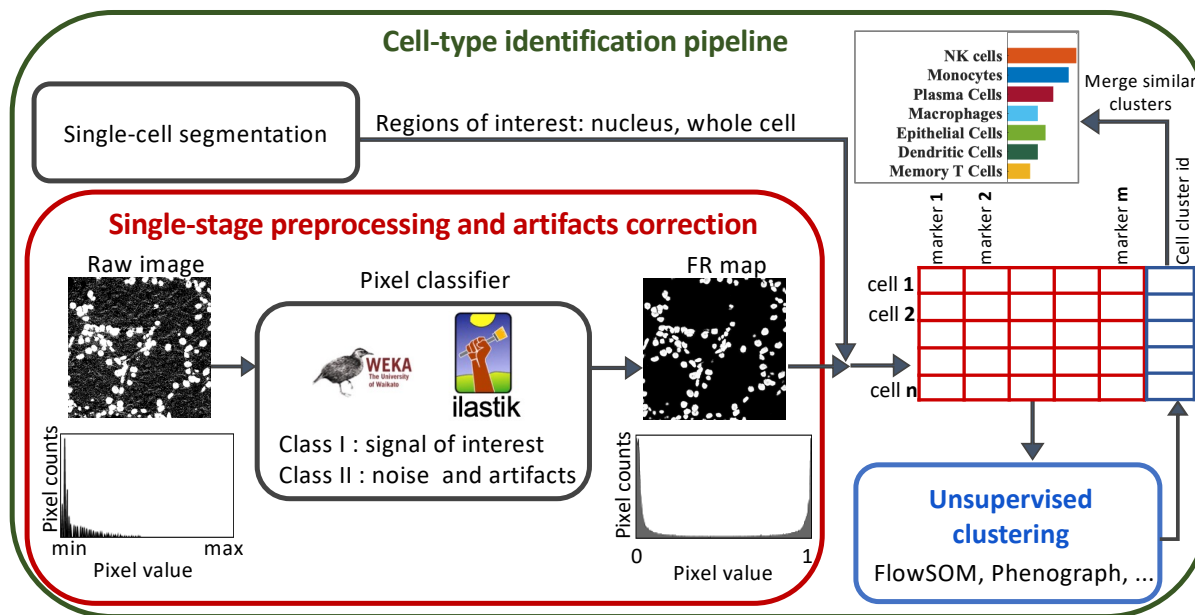
Figure 1: **A general computational pipeline for cellular phenotyping of multiplex tissue imaging data.** We replaced image denoising and preprocessing with our proposed framework as shown in the red box. A pixel classifier is used to classify the pixels in the raw image into two classes: I- desired signal, or II- noise and artifacts. Weka [34] and ilastik [35, 36] are well-established computational tools that can be used to perform the pixel classification. The output of the classifier are two feature representation maps, one for each class, with pixel values between 0 and 1. Marker expression within the border of each cell is then measured from the class I FR maps. The measured single-cell information data (a table with cells in rows and marker expression level in columns) is then used as the input for unsupervised clustering algorithms to identify the cell types.

fluorescence-based multiplex imaging analyses and b) MAUI [25], a publicly available computational pipeline used to analyze MIBI data. All code and data have been made available as a resource for the research community [33].

Results

A single-step platform-independent framework for preprocessing of multiplex tissue imaging data

Our proposed framework for cellular phenotyping of multiplex tissue imaging data formulates denoising and artifact removal as a pixel classification problem in which the pixels in an image are classified into two classes (Fig. 1):

- class I: pixels with signal of interest (marker signal/positive signal).

- class II: pixels with noise, artifacts, and other platform-specific properties that are not useful for downstream analyses.

The input of the classifier is a training set (labeled data) comprised of user-provided examples for both class I (desired signal) and class II (noise and artifacts) pixels. Features can be extracted by convolving various kernels with the image such as noise reduction filters, edge detectors, and texture kernels. Classifying algorithms, such as Random Forest [37], can then be used to classify each pixel given the extracted features. Several well-established and user-friendly tools have been developed for pixel classification such as Trainable

Weka Segmentation (TWS) [34], or ilastik [35] and can be used to interactively label training data for signal and noise during image preprocessing. The output of the classifier consists of two FR maps, one for each class. The pixel values in each FR map represent the probability of that pixel belonging to the corresponding class. Pixels with noise and artifacts have either zero or very low values in the FR map of class I, meaning FR maps of class I are free from noise and artifacts. In addition, the values in the FR maps range between 0 to 1. Therefore, the FR maps are normalized and this eliminates the need for further normalization processes.

Noise and artifact removal from multiplex imaging data

Image noise is generally defined as unwanted random variation that obscures the desired information in an image. Multiplex images are visualizations of bound antibodies, detected either by heavy metal ions conjugated to antibodies in mass spectrometry-based imaging or by fluorescent particles directly or indirectly associated with antibodies in light microscopy-based imaging. Thus, noise in multiplex imaging can be defined as any signal that differs from the biological structures to which antibodies are specifically bound [25]; and it can be classified into several categories: 1) channel crosstalk, 2) nonspecific antibody staining, and 3) aggregates.

Channel crosstalk or cross-channel contamination is the variable presence of signal from a contaminating channel in a target channel. For fluorescence microscopy imaging techniques, channel crosstalk is caused by the wide and often overlapping emissions spectra of fluorochromes. For mass-spectrometry imaging techniques, impurities in the heavy metal ion source used for antibody labeling, modification of the heavy metal ions (by hydrogenation, oxygenation, hydroxylation, etc), or for MIBI specifically, gold ions from the slide surface can introduce channel crosstalk to the image. For example, a low intensity contaminating signal that mirrors the structure of the real higher intensity signal in the Vimentin-163 channel (Fig. 2a-top) is visible in the neighboring SMA-164 channel (Fig. 2a-middle). To correct this cross-talk, a few examples of the contaminating Vimentin and real SMA signals were used to train the pixel classifier and generate an FR map for the positive signal (Fig. 2-bottom), in which the contaminating Vimentin signal is removed. As demonstrated in the histograms (Fig. 2a), the SNR is significantly enhanced in the FR map (green box) compared to the raw image (red box). That is, the values of pixels with positive signal in the FR map group to the far right of the histogram, whereas the rest of the pixels, including those with cross-talk contamination, group to the far left of the histogram and thus are cleanly separated. Another example of cross-talk contamination is gold background (Fig 2b) where signal from the exposed part of the gold slide (contaminating channel) is observed in the target channel, here Ki-67 (Fig 2b-left). Our approach allows easy removal of this artifact by training the pixel classifier with a few examples of contaminating pixels (pixels with low intensity that correlate with the bare gold slide) as well as the real Ki-67 signal. The contaminating artifact is completely removed in the resulting FR map (Fig 2b-right) and the Ki-67 signal is clearly separated from the noise pixels, as shown in the histograms.

Non-specific antibody binding is caused by cross-reactivity of antibodies, tissue features that non-specifically bind to antibodies, or instrumental noise. This artifact can appear as either random signal or as dim patterns superimposed on top of biological structures and may have correlation with the histological structure of the tissue [25]. Regardless of the source, the intensity and spatial distribution of this artifact is different from the true positive signal. Therefore, regions with signal from non-specific antibody binding can be easily labeled as undesired signal (class II) using our approach. An example of non-specific antibody staining is necrosis where a necrotic tissue region that exhibits non-specific signal in many channels appears in the pan-cytokeratin channel of a breast cancer tissue (Fig 2c-left). After pixel classification, the artifact is removed in the resulting FR map (Fig 2c-right) and the signal to noise ratio is enhanced, as shown in the histograms.

Aggregates are small high-intensity specks of signal caused by aggregation of antibodies, secondary reagents, or fluorochromes used in the staining process. Due to the concentration of high-intensity pixels, aggregates can be falsely interpreted as positive signal. Figure 2d shows an example of CD163 staining surrounded by noise and aggregates (left). To compare the impact of signal caused by aggregates, pseudo-cells of the same area were drawn by hand (610 pixels) and the mean expression of CD163 was calculated. While area A (m=1.18) is clearly a CD163+ cell, the mean signal for area B (m=0.43), which is likely a CD163+
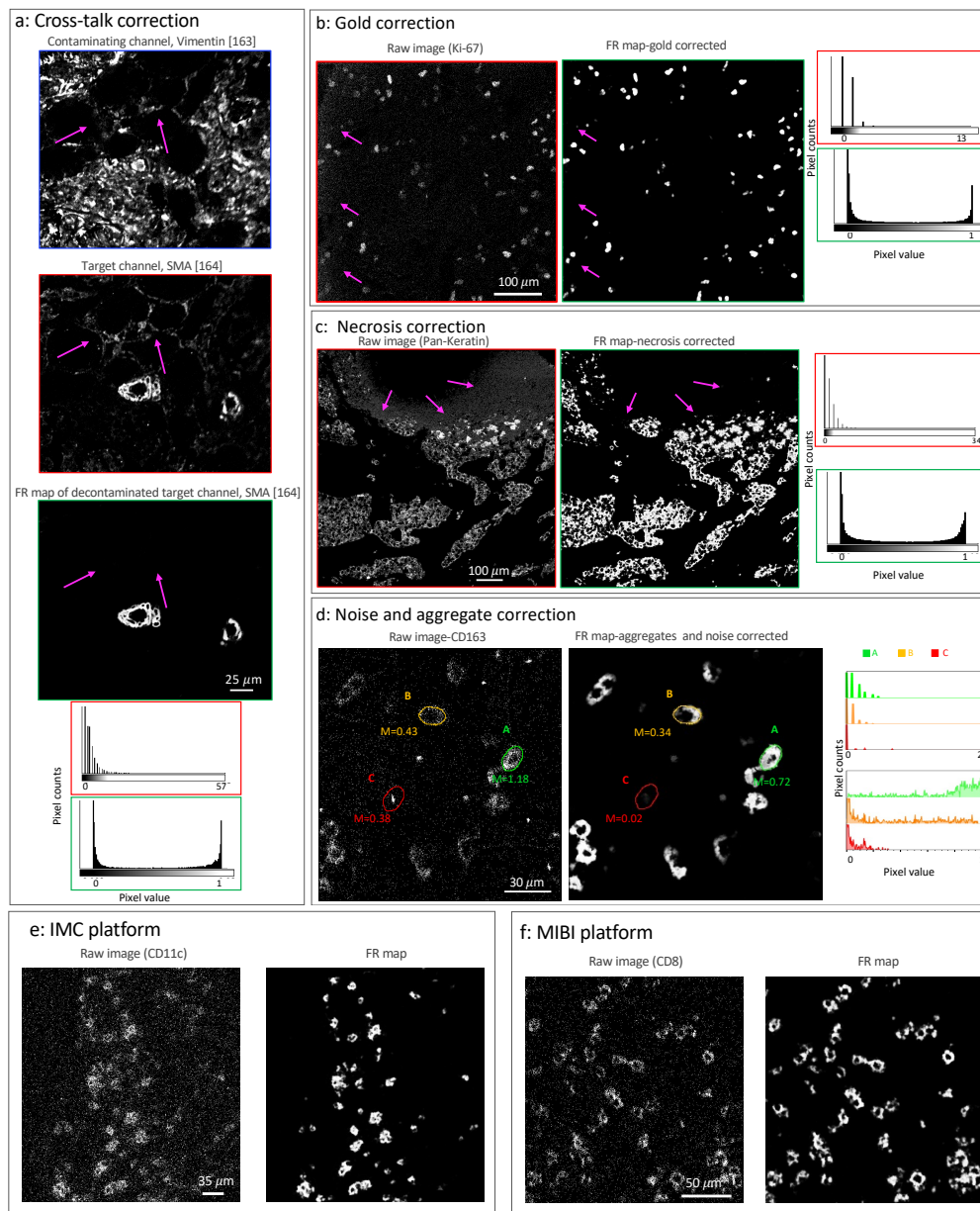
Figure 2: **Removal of various artifacts and noise from MIBI data.** **a**: An example of cross-channel contamination where a contaminating channel, Vimentin with mass isotope of 163 (left), contaminates a target channel, SMA with mass isotope of 164 (middle). Our approach removes this contamination by generating an FR map for the target channel (right). **b**: Section of Ki-67 marker from ovarian cancer tissue before (left) and after(right) gold removal (a MIBI platform artifact). **c**: Section of breast tissue stained with pan-cytokeratin antibodies before (left) and after (right) removal of necrotic tissue regions. Histograms of pixel values are included for all images (a-c) before and after artifacts correction. **d**: Section of ovarian cancer tissue stained with CD163 antibody before (left) and after noise and aggregates (right) correction. Corresponding histograms of pixel values for the pseudo-cells outlined in green, orange, and red are plotted before and after noise and aggregate correction. **e**: CD11c staining of lung tissue by IMC (left) and the corresponding FR map (right). **f**: CD8 staining of ovarian cancer tissue by MIBI (left) and the corresponding FR map (right).

5

cell, is difficult to distinguish from area C (m=0.38) that contains positive signal due to an aggregate in the raw image (left). However, after removing noise and aggregates (Fig. 2d-right), the mean signal for area C is greatly reduced (m=0.02) while the mean signal for area B remains high (m=0.34). Increased SNR after noise and aggregate removal is also demonstrated in the overlaid histograms, where the pixel intensities for each pseudo-cell are plotted for the raw image (top) and the FR map (bottom). Therefore, our approach can prevent identification of false positive cells from aggregate signal.

Mass based technologies generate pixelated imaging data with low SNR compared to fluorescence-based technologies. Another important advantage of our framework is that the sparse and pixelated signal of Mass-based imaging data (Fig. 2d-left) is converted to a continuous signal with high SNR (Fig. 2d-right). This property makes the processing of the such data a challenging task because the pixel intensity alone does not carry sufficient information to separate positive signal from noise. The pixel density should be considered along with the intensity [2]. That is, while area A (Fig. 2d-left) can be effortlessly called a CD163+ cell given its consistent and high pixel intensity, the signal intensity in area B is much lower and it is the density of the pixels that contributes to forming a cell-like structure. The low intensity signal values of the individual pixels in area B are difficult to distinguish from noise in the raw image (see the corresponding histograms in Fig. 2d). However, the spatial information of the pixels contributes to the extracted features in the FR map and thus the classification output. As a result, low-density and low-intensity pixels of noise have very low pixel values in the FR map, while high-density and low-intensity pixels of area B have high pixel values (see the Fig. 2d-right and the corresponding histograms). For further illustration, two additional examples of converting mass-based imaging raw data from IMC and MIBI platforms to FR maps are shown fig. 2(e-f).

Accurate phenotyping of tissue imaging data across platforms

To further evaluate the performance of our image preprocessing framework, we compared the results of unsupervised clustering after denoising and artifacts correction using two previously published datasets collected by MIBI and Vectra Polaris. The MIBI data are a set of publicly available images from a triple-negative breast cancer cohort (TNBC) with 41 patients [2]. This dataset was analyzed by a computational pipeline [2] that was later wrapped into a user-friendly graphical interface, MAUI [25]. The Vectra dataset includes a subset of 6 patients with ovarian cancer collected in-house and analyzed using inForm [38, 39], a widely-used commercial software package for analyses of multiplex fluorescence imaging data.

We compared the output of our analysis framework with the MAUI pipeline using t-SNE plots comprised of about 200,000 cells from the MIBI breast cancer dataset (Fig. 3a, Fig S1-a). The overlaid heatmaps indicate the signal intensities for a subset of non-immune (Keratin6, CD31, SMA, Vimentin) and immune (CD45, CD3, CD4, CD8, CD20, CD68) markers measured from the raw images (top row) or FR maps (middle row). FR maps generated signal intensities ranging between 0 and 1 with improved SNR, as indicated by the more differentiated positive regions in the heat map. The scatter plots (bottom row) demonstrate the non-linear mapping of the average expression values per cell measured from the raw image (x-axis) and the FR maps (y-axis). We then compared the outcomes of unsupervised clustering (Fig. 3b) using normalized marker expression scaled from zero to one (left) or measured from FR maps (right). To directly compare the two heatmaps and avoid discrepancies introduced by segmentation or clustering, we used the same single-cell segmentation maps and the same cellular phenotypes identified in the TNBC study [2]. Qualitative comparison of the two heatmaps confirms that similar cell populations can be identified using FR maps. For quantitative comparison, we applied the same unsupervised clustering algorithm, FlowSOM [40], to the single-cell information extracted using FR maps. The applicability of our analysis pipline is confirmed by the significant correlation in the counts of different cell types identified on a patient-level in the TNBC study and using our framework (Fig 3c). Furthermore, there is an 84% agreement between the predicted cell types in a comparison of individual cells (Fig.S1 b-c). The discrepancies between the predictions of the two pipelines can be further reduced by manual quality control of the clustering results.

Next, we analyzed ovarian cancer fluorescence imaging data from 6 patients (about 30,000 cells) using inForm software as a benchmark. To avoid any discrepancies caused by differences in the single-cell segmentation, we used the same cell segmentation maps produced by inForm. Figure 4a shows the signal intensities measured from the raw images (top row) or FR maps (middle row) overlaid on tSNE plots. Scatter plots
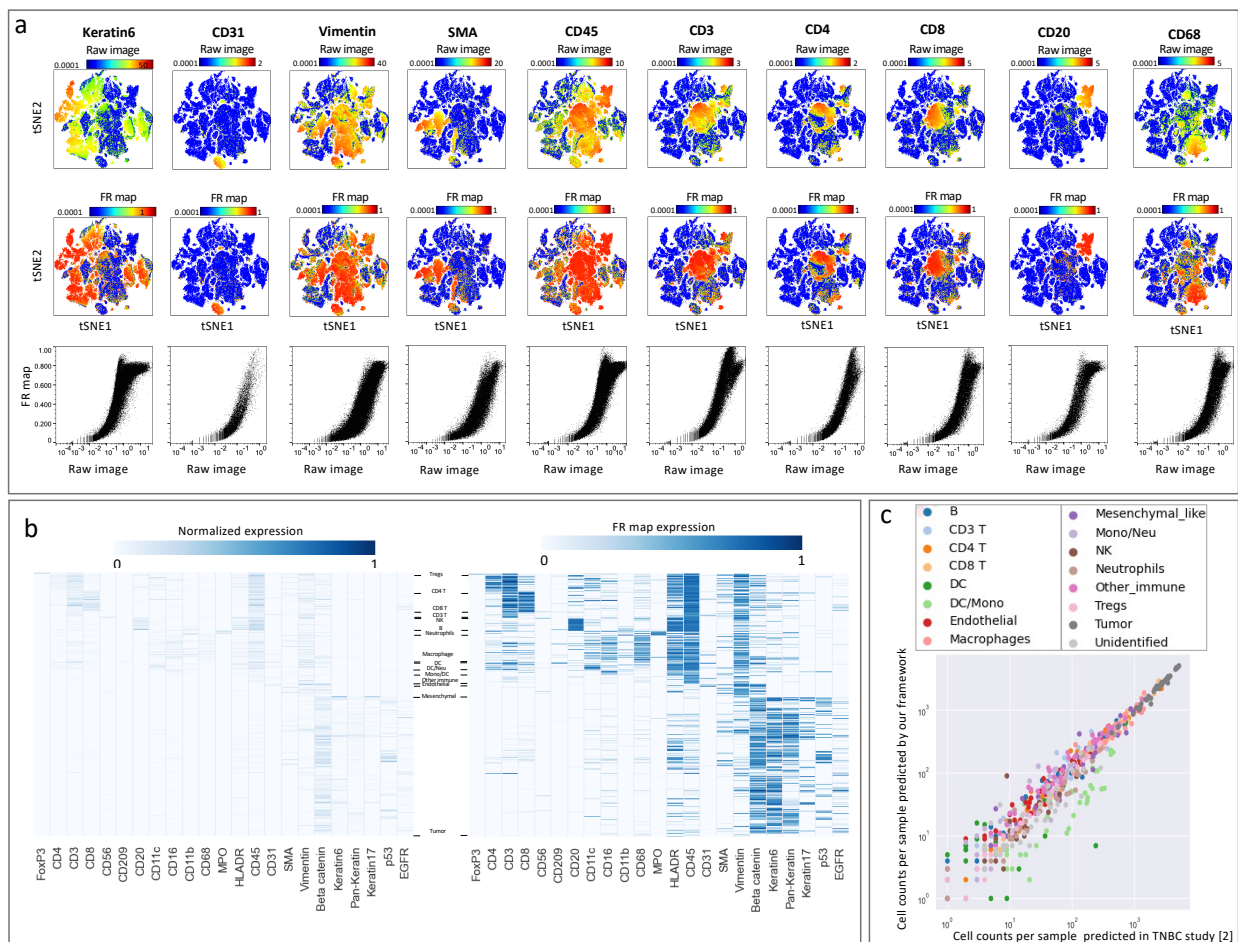
Figure 3: **Our framework delineates cell type compositions in mass ion beam imaging data consistent with MAUI[2]. a,** Marker expression measured per cell using the images (top row) and the FR maps (middle row) are overlaid on the tSNE plot for selected immune and tumor markers. The bottom row demonstrates correlation between marker expression per cell from raw images (x-axis) and FR maps (y-axis). **b,** Cell types clustered by marker expression. Expression values for each marker are scaled from zero to one (top) or are measured from the FR maps (bottom). **c,** Cell-cell comparison between the frequency of each cell type per patient predicted using MAUI in the TNBC study (x-axis) and our proposed framework (y-axis).

(bottom row) visualize the non-linear mapping of the average expression values per cell measured from the raw image (x-axis) to the FR maps (y-axis) and demonstrate the correlation of signal intensities between the methods. We extracted single-cell information from the FR maps as input to FlowSOM [40] and identified tumor cells and four groups of immune cells. Figure 4b shows a cell-cell comparison of the cell types identified by inForm (x-axis) and the cell types identified using our framework (y-axis). While there is 84% agreement between cellular phenotypes identified by the pipelines (shown in green), about 12% of the cells are unidentified by inForm but assigned to a cell type by our pipeline (shown in red). Many of the cells in inForm's unidentified cluster (Fig. 4b-right) have measurable signal intensities that are more clearly defined in the measured counts from the FR map. Therefore, our pipeline was better able to discern the phenotypes of these unidentified cells with enhanced SNR. To further investigate these discrepancies, we generated a color overlay of markers from the raw image and the corresponding FR map (Fig. 4top). Comparing the pseudo-coloring of cell populations predicted by inForm and our pipeline (Fig. 4bottom) with the protein
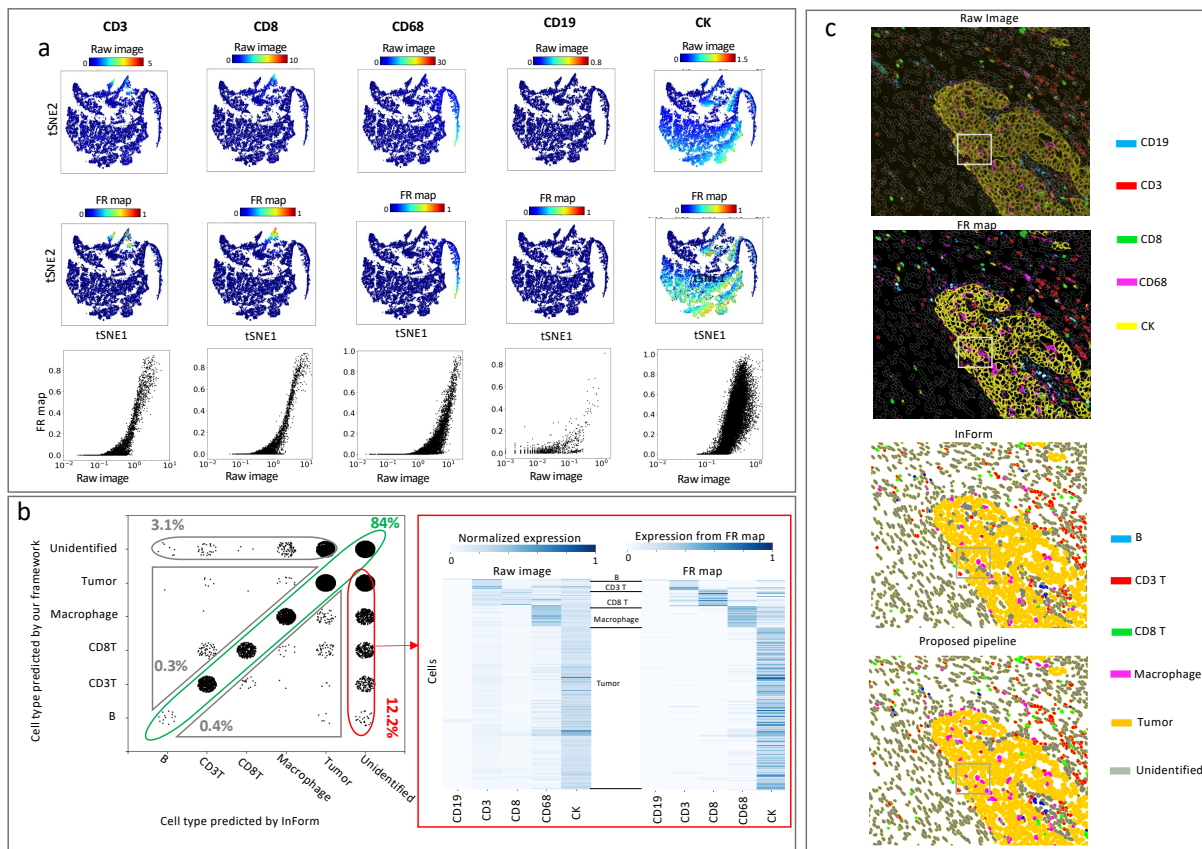
7

Figure 4: **Our framework delineates cell-type composition in fluorescence imaging data consistent with inForm. a,** Signal intensities for selected immune and tumor markers measured for individual cells using the raw images (top row) or the FR maps (middle row) were overlaid on tSNE plots. Scatter plots (bottom row) demonstrating the correlation between signal intensity per cell from raw images (x-axis) and FR maps (y-axis). **b,** Cell-cell comparison between the cell type predicted by inForm (x-axis) and our framework (y-axis). Heatmap of marker expression for the unidentified cluster by inForm (right); the expression level of markers per cell is measured using the raw image scaled between 0 and 1 (left) or measured from FR maps (right). In both heatmaps the expression level is computed as the summation of pixel values within the boundary of a cell divided by the total number of pixels comprising that cell. **c,** Color overlay of markers CD19, CD3, CD8, CD68, and CK (top); plots compare the stain from the raw image with the corresponding FR map (top). Pseudo-coloring of cell populations compares the predicted cell types by inForm with our framework (bottom).

expression on top, we can easily spot many CK+ tumor cells that are unassigned by inForm (Fig.4c).

Discussion

A goal of image quantification pipelines is to remove noise and artifacts and normalize signal intensities to eliminate batch variations so that imaging data can be combined. The amount of manual curation, dependency on user-defined thresholding parameters, and lack of a unified solution that works across different imaging platforms slows analyses and hinders robust and reproducible results. We developed a framework that overcomes these challenges by replacing the multi-step low-level image processing with a single-step pixel classification. We classify each image such that all categories of undesired signal are placed in a different class from the desired marker signal and continue the downstream analysis using the generated

feature representation map of the marker signal. This pipeline opens a new window for integrating data across multiplex imaging platforms and constructing and training generalizable cell-type annotators that can be used in a clinical study with patient and experimental variation.

Using FR maps instead of the raw images has several advantages beyond noise and artifact removal. First, the positive signal intensity values are mapped to normalized values between 0 and 1 in the FR map. Therefore, the artifact removal and normalization are combined into a single step and the normalization is completed at the image level in a very early stage of the analysis.

In addition to noise and artifacts, the platform-specific properties of images such as differences in signal-to-noise ratio, autofluorescence, and background staining are removed from the data. These platform-specific features do not carry useful information and only play confounding roles in downstream analysis. For instance, platform-specific features prevent single-cell segmentation models trained using MIBI data from generalizing on fluorescence-based imaging data [41]. Using our method, discretized and low-SNR signal of MIBI data is converted to continuous and high-SNR signal using FR maps, making the images more similar to fluorescence-based imaging data and making it easier to identify phenotypes. The general idea of pixel classification works across all imaging platforms and thus provides a unified solution for all the preprocessing required for robust cell clustering. Our platform-independent solution eliminates the need for alternative approaches that would otherwise be chosen according to which works best for each imaging technology. For instance, while fluorescence-based imaging data can be denoised using intensity-based methods, MIBI data require density-based methods for denoising.

Previous computational methods for denoising and artifact removal require substantial manual curation and parameter tuning and there is little consensus about which denoising pipeline produces high quality and reproducible results [42]. Many parameters of the algorithms used in the existing pipelines were tuned by hand and evaluated visually, leveraging the expert knowledge of pathologists and biologists. In addition, based on the level and composition of noise, custom-built approaches were optimized by the investigators for different imaging data [42]. As a result, those methods are prone to human bias and may lack reproducibility. Our proposed method provides a unified solution for denoising of MIBI data by redefining the problem as a pixel classification problem, where the final value of a pixel is decided by an algorithm based on a variety of imaging features extracted from the raw data instead of thresholding parameters inferred by user. Evaluation of functional markers can still be performed with manual thresholding of expression levels, however this process is easier and more robust using FR maps compared to thresholding the raw signal data.

Although advantageous in many ways, supervised pixel level classification is computationally expensive. For instance, generating FR maps for a 41-page image of $800\mu m$ by $800\mu m$ of a highly-expressed marker (e.g. tumor cell markers) in the TNBC data took 140 minutes using an Intel Core i9 CPU-3.30GHz with 128 GB RAM. This time was measured to be less than 90 minutes for low-count markers (e.g. NK cell marker). The time required for labeling the training data should also be considered as 10-15 minutes for adding 20-100 annotations for positive and negative classes depending on the variability in different channels of a marker in a data set. For instance for labeling a marker like Foxp3, 10 annotations per class is sufficient for the classifier to remove the noise from the data. For more ubiquitous signals such as Beta-tubulin or Vimentin more examples are needed. While the run time of our proposed framework is relatively high compared to conventional image preprocessing pipelines, the significantly enhanced SNR reduces manual efforts required downstream for quality control.

Methods

Ovarian tumor sample preparation

Images from 12 previously described ovarian tumors were analyzed [39, 43]. Tumors were formalin-fixed, paraffin embedded (FFPE) and assembled in a tissue microarray (TMA) consisting of 2 mm cores. TMA construction was approved by Colorado Multiple Institutional Review Board (#17-7788).

Computational framework

Data structure: the input to the classifier is a multipage TIFF where each page depicts the signal for a

single conjugated antibody. The output of the classifier is a multipage TIFF where each page depicts the corresponding FR map. The signal counts are then measured using the FR maps and are saved in a tabular format.

Training a classifier: pixel classification is extensively researched with versatile, interactive, and user-friendly tools available for microscopic imaging data [34–36, 44]. Therefore, we did not develop any tool for pixel classification and instead used TWS [34], a series of library methods that combines Fiji [44] and WEKA [45] and provides an interactive tool for extracting non-linear features and statistical properties of imaging data from user-provided examples. To use the TWS, two sets of inputs are needed from the user: a) a training set that includes examples from class I (signal) and class II (noise and artifacts), and b) the nonlinear features that should be extracted. For the training set, we provided annotations on the image using the TWS interface. For feature selection, TWS provides library functions to evaluate the effectiveness of the extracted features in classification as well as the accuracy of algorithms. Given the differences between statistical properties of noise and artifacts, and the marker signal, empirically we observed that using a subset of nonlinear features available in TWS is sufficient to generate FR maps that are free from noise and artifacts. List of extracted features and their parameters are given in Table S1. Maintaining the out-of-bag error below 1.3%, we trained a Random Forest algorithm with 1000 trees and visually inspected the resulting FR maps and corrected the pixel labels until satisfactory results were obtained. Once an appropriate model is trained for a randomly selected images of a marker, it can be applied to all images depicting that marker in a data set.

Integrating the Single-cell segmentation and FR maps counts: we have developed a series of code that generates the required data structure for the classifier and integrates the output of the classifier with the single-cell segmentation results. Using the the FR maps and the cell segmentation maps, we generate a cell table with cells in row and summary statistics of expression as well as spatial information and morphological properties of cells in columns.

Data Analysis

Ovarian tumor data collected by Vectra: using inForm software (Akoya Biosciences), the images were spectrally unmixed, individual cells were identified using DAPI+ nuclei, and the phenotyping algorithms were trained by marking over 100 cells as positive or negative for each of the phenotypic markers in the panel (CD3, CD8, CD68, CD19, and CK). The algorithms were applied to the entire dataset and the data were merged and consolidated in Phenoptr Reports, an open access software package by Akoya Biosciences. Using our proposed pipeline, we generated five stacks from the phenotypic markers where each stack contained 12 images. We then used the TWS tool [34] to train a pixel classifier for each stack and produced a set of FR maps for each marker. We selected Gaussian blur, Sobel filter, mean, median, and entropy kernels for noise reduction and texture filtering. Finally, we extracted single-cell expression values by measuring the signal counts on the FR maps of the positive class. To avoid any discrepancies caused by differences in the single-cell segmentation, we used the same cell segmentation maps produced by inForm. We used the extracted single-cell information as input to FlowSOM [40] and identified tumor cells and four groups of immune cells. The unidentified cluster includes cells with no positive signal for the present markers or cells with co-expression of mutually exclusive markers.

Breast cancer tissue collected by MIBI: We trained a Random Forest classifier using the following features: Gaussian blur, for denoising, mean, median, and entropy kernels for texture filtering, and Sobel filter and difference of Gaussians as edge detectors. We used the single-cell segmentation maps that were generated in TNBC study [2] to extract the single-cell information from our FR maps. Then we followed the exact hierarchical clustering scheme that was originally used to identify the cell types. In doing so, initially FlowSOM was used to cluster the cells into "Immune" and "Non-immune" using 16 markers (CD45, FoxP3, CD4, CD8, CD3, CD20, CD16, CD68, MPO, HLA-DR, Pan-Keratin, Keratin17, Keratin6, p53, Beta catenin, EGFR). Then, using 8 markers (Vimentin, SMA, CD31, Beta-catenin, EGFR, Keratin 17, Keratin 6, Pan-keratin) non-immune cells were clustered to into Epithelial, Mesenchymal, Endothelial and Unidentified. Immune cells were clustered into 12 groups using 13 markers (CD4, CD16, CD56, CD209, CD11c, CD68, CD8, CD3, CD20, HLA-DR, CD11b, MPO and FoxP3). We then manually labeled merged and labeled the clusters with different cell types.

Data and software availability

Two sets of data are described in this work, ovarian cancer data that was collected in-house, and a publicly available breast cancer data [2]. All the information including channel images, segmentation masks and cell identities for the ovarian cancer data can be accessed by request. The code for the analysis can be downloaded at `https://github.com/himsr-lab/CU-PhenoNorm`.

Acknowledgments

Competing interests

Authors declare no conflict of interest.

References

[1] Michael Angelo, Sean C Bendall, Rachel Finck, Matthew B Hale, Chuck Hitzman, Alexander D Borowsky, Richard M Levenson, John B Lowe, Scot D Liu, Shuchun Zhao, et al. Multiplexed ion beam imaging of human breast tumors. *Nature medicine*, 20(4):436–442, 2014.

[2] Leeat Keren, Marc Bosse, Diana Marquez, Roshan Angoshtari, Samir Jain, Sushama Varma, Soo-Ryum Yang, Allison Kurian, David Van Valen, Robert West, et al. A structured tumor-immune microenvironment in triple negative breast cancer revealed by multiplexed ion beam imaging. *Cell*, 174(6):1373–1387, 2018.

[3] Charlotte Giesen, Hao AO Wang, Denis Schapiro, Nevena Zivanovic, Andrea Jacobs, Bodo Hattendorf, Peter J Schüffler, Daniel Grolimund, Joachim M Buhmann, Simone Brandt, et al. Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nature methods*, 11(4):417–422, 2014.

[4] Yury Goltsev, Nikolay Samusik, Julia Kennedy-Darling, Salil Bhate, Matthew Hale, Gustavo Vazquez, Sarah Black, and Garry P Nolan. Deep profiling of mouse splenic architecture with codex multiplexed imaging. *Cell*, 174(4):968–981, 2018.

[5] Christian M Schürch, Salil S Bhate, Graham L Barlow, Darci J Phillips, Luca Noti, Inti Zlobec, Pauline Chu, Sarah Black, Janos Demeter, David R McIlwain, et al. Coordinated cellular neighborhoods orchestrate antitumoral immunity at the colorectal cancer invasive front. *Cell*, 182(5):1341–1359, 2020.

[6] Edward C Stack, Chichung Wang, Kristin A Roman, and Clifford C Hoyt. Multiplexed immunohistochemistry, imaging, and quantitation: a review, with an assessment of tyramide signal amplification, multispectral imaging and multiplex analysis. *Methods*, 70(1):46–58, 2014.

[7] Jia-Ren Lin, Mohammad Fallahi-Sichani, and Peter K Sorger. Highly multiplexed imaging of single cells using a high-throughput cyclic immunofluorescence method. *Nature communications*, 6(1):1–7, 2015.

[8] Jia-Ren Lin, Mohammad Fallahi-Sichani, and Peter K Sorger. Highly multiplexed imaging of single cells using a high-throughput cyclic immunofluorescence method. *Nature communications*, 6(1):1–7, 2015.

[9] Jia-Ren Lin, Benjamin Izar, Shu Wang, Clarence Yapp, Shaolin Mei, Parin M Shah, Sandro Santagata, and Peter K Sorger. Highly multiplexed immunofluorescence imaging of human tissues and tumors using t-cycif and conventional optical microscopes. *Elife*, 7, 2018.

[10] Mikhail Binnewies, Edward W Roberts, Kelly Kersten, Vincent Chan, Douglas F Fearon, Miriam Merad, Lisa M Coussens, Dmitry I Gabrilovich, Suzanne Ostrand-Rosenberg, Catherine C Hedrick, et al. Understanding the tumor immune microenvironment (time) for effective therapy. *Nature medicine*, 24(5):541–550, 2018.

[11] Takahiro Tsujikawa, Sushil Kumar, Rohan N Borkar, Vahid Azimi, Guillaume Thibault, Young Hwan Chang, Ariel Balter, Rie Kawashima, Gina Choe, David Sauer, et al. Quantitative multiplex immunohistochemistry reveals myeloid-inflamed tumor-immune complexity associated with poor prognosis. *Cell reports*, 19(1):203–217, 2017.

[12] Erik Gerdtsson, Milind Pore, Jana-Aletta Thiele, Anna Sandström Gerdtsson, Paymaneh D Malihi, Rafael Nevarez, Anand Kolatkar, Carmen Ruiz Velasco, Sophia Wix, Mohan Singh, et al. Multiplex protein detection on circulating tumor cells from liquid biopsies using imaging mass cytometry. *Convergent science physical oncology*, 4(1):015002, 2018.

[13] Xavier Rovira-Clave, Sizun Jiang, Yunhao Bai, Graham Barlow, Salil Bhate, Ahmet F Coskun, Guojun Han, Bokai Zhu, Chin-Min Kimmy Ho, Chuck Hitzman, et al. Subcellular localization of drug distribution by super-resolution ion beam imaging. *bioRxiv*, page 557603, 2019.

[14] Qing Chang, Olga I Ornatsky, Iram Siddiqui, Rita Straus, Vladimir I Baranov, and David W Hedley. Biodistribution of cisplatin revealed by imaging mass cytometry identifies extensive collagen binding in tumor and normal tissues. *Scientific reports*, 6(1):1–11, 2016.

[15] Yuan Zhao, Mohamed Uduman, Jacqueline HY Siu, Thomas J Tull, Jeremy D Sanderson, Yu-Chang Bryan Wu, Julian Q Zhou, Nedyalko Petrov, Richard Ellis, Katrina Todd, et al. Spatiotemporal segregation of human marginal zone and memory b cell populations in lymphoid tissue. *Nature communications*, 9(1):1–15, 2018.

[16] Na Li, Vincent van Unen, Tamim Abdelaal, Nannan Guo, Sofya A Kasatskaya, Kristin Ladell, James E McLaren, Evgeny S Egorov, Mark Izraelson, Susana M Chuva de Sousa Lopes, et al. Memory cd4+ t cells are generated in the human fetal intestine. *Nature immunology*, 20(3):301–312, 2019.

[17] Berend Snijder and Lucas Pelkmans. Origins of regulated cell-to-cell variability. *Nature reviews Molecular cell biology*, 12(2):119–125, 2011.

[18] Bhavna Rani, Yuan Cao, Andrea Malfettone, Ciprian Tomuleasa, Isabel Fabregat, and Gianluigi Giannelli. Role of the tissue microenvironment as a therapeutic target in hepatocellular carcinoma. *World Journal of Gastroenterology: WJG*, 20(15):4128, 2014.

[19] Stéphane Chevrier, Helena L Crowell, Vito RT Zanotelli, Stefanie Engler, Mark D Robinson, and Bernd Bodenmiller. Compensation of signal spillover in suspension and imaging mass cytometry. *Cell systems*, 6(5):612–620, 2018.

[20] Nicolas Damond, Stefanie Engler, Vito RT Zanotelli, Denis Schapiro, Clive H Wasserfall, Irina Kusmartseva, Harry S Nick, Fabrizio Thorel, Pedro L Herrera, Mark A Atkinson, et al. A map of human type 1 diabetes progression by imaging mass cytometry. *Cell metabolism*, 29(3):755–768, 2019.

[21] Yue J Wang, Daniel Traum, Jonathan Schug, Long Gao, Chengyang Liu, Mark A Atkinson, Alvin C Powers, Michael D Feldman, Ali Naji, Kyong-Mi Chang, et al. Multiplexed in situ imaging mass cytometry analysis of the human endocrine pancreas and immune system in type 1 diabetes. *Cell metabolism*, 29(3):769–783, 2019.

[22] Coleman R Harris, Eliot T McKinley, Joseph T Roland, Qi Liu, Martha J Shrubsole, Ken S Lau, Robert J Coffey, Julia Wrobel, and Simon N Vandekar. Quantifying and correcting slide-to-slide variation in multiplexed immunofluorescence images. *Bioinformatics*, 38(6):1700–1707, 2022.

[23] Shan E Ahmed Raza, Daniel Langenkämper, Korsuk Sirinukunwattana, David Epstein, Tim W Nattkemper, and Nasir M Rajpoot. Robust normalization protocols for multiplexed fluorescence bioimage analysis. *BioData mining*, 9(1):1–13, 2016.

[24] John W Hickey, Yuqi Tan, Garry P Nolan, and Yury Goltsev. Strategies for accurate cell type identification in codex multiplexed imaging data. *Frontiers in Immunology*, page 3317, 2021.

[25] Alex Baranski, Idan Milo, Shirley Greenbaum, John-Paul Oliveria, Dunja Mrdjen, Michael Angelo, and Leeat Keren. Maui (mbi analysis user interface)—an image processing pipeline for multiplexed mass based imaging. *PLoS computational biology*, 17(4):e1008887, 2021.

[26] Stéphane Chevrier, Helena L Crowell, Vito RT Zanotelli, Stefanie Engler, Mark D Robinson, and Bernd Bodenmiller. Compensation of signal spillover in suspension and imaging mass cytometry. *Cell Systems*, 6(5):612–620, 2018.

[27] Yue J Wang, Daniel Traum, Jonathan Schug, Long Gao, Chengyang Liu, Mark A Atkinson, Alvin C Powers, Michael D Feldman, Ali Naji, Kyong-Mi Chang, et al. Multiplexed in situ imaging mass cytometry analysis of the human endocrine pancreas and immune system in type 1 diabetes. *Cell metabolism*, 29(3):769–783, 2019.

[28] Nicolas Damond, Stefanie Engler, Vito RT Zanotelli, Denis Schapiro, Clive H Wasserfall, Irina Kusmartseva, Harry S Nick, Fabrizio Thorel, Pedro L Herrera, Mark A Atkinson, et al. A map of human type 1 diabetes progression by imaging mass cytometry. *Cell metabolism*, 29(3):755–768, 2019.

[29] Ximo Pechuan Jorge, Xiao Li, Tyler Risom, Artem Zubkov, Evgeniy Tabatsky, Aleksandr Prilipko, Xin Ye, Zhen Shi, Frank Peale, Derrek Hibar, et al. Spex: A modular end-to-end analytics tool for spatially resolved omics of tissues. *bioRxiv*, 2022.

[30] Nicholas Hamilton. Quantification and its applications in fluorescent microscopy imaging. *Traffic*, 10(8):951–961, 2009.

[31] Jiyun Byun, Mark R Verardo, Baris Sumengen, Geoffrey P Lewis, BS Manjunath, and Steven K Fisher. Automated tool for the detection of cell nuclei in digital microscopic images: application to retinal images. *Mol Vis*, 12(105-07):949–60, 2006.

[32] Hui Kong, Hatice Cinar Akakin, and Sanjay E Sarma. A generalized laplacian of gaussian filter for blob detection and its applications. *IEEE transactions on cybernetics*, 43(6):1719–1733, 2013.

[33] Mehmet Sezgin and Bülent Sankur. Survey over image thresholding techniques and quantitative performance evaluation. *Journal of Electronic imaging*, 13(1):146–165, 2004.

[34] Ignacio Arganda-Carreras, Verena Kaynig, Curtis Rueden, Kevin W Eliceiri, Johannes Schindelin, Albert Cardona, and H Sebastian Seung. Trainable weka segmentation: a machine learning tool for microscopy pixel classification. *Bioinformatics*, 33(15):2424–2426, 2017.

[35] Christoph Sommer, Christoph Straehle, Ullrich Koethe, and Fred A Hamprecht. Ilastik: Interactive learning and segmentation toolkit. In *2011 IEEE international symposium on biomedical imaging: From nano to macro*, pages 230–233. IEEE, 2011.

[36] Stuart Berg, Dominik Kutra, Thorben Kroeger, Christoph N Straehle, Bernhard X Kausler, Carsten Haubold, Martin Schiegg, Janez Ales, Thorsten Beier, Markus Rudy, et al. Ilastik: interactive machine learning for (bio) image analysis. *Nature Methods*, 16(12):1226–1232, 2019.

[37] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[38] Benjamin Steinhart, Kimberly R Jordan, Jaidev Bapat, Miriam D Post, Lindsay W Brubaker, Benjamin G Bitler, and Julia Wrobel. The spatial context of tumor-infiltrating immune cells associates with improved ovarian cancer survivalspatial interactions in the tumor immune microenvironment. *Molecular Cancer Research*, 19(12):1973–1979, 2021.

[39] Kimberly R Jordan, Matthew J Sikora, Jill E Slansky, Angela Minic, Jennifer K Richer, Marisa R Moroney, Junxiao Hu, Rebecca J Wolsky, Zachary L Watson, Tomomi M Yamamoto, et al. The capacity of the ovarian cancer tumor microenvironment to integrate inflammation signaling conveys a shorter disease-free intervaltherapy-induced remodeling of the tumor microenvironment. *Clinical Cancer Research*, 26(23):6362–6373, 2020.

[40] Sofie Van Gassen, Britt Callebaut, Mary J Van Helden, Bart N Lambrecht, Piet Demeester, Tom Dhaene, and Yvan Saeys. Flowsom: Using self-organizing maps for visualization and interpretation of cytometry data. *Cytometry Part A*, 87(7):636–645, 2015.

[41] Noah F Greenwald, Geneva Miller, Erick Moen, Alex Kong, Adam Kagel, Christine Camacho Fullaway, Brianna J McIntosh, Ke Leow, Morgan Sarah Schwartz, Thomas Dougherty, et al. Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. *bioRxiv*, 2021.

[42] Heeva Baharlou, Nicolas P Canete, Anthony L Cunningham, Andrew N Harman, and Ellis Patrick. Mass cytometry imaging for the study of human diseases—applications and data analysis strategies. *Frontiers in immunology*, 10:2657, 2019.

[43] Zachary L Watson, Tomomi M Yamamoto, Alexandra McMellen, Hyunmin Kim, Connor J Hughes, Lindsay J Wheeler, Miriam D Post, Kian Behbakht, and Benjamin G Bitler. Histone methyltransferases ehmt1 and ehmt2 (glp/g9a) maintain parp inhibitor resistance in high-grade serous ovarian carcinoma. *Clinical epigenetics*, 11(1):1–16, 2019.

[44] Caroline A Schneider, Wayne S Rasband, and Kevin W Eliceiri. Nih image to imagej: 25 years of image analysis. *Nature methods*, 9(7):671–675, 2012.

[45] Mark Hall, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H Witten. The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18, 2009.

Figure S1: Marker expression per cell measured using the images (top row) and the FR maps (middle row) overlaid on the tSNE plot for rest of the immune and tumor markers. The bottom row demonstrate correlation between marker expression per cell from raw images (x-axis) and FR maps (y-axis). **b,** Cell types clustered by marker expression. Expression values for each marker are measured from the FR maps. **c,** Cell-cell comparison between the cell type predicted by inForm (x-axis) and our framework (y-axis).

Table S1: Extracted features for classification of TNBC and ovarian cancer data sets.

Feature	Number of extracted features	Notes [34]
Gaussian blur	5 (σ = 1, 2, 4, 8, 16)	Performs 5 convolutions of Gaussian kernels with the 5 variations of σ and generates 5 features for each image.
Sobel filter	5	Calculates an approximate of the gradient of the image intensity at each pixel for each image after applying 5 variations of Gaussian blurs; generates 5 features for each image.
Difference of Gaussians	10	Calculates two Gaussian blur images from the original image and subtracts one from the other. Thus, with 5 variations of Gaussian blur, 10 features are generated for each image.
Mean	5	Mean is calculated from the pixels within a radius of σ pixels from the target pixel and the target pixel is set to that mean value. With 5 variations of Gaussian blur, 5 features are generated for each image.
Median	5	Median is calculated from the pixels within a radius of σ pixels from the target pixel and the target pixel is set to that median value. With 5 variations of Gaussian blur, 5 features are generated for each image.
Entropy	20	Within radius σ around each pixel, generates the histogram of that circle using n = 32, 64, 128, 256 as the number of bins and calculates the entropy as $\sum -p*\log2(p)$, where p is the probability of each collection in the histogram. With 5 variations of Gaussian blurs adds 20 features for each image.