

Statistical Analysis of Multiplex Immunofluorescence and Immunohistochemistry Imaging Data

Julia Wrobel^{1*}, Coleman Harris², Simon Vandekar²

¹Department of Biostatistics and Informatics, University of Colorado Anschutz Medical Campus, Aurora, CO 80045, USA

²Department of Biostatistics, Vanderbilt University Medical Center, Nashville, TN 37203, USA

*Corresponding Author

Julia Wrobel

13001 E 17th St

Aurora, CO USA 80045

julia.wrobel@cuanschutz.org

Abstract

Advances in multiplexed single-cell immunofluorescence (mIF) and multiplex immunohistochemistry (mIHC) imaging technology have enabled analysis of cell-to-cell spatial relationships that promise to revolutionize our understanding of tissue-based diseases and autoimmune disorders. Multiplex images are collected as multichannel TIFF files, then must be de-noised, segmented to identify cells and nuclei, normalized across slides and protein markers to correct for batch effects, phenotype, then tissue composition and spatial context at the cellular level is analyzed. This chapter discusses methods and software infrastructure for image processing and statistical analysis of mIF/mIHC data.

Key Words: multiplex imaging, bioimage analysis, spatial statistics, cell phenotyping, image normalization

1. Introduction

The development of multiplexed immunofluorescence (mIF) and multiplex immunohistochemistry (mIHC) imaging has revolutionized researchers' ability to study tissue structure and function at a cellular level while preserving the original spatial context of the sample. Functionally, mIF/mIHC is a hybrid between multiplex single-cell techniques like flow and mass cytometry [1, 2], which allow for staining of multiple antibody markers but require the dissolution of tissue in a suspended solution, and traditional immunohistochemistry, which preserves the original tissue structure but staining is limited to 2-3 antibodies [3]. Along with similar single cell imaging technologies, including multiplexed ion beam imaging (MIBI) [4, 5], and imaging mass cytometry (IMC) [6, 7], mIF platforms have particular promise in the field of cancer immunology, where they have been used to characterize the tumor immune microenvironment (TIME). This chapter surveys statistical methods for preprocessing and analysis of mIF data using examples motivated by scientific questions about the TIME. Specifically, the study of abundance, spatial location, and protein expression of immune cells in and around tumors and how immune profiling relates to improved prognosis for cancer patients. However, the methods highlighted can be directly applied or adapted for mIF data from many other research areas; recent mIF advances in non-cancer domains include the study of neurodegenerative diseases [8, 9], type 1 diabetes [10], inflammatory bowel disease [11], and multiple sclerosis [12].

The collection of mIF data follows a complex data acquisition, image processing, and analysis pipeline depicted at a high level in Figure 1. First, tissue is placed on a slide. An advantage of mIF

imaging is that it can be applied to regions of interest (ROIs) of a whole tissue slide, or to tissue microarrays (TMAs), allowing for analysis across large tissue areas or a large number of samples [13]. The slide is then stained with multiple, or *multiplex*, primary antibodies, referred to as *markers*, that attach to biomarkers of interest. Markers can be broadly categorized as either *phenotypic markers*, which are used to categorize cell type, or *functional markers* that inform cell function and can be present across multiple cell types. The marker antibodies are attached to fluorophores, which emit light at specific wavelengths when imaged, and the signal of light intensity at a particular wavelength is used to quantify biomarker abundance. Then intensity spectra must be unmixed to give separate intensity values for each marker. Resulting images are stored as multichannel tiff files where each channel provides the tissue intensities for a particular marker.

These images are then segmented to identify tissue compartments (e.g. tumor vs. stroma), cells, and nuclei. Single-cell segmentation is an important and rapidly advancing area of research that benefits from recent success in neural networks and deep learning. Many algorithms have been developed as part of proprietary software packages such as Inform and HALO, but are also being developed by academic research groups [14–16]. These algorithms are often tissue specific, can be influenced by the image production process, so sometimes must be retrained for new data. Research in this area works to make the algorithms more general and reduce bias. Because research in single-cell segmentation is broad in itself, the topic is not discussed further in this chapter.

After segmentation, cells are *phenotyped*, or given biological label(s) based on their mean marker intensities. Segmentation and phenotyping converts images into a tabular dataset where each row is a cell and columns are features identifying cell spatial location, tissue type, phenotype, marker intensities, and other cell or image characteristics. Finally, the datasets are analyzed, potentially in combination with patient level outcomes such as survival time, disease subtype, or cancer stage. While examples in this chapter focus cell-level data, the images can also be analyzed directly using the same methodology, using pixels rather than cells as the unit of observation. In our more detailed discussion below, we focus on

steps of the mIF/mIHC pipeline downstream of image acquisition that can be applied and improved upon by statisticians and bioinformaticians.

2. Ovarian and lung cancer datasets

With 184,000 deaths per year, globally, ovarian cancer is the most lethal reproductive cancer [17]. In addition, 75% of patients with high-grade serous ovarian cancer have tumor recurrence and develop therapy resistant disease [18]. Lung cancer is even more deadly, making up 25% of all cancer deaths [19]. There is mounting evidence of the prognostic significance of the tumor-immune composition in lung and ovarian cancers [20, 21]. However, the biological significance of spatial relationships between cell types in the TIME of these cancers is not yet well understood, and mIF imaging may provide key scientific insights. To address gaps in understanding of the biology of these cancers, a high-grade serous ovarian cancer (HGSOC) and a non-small cell lung carcinoma (NSCLC) mIHC dataset were collected at the University of Colorado Anschutz Medical Campus, and will be used to illustrate the methods in this chapter.

Images were collected on Vectra 3.0TM and Vectra PolarisTM instruments for the ovarian and lung datasets, respectively, and image processing was performed using InForm software [22–25]. The Vectra 3.0TM and Vectra PolarisTM by Akoya Biosciences are currently the most widely used mIF/mIHC platforms, and InForm is accompanying proprietary software that provides a graphical user interface (GUI) for image processing steps including spectral unmixing, segmentation, and phenotyping. For our lung and ovarian datasets, InForm was used for spectral unmixing, nucleus, cell, and tissue segmentation, and cell phenotyping; however, it is also common to use a mixture of InForm and other GUI-based software such as HALO and QuPATH to convert images into tabular datasets [26]. Both datasets are freely available in R in a tabular post-segmentation format as part of the Bioconductor ExperimentHub package VectraPolarisData [27]. Specifically, these data are stored as SpatialExperiment class objects[28] and code used to reproduce figures in this chapter using these data is available at <https://github.com/julia->

[wrobel/mIF_chapter_code](#). Each dataset provides cell spatial locations and marker channel intensities for over 1.5 million cells in each dataset, as well as patient-level clinical information.

The HGSOC ovarian dataset comes from a spatial analysis of a tumor microarray of 128 subjects with ovarian cancer collected at the University of Colorado [29] and contains one image per subject. Images were stained for DAPI, five phenotypic markers (CD3, CD8, CD68, CD19, cytokeratin), and two functional markers (phospho-stat3 and IER3), allowing for identification of CD4+, CD8+, CD68+, and CD19+ immune cells. Patient-level variables include age, cancer stage at diagnosis, BRCA mutation status, debulking status, survival time, survival status, and time-to-recurrence.

The NSCLC lung data was collected to analyze the relationship between tumor-infiltrating immune cells and major histocompatibility II expressing cancer cells in the NSCLC tumor immune microenvironment, and consists of 761 mIF-imaged regions of interest (ROIs) from 153 patients [30]. Images were stained for DAPI, five phenotypic markers (CD3, CD8, CD14, CD19, cytokeratin) and one functional marker (MHCII), allowing for identification of CD4+, CD8+, CD14+, and CD19+ immune cells. Patient-level variables include age, gender, cancer stage at diagnosis, pack-years of cigarette smoking, survival time, survival status, and time-to-recurrence. Because each sample consist of multiple ROIs that are not spatially adjacent, these images cannot be stitched together and treated as one image. Instead, repeated measures approaches must be considered when analyzing image data in models for patient level outcomes. Figure 2 (far left) shows an example image from this dataset which combines intensities from all 8 channels. The three right columns are the DAPI, cytokeratin, and CD8 channels for this image, respectively; these are used to identify cell nuclei, tumor areas, and CD8 T cells.

3. Methods

Multiplex images are typically provided as multichannel tiff files and transformed into tabular sets after a cell segmentation step. In the following sections we direct our focus towards analyzing the cell-level tabular data after single-cell segmentation. These methods applied at the cell level can also be applied to pixels of a non-tabular mIF image directly, but at a greater computational cost. We divide our review of

methods in the following subsections that address different areas of the processing pipeline depicted in Figure 1. In Section 3.1 we cover marker channel processing steps that must be addressed before data is appropriate for answering questions pertaining to scientific hypotheses. In Section 3.2 we cover methods for cell phenotyping, the process of classifying cell types. In Section 3.3 we discuss analysis of image summary data using statistics that capture the cellular composition of tissue samples. An increasing focus of the literature on mIF data is dedicated to spatial analysis, and in Section 3.4 we detail existing methods; in particular we emphasize approaches grounded in point process theory because this is an area of mIF where we believe expertise from statisticians and bioinformaticians can have greatest impact. In Section 3.5 we conclude with a review of available open-source software tools for the visualization, preprocessing, and analysis of multiplex imaging data.

3.1: Image Transformation, Normalization, and Batch Correction

Image transformation, normalization, and batch correction are, in general, used to make the data more appropriate for downstream analysis by adjusting intensity distributions or removing systematic biases due to processing variables. Image transformations (e.g. log, arcsinh, or root) are used to make the data more amenable to modeling, but do not explicitly adjust for systematic effects due to image acquisition variables. Image normalization is the process of making the same tissue type appear similar in images that were acquired across different sites, times, imaging parameters, or platforms [31–35]. Normalization adjusts the distribution of marker intensities in each slide, batch, channel, or image separately to make the distributions appear more similar. Batch correction (often called harmonization in the imaging literature) is the process of removing systematic effects in groups of cells or images that are due to processing steps or other non-biological sources [36]. Batch correction methods use a single model across slides to account for covariates while estimating and removing batch effects across slides. Transformation, normalization, and batch correction methods are applied to each channel of the mIF data, so it is important to have procedures that are effective for a wide range of different marker types, including both phenotypic markers and functional markers. The top two rows of Figure 3 show evidence of potential batch effects in

the NSCLC data. In this figure, distributions of unnormalized marker intensities are shown for three subjects and two immune cell markers (CD14 and CD19). Each subject has densities plotted for multiple ROIs (dotted lines) as well as the subject-specific mean density (solid lines). For this dataset each slide contains data for a single subject, and the range of intensity values differ substantially for each slide.

Normalization and batch correction algorithms can be applied at the image or cell level [37]. The goals of the two procedures are methodologically similar and batch correction methods from genomics have been very effectively applied in imaging datasets [34, 38, 39]. There is an emerging literature on batch correction in mIF data [31, 32, 34, 37, 40, 41], and this research area is likely to grow as the effects of imaging parameters and platforms are better understood. It is well-accepted that there are effects of slide, staining round, and batch on the distribution of mIF intensities at the image and cell level. mIF data are unique from magnetic resonance imaging (where batch correction is being widely applied) because the tissue populations differ significantly in their composition across samples, so close alignment of the image histograms is not always an indication of effective normalization (Figure 2 top rows). Batch adjustment methods developed for other genomic assays, such as sc-RNAseq are promising, but cannot be directly applied due to greater overlap in the distributions of expressed and unexpressed cells. For these reasons, mIF normalization and batch correction is a rapidly growing area of research. Here, we provide a brief overview of recently proposed methods and software for normalization and batch correction in mIF data.

Image Normalization

Raza et al. were one of the earliest groups to explore normalization methods for mIF data applied at the image level [31]. They compared a linear normalization,

$$\tilde{I}_{ic}(p) = \frac{I_{ic}(p) - \min_p I_{ic}(p)}{\max_p I_{ic}(p) - \min_p I_{ic}(p)},$$

where $I_{ic}(p)$ denotes the image value for pixel p of slide i and channel c , and $\tilde{I}_{ic}(p)$ denotes the normalized data; and a tanh transformation after applying mean division normalization,

$$\tilde{I}_{ic}(p) = \tanh \left\{ \frac{I_{ic}(p)}{2 \times \bar{I}_{ic}} \right\},$$

where \bar{I}_{ic} is the slide mean. The methods were evaluated based on expert ranking in consistency between tissue compartments across images and KL-divergence within and between tissue classes. Linear normalization almost always scored better across their evaluation procedures.

Graf et al. proposed several quantile based normalization methods and applied the methods at the pixel and segmented cell levels of analysis. They compared linear

$$\tilde{I}_{ic}(p) = I_{ic}(p) + \text{Quantile}(I_i(p)) - \text{Quantile}(I_{ic}(p))$$

and multiplicative normalizations

$$\tilde{I}_{ic}(p) = \frac{I_{ic}(p) \times \text{Quantile}(I_c(p))}{\text{Quantile}(I_{ic}(p))},$$

where $\text{Quantile}(I_c(p))$ is the quantile across all images and $\text{Quantile}(I_{ic}(p))$ is the quantile in slide i only. They considered the median for the linear transformations and the median and 75 quantile for the multiplicative normalization as well as several other quantile based normalizations (see Table 1 of [37]).

They used a novel approach to evaluate normalization by leveraging the DAPI channel, which was repeatedly stained through multiple rounds and serves as a “virtual” ground truth because it marks the nucleus and should be identical across staining rounds. Acquisition parameters were varied through the rounds, but the same tissues were used, so that they could assume (in the absence of batch effects) the tissue should have consistent image intensity through the rounds of staining. Their normalization procedure and methods assumed control regions that are expected to be consistent across the staining rounds and they used coefficient of variation to quantify batch effects. The authors found that the multiplicative normalization applied on log transformed data had the best performance (lowest coefficient of variation).

Other normalization methods have relied on control regions for estimation [32] or evaluation [40] of batch effects. These assume adjacent control sections are used across slides and that alignment of these control regions implies alignment of the tissue of interest. These are not possible if a good control region is used and may not be effective if sections of the control tissue are not adjacent sections. In general, prior

to transforming the data normalization by dividing by the mean, quantile/median, or range of the data seems to work, but more evaluation across a greater range of data sets is merited.

Batch correction

Harris et al. compared several normalization batch correction methods to adjust for slide effects in mIF data at the cell-level [34]. They compared an adaptation of the widely used ComBat method [36, 38], and a functional data registration method to perform histogram matching across slides [34, 42] as well as several transformations and a mean division normalization. To assess the methods in the absence of a ground truth, they established a framework for evaluating batch effects for mIF data, that includes a) quantification of misalignment of the expression histograms for each marker channel b) quantification of discordance between slide-level phenotyping (no batch effects present) to across slide phenotyping (batch effects may be present) before and after normalization c) Quantification of proportion of variance due to slide d) visualization of batch and biological effects in UMAP embedding. Across these metrics, applying mean division, mean division with a log transformation, and using functional data registration all performed well. The authors recommended mean division and mean division with log transformation as a balance between efficacy and simplicity.

Future research and software development

Based on the current research, normalization is demonstrated to be reliable and effective, however the simplistic procedures currently used are unlikely to remain as the top performers as we learn more about sources of systematic variation in mIF data and more public data sets are released that can be used for evaluation [43, 44]. Future directions may more rigorously consider batch adjustment methods. For example, ComBat has been demonstrated to fail in mIF data due to zero inflation [34], but a recently developed method using a negative binomial model may hold more promise [45]. In addition, the normalization methods used currently are robust across channels, but do not explicitly leverage information about the joint distribution across channels, the staining round, or the spatial information available in the image. These features may be used to better understand and quantify systematic effects in mIF.

Software designed for other assays (e.g. Seurat, Giotto) can be used for mIF normalization or batch correction. In addition, most of the recently proposed normalization procedures provide code to implement their methods [37, 40, 41]. In particular, *mxnorm* is a recently published R package devoted to normalization of cell-level mIF data [41]. This package provides the normalization and batch correction algorithms used in [34] and also allows users to specify new algorithms which can be compared using the authors' proposed evaluation framework. The third row of Figure 3 shows plots of discordance scores of phenotypic marker channels for 15 subjects in the NSCLC data before and after normalization. Simple log10 and mean division transform-based normalization was applied using the *mxnorm* package. The discordance score is a novel evaluation metric proposed by [34] that quantifies the difference between threshold-based phenotyping at the slide and population level. Smaller discordance scores indicate a reduction in batch effects, and smaller discordance is seen for the normalized as compared to the raw marker intensity values.

3.2: Cell Phenotyping

Cell phenotyping is the process of identifying discrete cell types from marker expression values. Phenotyping can use unsupervised algorithms, but often requires expert intervention to label and validate cell annotations. This step in the analysis pipeline is not exclusive to mIF and has been well studied in other fields, including flow and mass cytometry[46, 47]. Typical preprocessing before phenotyping for other assays includes batch correction, transformation, (e.g. log or arcsinh) Z-scaling, and dimension reduction. These steps have not been rigorously evaluated in mIF data. For example, dimension reduction may not be necessary due to the relatively small number of image marker channels (10s as opposed to 100s), but data whitening may be beneficial. Here, we provide a brief overview of the approaches used and proposed for cell phenotyping mIF data. These include marker gating, and unsupervised or semisupervised clustering methods.

Marker Gating

Marker gating is a widely used procedure of determining a threshold to define marker positive cells that have expression values above the given threshold [32, 43, 48]. This procedure is performed manually by visualizing univariate or bivariate histograms of cell expression values and image overlays to determine a threshold that separates modes in the data. If three or more channels are necessary to determine cell phenotypes, marker gating can be performed iteratively for each channel or pairs of channels [49–51]. Software with features for performing marker gating of mIF data include Cellprofiler Analyst [50, 51], histoCAT [52], CyLinter [53], and Advanced Cell Classifier [54]. Marker gating is useful for phenotyping cells of known type that can be identified based on a subset of known marker channels. Otherwise, unsupervised algorithms are more convenient for identifying known and unknown cell classes.

Unsupervised clustering algorithms

K-means, Louvain, and Leiden algorithms are unsupervised clustering/community detection algorithms that have been widely implemented for single cell assays because they are computationally efficient in large data sets [55–58] and are included in most software programs for single-cell analysis, such as Seurat, Giotto, and Phenograph [15, 57, 59, 60]. After clustering, manual annotations are assigned for each cluster by visualizing the clusters in a dimension reduced projection (e.g. UMAP, tSNE), visualizing the cluster memberships on the slides, and visualizing expression profiles (Figure 4). K-means is a distance-based clustering; in contrast, Louvain and Leiden clustering are popular graph-based clustering methods that group cells based on the similarity of the cells in a local neighborhood, where the neighborhoods are determined by the similarity of the marker expression values.

Figure 4 shows clustering results applied to five ROIs from one subject in the NSCLC dataset. TSNE was first applied to the raw marker values to reduce dimensions for visualization purposes. The cells for this subject were plotted along the first two TSNE directions and colored by phenotypes calculated using Inform software (left), colored by clusters produced by applying Phenograph to unnormalized marker intensities (center) and colored by clusters produced by applying Phenograph to log₁₀ mean division normalized marker intensities (right). Phenograph produced 17 and 14 clusters for

the unnormalized and normalized data respectively, more than are identifiable by the markers collected in the dataset. However, many map on to clusters labeled in Inform.

These clustering algorithms were not specifically developed for single-cell analysis and adaptations have been developed to address characteristic features of single-cell data such as the dimensionality and rare cell types. For example, RaceID uses outlier detection to revise K-means cluster identities because K-means clustering tends to form clusters of similar sizes, so rare cell types are often mixed with discrete cell types [61]. In addition, methodological and computational modifications to the analysis process have been developed to speed computation in large data sets (> 1 million cells) [57, 62–64]. Two examples are Phenograph and FastPG for clustering cells that are scalable to large datasets [57, 62]. Both make modifications to the k-nearest neighbors computation, weighting estimation, and computational processing of the Louvain algorithm to improve processing speed. Many of the clustering methods mentioned here have been evaluated in single-cell RNA sequencing [46, 47], but these algorithms have not been rigorously evaluated in mIF data. Software tools specifically designed for mIF or other spatial assays may include the above mentioned phenotyping algorithms [59, 65], however it is likely that the phenotyping pipelines for other single-cell assays are not directly applicable here, given differences in acquisition and processing [66].

Recently, several phenotyping algorithms have been proposed specifically for microscopy data that can be applied to mIF data at the cell or image-pixel level. Astir is a cell-level scalable semisupervised phenotyping algorithm using a probabilistic model and variational inference with neural network that was specifically designed for mIF data. As input, Astir takes the mIF expression matrix and a user-specified marker file which defines combinations of marker values attributed to each cell type. Astir can be used to broadly classify cells before applying unsupervised clustering on the subpopulations or to identify known cell types automatically. Cell phenotyping has been performed using morphological summary features in addition to median image intensities within each cell to perform unsupervised clustering [14, 67]. These methods still rely on single-cell segmentation as a preprocessing step but use spatial information through cell shape, size, composition, and texture. When single-cell segmentation is

difficult, then phenotyping can be applied directly to the images in order to identify tissue classes or neighborhoods in mIF analysis that are spatially contiguous [67–72]. Some of these methods, which have been developed in other similar fields, such as imaging mass spectrometry, will likely be useful in tissue where cell segmentation is difficult. After defining neighborhoods they can be used to study spatial interactions between these tissue compartments.

To summarize, phenotyping in mIF data is not well evaluated, but it is possible that algorithms effective in other single-cell assays will be useful in mIF data with methodological modifications to the phenotyping analysis pipeline. Quality assessment and visualization tools that allow users to evaluate cell phenotypes *in situ* are critical for evaluation [15, 53]. Given the variability in possible markers, unsupervised or semisupervised methods are likely to be most applicable. Finally, mIF data accommodate image level clustering, which allows analysts to circumvent single-cell segmentation, which could be biased or difficult depending on the tissue type. Cell and image level phenotypes can then be passed to downstream spatial analytic methods.

3.3 Analysis of Cellular Composition and Marker Expression

As a first step of probing the tissue microenvironment it is common to analyze the cellular composition of the overall sample and tissue sub-compartments. One typical approach is estimating the number, percentage, or density of cell subtypes in different tissue regions; for example, summarizing immune cell proportions within tumor and stroma regions [73]. It is also common to characterize cell composition across levels of a covariate, such as responders and non-responders to a chemotherapy. Summary statistics including percentage or proportion of positive cells for each phenotypic form can be easily calculated in image analysis softwares such as Halo and Inform. However, these quantities are not normally distributed, even after normality transformations, so modeling choices need to be carefully considered [13]. In addition, zero inflation is common in mIF studies where rare cell populations contribute to zero cell counts for some subjects or ROIs in a sample. This can occur, for example, in studies of immune cold tumors such as ovarian cancer. Figure 5 (left panel) shows empirical proportion of

total cells that belong to each type of tumor-infiltrating immune cell, including B cells, CD4 T cells, CD8 T cells, and macrophages, for women in the HGSOC ovarian data with and without a BRCA genetic mutation. Note that there is substantial right skew in the distributions of proportions for all immune cell subtypes, and zero inflation is present in the proportions of B-cells because many images do not contain B-cells. In addition, mIF data often exhibit overdispersion when modeled using common statistical methods for counts and proportions. Overdispersion is when the data exhibit greater variance than is dictated by the chosen model and can lead to Type I error and artificially narrow p-values. Failing to account for non-normality, overdispersion, and zero-inflation can lead to erroneous results.

Modeling counts or proportions with overdispersion and zero inflation

Count data can be effectively modeled using a generalized linear modeling (GLM) framework with a data-informed choice of statistical distribution [74]. A GLM also enables straightforward modeling of how summary statistics, such as cell proportions, differ across levels of a covariate (e.g. treatment response). The binomial and Poisson distributions are common distributions for modeling proportions; however for cell count data from mIF these models will likely be overdispersed and underestimate the variance of model parameters, in part because they implicitly treat multiple cells within the same image as independent. One way to account for overdispersion is to use heteroskedasticity-robust standard errors, which have several standard implementations in R including the *sandwich* library [75, 76]. Another approach is to introduce an additional model parameter that explicitly captures overdispersion, and methods that account for overdispersion in counts or proportions in this way include quasi-binomial, quasi-Poisson, negative binomial, and beta binomial regression [74].

When an excess of zero counts is present, usually due to rare cell types, zero-inflated versions of Poisson, beta binomial, or negative binomial count models can be used. Zero-inflated models have two parts, for example, zero-inflated Poisson regression models zeros using a binomial distribution observations greater than zero using a Poisson distribution. These models have become popular for other types of biological data where excess zeros are present, including the microbiome composition [77, 78] and RNA-seq data [79–81]. Because mIF data that is zero-inflated is also likely overdispersed, zero-

inflated beta binomial and zero-inflated negative binomial are recommended over zero-inflated Poisson. For datasets with multiple tissue samples or ROIs per subject, within-subject correlation induced by repeated measures needs to be accounted for, and generalized mixed effects or estimating equation models can provide summaries of cellular composition at experiment-wide levels [82, 83]. Generalized mixed effects models additionally provide summaries at the subject level, and both frequentist and Bayesian implementations of mixed-effect models for overdispersed and zero-inflated count data are available [84–86].

Figure 5 (right panel) shows results from generalized linear models of proportions of B cells (top) and macrophages (bottom), controlling for BRCA mutation status. Specifically, BRCA mutation status β coefficient estimates and 95% confidence intervals are shown for six different models. Black stars above estimates indicate statistical significance at the 5% α level. The “naive” model applies linear regression to empirical proportions that have been square-root transformed, “Poisson robust” and “binomial robust” indicate Poisson and binomial regression with robust standard errors, and ZIP indicates zero-inflated Poisson regression. The naive model, though it accounts for overdispersion and provides inference about model covariates, is otherwise limited because it does not account for the total number of cells in each image and cannot be back transformed to obtain estimates of cell counts or proportions. The binomial and Poisson models show significant differences by BRCA status for both B cells and macrophages, however these results are likely spurious because these models do not account for overdispersion. When robust standard errors are used for the binomial and Poisson models, the proportion of B cells or macrophages are no longer significantly different across BRCA groups. Zero-inflated Poisson regression was only used to model B cell proportions because the HGSOC data did not contain excess zero counts of macrophages. While the ZIP coefficient for BRCA is significant in Figure 5, this is also likely due to overdispersion.

As Figure 5 demonstrates, failing to properly account for overdispersion can lead to artificially narrow confidence intervals and spurious p-values. To avoid this pitfall, we recommend using models with robust standard errors or others, such as beta binomial, which explicitly model overdispersion. When an excess of zeros is present it is instead more appropriate to use a model that accounts for both zero

inflation and overdispersion, such as zero-inflated beta binomial or zero-inflated negative binomial. If one is unsure whether data has enough zeros to constitute “excess”, hypothesis tests can be used to compare zero-inflated models with their standard counterpart [87–89]. Finally, when multiple images/ROIs are available for each subject, mixed effects versions of these models should be used to account for repeated measures.

Analysis of functional markers

Differences in the distribution and expression of functional markers across cell or patient populations is also of interest. Recent examples include whether abundance of PD-L1 positive cells differ between responders and non-responders to an immunotherapy in ovarian cancer [90] and whether high or low expression of major histocompatibility complex II (MHCII) in tumor cells affects overall survival in lung cancer patients [30]. To simplify analysis of functional markers, they are often categorized into groups by thresholding continuously measured values. The choice of threshold can be arbitrary; some typical approaches include quantile-based groupings, biologically driven cutpoints, and optimal thresholds that maximize some test statistic of interest. For quantile based groupings, a median threshold is most common. Biologically relevant thresholds are chosen based on evidence from previous studies and are often classified at the patient level; for example, defining a patient with estrogen receptor (ER) expression on more than 1% of cells as ER+ [91], or a patient with MHCII expression on more than 5% of tumor cells as MHCII high [30]. However, biologically relevant thresholds are not always available and are sometimes inconsistently applied in the literature across studies and disease types. Finally, optimal cutpoints are chosen to maximize a test statistic of interest; for example, choosing high or low immune cell groups to assess the ovarian cancer tumor microenvironment in African American women [92]. This approach can be appropriate for data exploration and hypothesis generation, but the threshold obtained must be evaluated on external datasets to ensure its biological validity and avoid statistical bias [93, 94]. Recently, [95] proposed a threshold-free approach that classifies subjects into different categories by clustering on Jensen-Shannon distances between marker densities, and may mitigate problems that arise when thresholding functional markers. All of these methods will be sensitive to

upstream marker intensity normalization and phenotyping so it is critical to carefully evaluate earlier results in the data processing pipeline.

While thresholding functional markers is common, categorizing observations from a continuous marker may reduce power and has been criticized within statistics [96]. This can be particularly problematic when the marker distribution is not bimodal and cannot be easily grouped into unexpressed and expressed cell populations. A more reproducible approach is to model marker values continuously, possibly accounting for nonlinearities with covariates using splines.

3.4 Spatial Analysis

Previous steps in the mIF data analysis pipeline are necessary for reducing bias due to technical variation, extracting interpretable image features like cell locations and cell types, and summarizing experiment-level characteristics. However, the explosion in popularity of multiplex imaging has mainly been driven by the unprecedented ability to study spatial relationships in samples with a large and growing number of markers. Many spatial analytic approaches have been adapted from ecology and geostatistics, and new spatial analytic methods specifically for multiplex imaging data are rapidly developing [13]. Below we discuss different classes of spatial analytic tools and techniques, with a particular focus on methods grounded in point process theory. While methods for spatial transcriptomics data may be successfully leveraged for mIF data as well, these are reviewed in more detail elsewhere, for example [97].

Basic tools for data exploration and visualization

There are several recent software packages for high-level data exploration and visualization. Packages like Seurat and Giotto [59, 60] are multiuse toolkits that incorporate dozens of visualization methods and spatial metrics. Both of these packages were introduced for single-cell analysis of RNA sequencing data, and have evolved to accommodate spatial transcriptomics and multiplex imaging data. In contrast, the histoCAT toolbox [52] was introduced specifically to explore and visualize multiplex imaging data, and provides interactive visualizations of cell phenotypes and nearest-neighbor analyses. Similarly, cytoMAP [98] is a user-friendly comprehensive toolkit for quantitative image analysis of

multiplexed tissues, and Squidpy [99] is a modular, scalable Python-based platform for spatial single cell analysis. Each of these software platforms provides similar features; specifically, a united framework to visualize and perform basic spatial analysis of multiplexed imaging data within an interactive graphical user interface or programming environment (usually R or Python). A more comprehensive list of software tools for multiplex imaging data is provided in Table 1.

Spatial statistical learning models

A few approaches that have emerged adapt statistical learning procedures to either identify spatial gene expression, which is used in spatial transcriptomics data because of the coarser spatial resolution, or to identify spatial neighborhoods [69, 100]. These approaches cluster pixels or wells while incorporating information about the spatial dependence, e.g. by smoothing or with a Markov random field. The Giotto package contains a method by [101] that integrates RNA sequencing data with mIF data using hidden Markov random fields. The BayesSpace model [102], which also uses Markov random fields, can be applied to spatial transcriptomics data to enhance resolution and cluster the image, and implements a fully Bayesian approach. [103] build gene or marker networks using nonnegative matrix factorization. Another approach uses spatial latent Dirichlet allocation to discover novel tissue subcompartments [104]. Many of these machine learning based techniques for spatial ‘omics are intended for spatial transcriptomics data, which have lower spatial resolution than mIF data, but can have read counts on hundreds or thousands of genes. Because of this format, these spatial clustering methods incorporate a dimension reduction step in the marker dimension, however, when applied to mIF datasets this dimension reduction may not be necessary and these methods may be well adapted for handling the high spatial resolution of mIF data.

Spatial metrics based on point processes

The most relevant approaches for analyzing spatial relationships between cell types or quantifying spatial co-expression of biological markers in multiplex imaging data are based on point process theory. The location of cells in mIF data are treated as following a point process, realizations of a point process are called “point patterns”, and point process models seek to understand correlations in the spatial distributions of cells. Under the assumption that the rate of a cell is constant over an entire region

of interest a point pattern will exhibit complete spatial randomness (CSR), and it is often of interest to model whether cells deviate from CSR either through clustering or repulsion. In marked point process theory points, in a point pattern may carry attributes. Example attributes include cell phenotype, cell size, or expression of functional marker PDL1, and this auxiliary information is called a “mark” [105]. Marks can be qualitative (cell phenotype) or quantitative (expression of a functional marker), and marked point processes are particularly useful in the mIF data setting.

Spatial summary statistics can be calculated to quantify the clustering and co-occurrence of cells in a circular region with a particular radius r . Typically univariate (one cell type) or bivariate (two cell types) summary statistics are reported, and inference is obtained by comparing the observed spatial summary statistic to that obtained under CSR. A popular quantity is Ripley’s $K(r)$, which studies the number of neighbors to a particular point within radius r , and has univariate and bivariate implementations in the *spatstat* package [106, 107]. Ripley’s K is characterized by clustering or repulsion depending on whether it is above or below the theoretical value of $2\pi r^2$. Modifications of Ripley’s K , Besag’s $L(r)$ and Marcon’s $M(r)$, are interpreted as clustering or repulsion for values above or below their theoretical limits of r and l , respectively. K and L increase with increasing radius, while M does not [108, 109]. A similar statistic, called the *hypothesized interaction distribution* measures the co-localization between immune cells, and authors found that high co-clustering of certain cell types was associated with worse survival in oroharyngeal squamous cell carcinoma [110, 111]. A popular paper by [5] studying the TME of breast cancer introduced a bivariate adaptation of $K(r)$ that can be interpreted as the number of close interactions between marker positive cells in a given region.

Other spatial summary statistics analyze the distance to a neighbor, and can be interpreted as probabilities of observing a particular cell type within a radius r . One of these metrics, $G(r)$, or the nearest neighbor distance distribution, is the cumulative distribution function of an exponential random variable [105]. Similar quantities include the empty space function $F(r)$, pair correlation function $g(r)$, and radial distribution function $J(r)$ [112, 113]. These types of spatial summaries have been used to study

interactions between T-regulatory cells and tumor cells in NSCLC lung cancer [114], and to study CD68+ macrophages in human head and neck tumors [115]. An excellent and more detailed overview of these metrics, including a table with mathematical formulations, is provided in [13].

Figure 6 shows tumor infiltrating immune cells with their respective Ripley's K functions for three images from the HGSOC dataset. The top row shows TMA cores and cell locations, with immune and other cells colored in red and gray, respectively. In the bottom row are Ripley's $K(r)$ for each image. The estimated $K(r)$ is indicated by the black bold line, and the dotted blue lines show theoretical values of $K(r)$ under complete spatial randomness. The subject in the first column has $K(r)$ values above the CSR line, indicating clustering of tumor infiltrating immune cells; in contrast, the subject in the second column has a similar density of cells but the immune cells (in red) are more randomly dispersed, leading to a $K(r)$ close to the theoretical value under CSR.

For multiplex imaging data it can be highly advantageous to incorporate quantitative marker intensity values with information about point locations, which can be done using second-order summary statistics. At a high level this is typically done by incorporating measures of correlation. Moran's I is a popular global (image-level) measure of spatial autocorrelation which has bivariate and local adaptations [116, 117]. Lee's L is the spatially weighted and smoothed correlation coefficient of two markers [118]. Recently [119] introduced a bivariate summary statistic based on the marker conditional mean. The pairwise correlation function also incorporates continuous marker information.

Point process adaptations for mIF

In mIF data a major goal of estimating region level spatial summaries is to perform analyses using the above spatial indices as outcome or predictor variables [92, 120–122]. It is critical to obtain summary measures that are informative about cell level interactions, but are not sensitive to gross tissue level features such as tissue orientation or shape. Many of the well established spatial summary metrics have corrections for edge effects, which arise when cells are close to the border of the image and are corrected for automatically by functions in the *spatstat* package [106]. However, this is not sufficient in mIF, where estimation issues can arise when the tissue has holes due to the shape of the tissue, folds, or

tears, resulting in patches of areas on the slide where no cells are present. This can bias the estimation of spatial summary statistics, and is common for whole sample mIF data as well as tumor microarrays (see Figure 5, third column). One recent methodological development accounts for regions where no cells were present by permuting an empirical value of complete spatial randomness, and then comparing observed spatial summary statistic values to that obtained by this empirical null distribution [92]. Implementations of this “patch” correction for Ripley’s K and nearest neighbor-based metrics are available in the `spatialTIME` package [120]. In addition, many spatial summary statistics are calculated at a particular distance, r , and the selection of this distance can be arbitrary. An alternative to calculating a single spatial value for each image is to perform inference using the entire spatial summary function evaluated over a range of distances covering the image spatial domain. Functional data analysis regression methods, which have been developed to handle functions as predictors and/or outcomes in linear models, are well suited for this. The additive functional Cox regression model [123] can be used to model spatial summary functions as covariates when the outcome is patient survival or time to disease recurrence, and this approach has been taken to study tumor-immune cross correlation functions in mIF and MIBI imaging data [122].

3.5 Software

Many software tools are in active development to assist in analysis of bioimage, spatial transcriptomics, and spatial proteomics data. Halo and Inform are proprietary softwares with a graphical user interface (GUI) to allow point-and-click analysis. Most other tools are open source, and primarily available in R or Python, though some open-source softwares are GUI-based as well. Some are specialized to address one step of the mIF processing pipeline (e.g., normalization) while others can be used for visualization, normalization, phenotyping, segmentation, and/or spatial analysis. Existing software packages that can be used to analyze multiplex imaging data are highlighted in Table 1.

4. Conclusions

This chapter is a brief overview on steps of the mIF analysis pipeline after single-cell segmentation, including normalization and batch correction, cell phenotyping, analysis of cell proportions and expression, and spatial analysis. We focused on reviewing these components of the pipeline because we believe these are areas where biostatistics can have the greatest impact. There are several areas of research that are beyond the scope of this review, but are important to mention. While we focused on steps of the pipeline downstream of slide staining and image acquisition, many of the image quality issues, in particular batch effects, can be mitigated in the image acquisition process. Some work has focused on maximizing reproducibility in the image acquisition process [124]. Secondly, we assumed effective single-cell segmentation prior to many of the analyses discussed here, but estimation, bias, and variability in single-cell segmentation as well as a survey of current methodologies is, itself, an entire literature to review. In addition, there is increasing attention in multimodal analysis, such as integrating mIF data with spatial transcriptomics, single-cell RNA-seq, and other ‘omics data types to perform patient-level analyses that draw information across these different data types. Finally, the steps of the mIF pipeline are modular, and can each be improved upon independently or in parallel; however, it is critical to evaluate how these methodological developments influence biological interpretations in downstream analysis steps. As long as statistical methods are thoughtfully and carefully applied at each step of the analysis pipeline, there is much to be learned from multiplex data and it will be exciting to see what discoveries will be made with these new technologies.

Software Information			Functionality					
<i>Name</i>	<i>Purpose</i>	<i>Language</i>	<i>Store</i>	<i>Viz</i>	<i>Norm</i>	<i>Seg</i>	<i>Ptype</i>	<i>Spatial</i>
VectraPolarisData	Example mIF datasets	R (Bioc)	x					
histoCAT	Computational histology topography cytometry analysis toolbox.	MATLAB, GUI		x	x			x
Squidpy	Tool for the analysis and visualization of spatial molecular data.	Python		x		x		x
Giotto	Tools to process, analyze and visualize single-cell spatial expression data.	R, GUI, Python depends		x	x			x
imcRTools	Analysis of imaging mass cytometry and other highly multiplexed data.	R (Bioc)	x	x				x
SVCA	Computational framework for the analysis of spatial molecular data.	Python, R depends		x	x			x
SPIAT	Data processing, quality control, visualization, data analysis tools.	R	x					
spatialTIME	Spatial analysis of mIF data	R	x	x				x

mxnorm	Normalization techniques for mIF data	R, Python depends	x	x	x			x
Seurat	Toolkit for QC, analysis, and exploration of single-cell 'omics data.	R		x	x			x
QuPath	Powerful, flexible, extensible platform for bioimage analysis.	GUI, Java		x	x	x	x	x
HALO	Platform for quantitative tissue analysis in digital pathology.	Closed source, GUI		x		x	x	x
inForm	Visualizing, segmenting, and phenotyping images from Vectra instruments	Closed source, GUI		x		x	x	
Cytomapper	Visualization of cell-level info obtained by multiplexed imaging cytometry.	R (Bioc)		x				
CyLinter	Quality control software for multiplex microscopy.	Python				x		
CellProfiler Analyst	Interactive data analysis for high-throughput, image-based experiments.	Python				x		

Table 1: Software tools for data infrastructure (Store), visualization (Viz), normalization (Norm), segmentation (Seg), phenotyping (Ptype), and spatial analysis (Spatial) of multiplex imaging data.

Figure Captions

[Figure 1]

Figure 1: Typical image acquisition, processing, and analysis workflow for mIF/mIHC data.

[Figure 2]

Figure 2: One image from the NSCLC dataset. Left panel is the composite image combining all 8 channels collected. Next three images are three individual channels from the same image, shown, from left to right, the nucleus (DAPI) channel, the cytokeratin channel, and the CD8 channel.

[Figure 3]

Figure 3: Batch effects and results of normalization for the NSCLC data. Top and middle rows: unnormalized marker intensity distributions for the immune cell markers CD14 (top row) and CD19 (middle row). Densities are plotted for three randomly selected subjects, each shown in a different color. Each subject has densities plotted for multiple ROIs (dotted lines) as well as the subject-specific mean density (solid lines). Differences across subjects represent slide-level batch effects. Bottom row: plots of discordance scores of phenotypic marker channels for 15 subjects in the NSCLC data before and after normalization. Simple log₁₀ and mean division transform-based normalization was applied using the *mxnorm* package. Smaller discordance score indicates a reduction in batch effects.

[Figure 4]

Figure 4: Clustering results applied to five ROIs from one subject in the NSCLC dataset. TSNE was first applied to the raw marker values to reduce dimensions for visualization purposes. The cells for this subject were plotted along the first two TSNE directions and colored by phenotypes calculated using Inform software (left), colored by clusters produced by applying Phenograph to log₁₀ mean division normalized marker intensities (right). Phenograph produced 12 clusters for the normalized data, more than are identifiable by the markers collected in the dataset, which must be combined to identify clusters comparable with those obtained by Inform.

[Figure 5]

Figure 5: Cell type composition across groups of women in the HGSOc data with (BRCA+) and without (BRCA-) a BRCA mutation. Left column shows histograms of proportions of tumor-infiltrating immune cells, including B cells, CD4 T cells, CD8 T cells, and macrophages, across subjects in the HGSOc data. Dotted vertical black line is drawn through zero. Note that there is substantial right skew in the distributions of proportions, and many images have zero B-cells. Right panel shows results from generalized linear models of proportions B cells (top) and macrophages (bottom), controlling for BRCA status. Estimates shown are β and 95% confidence intervals for the BRCA covariate from each models. Black stars above estimates indicate statistical significance at the 5% α level. The “naive” model applies linear regression to empirical proportions that have been square-root transformed, “Poisson robust” and “binomial robust” indicate Poisson and binomial regression with robust standard errors, and ZIP indicates zero-inflated Poisson regression. The ZIP model was not used to model proportions of macrophages because all images contained at least one macrophage. For the ZIP model, β and 95% confidence intervals are reported for the count model.

[Figure 6]

Figure 6: Spatial summaries of tumor infiltrating immune cells in three images from the HGSOc ovarian data. Top row shows plots of spatial locations of all cells (gray) and tumor infiltrating immune cells (red). Bottom row shows each estimated Ripley’s K function $K(r)$ for corresponding images in the top row. Estimated $K(r)$ is shown by black bold line, and dotted blue line shows theoretical values of $K(r)$ under complete spatial randomness. The first and second columns each show a subject with representative high or low K values at $r = 100$. The third column shows an image with areas of sparse cell density due to tearing or folding of the TMA tissue.

Acknowledgements

This work was supported by the National Institutes of Health (U2CCA233291), the Colorado Clinical and Translational Sciences Institute (UL1TR002535), and the Lupus Research Alliance (Lupus Innovation Award).

References

1. Bendall SC, Nolan GP, Roederer M, Chattopadhyay PK (2012) A deep profiler’s guide to cytometry. *Trends Immunol* 33:323–332. <https://doi.org/10.1016/j.it.2012.02.010>
2. Young YK, Bolt AM, Ahn R, Mann KK (2016) Analyzing the Tumor Microenvironment by Flow Cytometry. *Methods Mol Biol* 1458:95–110. https://doi.org/10.1007/978-1-4939-3801-8_8
3. Hedvat CV (2010) Digital microscopy: past, present, and future. *Arch Pathol Lab Med* 134:1666–1670. <https://doi.org/10.5858/2009-0579-RAR1.1>
4. Angelo M, Bendall SC, Finck R, et al (2014) Multiplexed ion beam imaging of human breast tumors. *Nat Med* 20:436–442. <https://doi.org/10.1038/nm.3488>
5. Keren L, Bosse M, Marquez D, et al (2018) A Structured Tumor-Immune Microenvironment in Triple Negative Breast Cancer Revealed by Multiplexed Ion Beam Imaging. *Cell* 174:1373–1387.e19. <https://doi.org/10.1016/j.cell.2018.08.039>
6. Giesen C, Wang HAO, Schapiro D, et al (2014) Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nat Methods* 11:417–422. <https://doi.org/10.1038/nmeth.2869>

7. Martinez-Morilla S, Villarroel-Espindola F, Wong PF, et al (2021) Biomarker Discovery in Patients with Immunotherapy-Treated Melanoma with Imaging Mass Cytometry. *Clin Cancer Res* 27:1987–1996. <https://doi.org/10.1158/1078-0432.CCR-20-3340>
8. Baharlou H, Canete NP, Cunningham AL, et al (2019) Mass Cytometry Imaging for the Study of Human Diseases-Applications and Data Analysis Strategies. *Front Immunol* 10:2657. <https://doi.org/10.3389/fimmu.2019.02657>
9. Ehrenberg AJ, Morales DO, Piergies AMH, et al (2020) A manual multiplex immunofluorescence method for investigating neurodegenerative diseases. *J Neurosci Methods* 339:108708. <https://doi.org/10.1016/j.jneumeth.2020.108708>
10. Apaolaza PS, Petropoulou P-I, Rodriguez-Calvo T (2021) Whole-Slide Image Analysis of Human Pancreas Samples to Elucidate the Immunopathogenesis of Type 1 Diabetes Using the QuPath Software. *Front Mol Biosci* 8:689799. <https://doi.org/10.3389/fmolb.2021.689799>
11. Kondo A, Ma S, Lee MYY, et al (2021) Highly Multiplexed Image Analysis of Intestinal Tissue Sections in Patients With Inflammatory Bowel Disease. *Gastroenterology* 161:1940–1952. <https://doi.org/10.1053/j.gastro.2021.08.055>
12. Ramaglia V, Sheikh-Mohamed S, Legg K, et al (2019) Multiplexed imaging of immune cells in staged multiple sclerosis lesions by mass cytometry. *Elife* 8:e48051. <https://doi.org/10.7554/eLife.48051>
13. Wilson CM, Ospina OE, Townsend MK, et al (2021) Challenges and Opportunities in the Statistical Analysis of Multiplex Immunofluorescence Data. *Cancers (Basel)* 13:3031. <https://doi.org/10.3390/cancers13123031>
14. McKinley ET, Shao J, Ellis ST, et al (2022) MIRIAM: A machine and deep learning single-cell segmentation and quantification pipeline for multi-dimensional tissue images. *Cytometry A*. <https://doi.org/10.1002/cyto.a.24541>
15. Schapiro D, Sokolov A, Yapp C, et al (2022) MCMICRO: a scalable, modular image-processing pipeline for multiplexed tissue imaging. *Nat Methods* 19:311–315. <https://doi.org/10.1038/s41592-021-01308-y>
16. Vizcarra JC, Burlingame EA, Hug CB, et al (2022) A community-based approach to image analysis of cells, tissues and tumors. *Computerized Medical Imaging and Graphics* 95:102013
17. Bray F, Ferlay J, Soerjomataram I, et al (2018) GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *Ca Cancer J Clin* 68:394–424
18. Fabbro M, Colombo P-E, Leaha CM, et al (2020) Conditional Probability of Survival and Prognostic Factors in Long-Term Survivors of High-Grade Serous Ovarian Cancer. *Cancers (Basel)* 12:E2184. <https://doi.org/10.3390/cancers12082184>
19. Siegel RL, Miller KD, Fuchs HE, Jemal A (2022) Cancer statistics, 2022. *CA: a cancer journal for clinicians*
20. Remark R, Becker C, Gomez JE, et al (2015) The non-small cell lung cancer immune contexture. A major determinant of tumor characteristics and patient outcome. *American journal of respiratory and critical care medicine* 191:377–390
21. Jordan KR, Sikora MJ, Slansky JE, et al (2020) The capacity of the ovarian cancer tumor microenvironment to integrate inflammation signaling conveys a shorter disease-free interval. *Clinical Cancer Research* 26:6362–6373

22. Shakya R, Nguyen TH, Waterhouse N, Khanna R (2020) Immune contexture analysis in immuno-oncology: applications and challenges of multiplex fluorescent immunohistochemistry. *Clinical & Translational Immunology* 9:e1183
23. Gorris MA, Halilovic A, Rabold K, et al (2018) Eight-color multiplex immunohistochemistry for simultaneous detection of multiple immune checkpoint molecules within the tumor microenvironment. *The Journal of Immunology* 200:347–354
24. Mezheyeuski A, Bergsland CH, Backman M, et al (2018) Multispectral imaging for quantitative and compartment-specific immune infiltrates reveals distinct immune profiles that classify lung cancer patients. *The Journal of pathology* 244:421–431
25. Mori H, Bolen J, Schuetter L, et al (2020) Characterizing the Tumor Immune Microenvironment with Tyramide-Based Multiplex Immunofluorescence. *J Mammary Gland Biol Neoplasia* 25:417–432. <https://doi.org/10.1007/s10911-021-09479-2>
26. Viratham Pulsawatdi A, Craig SG, Bingham V, et al (2020) A robust multiplex immunofluorescence and digital pathology workflow for the characterisation of the tumour immune microenvironment. *Molecular oncology* 14:2384–2402
27. Wrobel J, Ghosh T (2022) VectraPolarisData: Vectra Polaris and Vectra 3 multiplex single-cell imaging data
28. Righelli D, Weber LM, Crowell HL, et al (2022) SpatialExperiment: infrastructure for spatially-resolved transcriptomics data in R using Bioconductor. *Bioinformatics* 38:3128–3131
29. Steinhart B, Jordan KR, Bapat J, et al (2021) The spatial context of tumor-infiltrating immune cells associates with improved ovarian cancer survival. *Molecular Cancer Research* 19:1973–1979
30. Johnson AM, Boland JM, Wrobel J, et al (2021) Cancer cell-specific MHCII expression as a determinant of the immune infiltrate organization and function in the non-small cell lung cancer tumor microenvironment. *bioRxiv*
31. Ahmed Raza SE, Langenkämper D, Sirinukunwattana K, et al (2016) Robust normalization protocols for multiplexed fluorescence bioimage analysis. *BioData Mining* 9:11. <https://doi.org/10.1186/s13040-016-0088-2>
32. Berry S, Giraldo NA, Green BF, et al (2021) Analysis of multispectral imaging with the AstroPath platform informs efficacy of PD-1 blockade. *Science* 372:. <https://doi.org/10.1126/science.aba2609>
33. Fortin J-P, Sweeney EM, Muschelli J, et al (2016) Removing inter-subject technical variability in magnetic resonance imaging studies. *NeuroImage* 132:198–212
34. Harris CR, McKinley ET, Roland JT, et al (2022) Quantifying and correcting slide-to-slide variation in multiplexed immunofluorescence images. *Bioinformatics* btab877. <https://doi.org/10.1093/bioinformatics/btab877>
35. Shinohara RT, Sweeney EM, Goldsmith J, et al (2014) Statistical normalization techniques for magnetic resonance imaging. *NeuroImage: Clinical* 6:9–19
36. Johnson WE, Li C, Rabinovic A (2007) Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* 8:118–127. <https://doi.org/10.1093/biostatistics/kxj037>
37. Graf J, Cho S, McDonough E, et al (2022) FLINO: a new method for immunofluorescence bioimage normalization. *Bioinformatics* 38:520–526. <https://doi.org/10.1093/bioinformatics/btab686>
38. Fortin J-P, Parker D, Tunç B, et al (2017) Harmonization of multi-site diffusion tensor imaging data. *Neuroimage* 161:149–170. <https://doi.org/10.1016/j.neuroimage.2017.08.047>

39. Pomponio R, Erus G, Habes M, et al (2020) Harmonization of large MRI datasets for the analysis of brain imaging patterns throughout the lifespan. *Neuroimage* 208:116450. <https://doi.org/10.1016/j.neuroimage.2019.116450>
40. Chang YH, Chin K, Thibault G, et al (2020) RESTORE: Robust intEnSiTy nORmalization mEthod for multiplexed imaging. *Commun Biol* 3:111. <https://doi.org/10.1038/s42003-020-0828-1>
41. Harris C, Wrobel J, Vandekar S (2022) mxnorm: An R Package to Normalize Multiplexed Imaging Data. *Journal of Open Source Software* 7:4180
42. Wrobel J, Martin ML, Bakshi R, et al (2020) Intensity warping for multisite MRI harmonization. *Neuroimage* 223:117242. <https://doi.org/10.1016/j.neuroimage.2020.117242>
43. Chen B, McKinley ET, Simmons AJ, et al (2021) Human colorectal pre-cancer atlas identifies distinct molecular programs underlying two major subclasses of pre-malignant tumors
44. Rozenblatt-Rosen O, Regev A, Oberdoerffer P, et al (2020) The Human Tumor Atlas Network: Charting Tumor Transitions across Space and Time at Single-Cell Resolution. *Cell* 181:236–249
45. Zhang Y, Parmigiani G, Johnson WE (2020) ComBat-Seq: batch effect adjustment for RNA-Seq count data | bioRxiv. <https://www.biorxiv.org/content/10.1101/2020.01.13.904730v1>. Accessed 18 Aug 2020
46. Freytag S, Tian L, Lönnstedt I, et al (2018) Comparison of clustering tools in R for medium-sized 10x Genomics single-cell RNA-sequencing data. *F1000Research* 7:
47. Menon V (2018) Clustering single cells: a review of approaches on high-and low-depth single-cell RNA-seq data. *Briefings in Functional Genomics* 17:240–245. <https://doi.org/10.1093/bfpg/elx044>
48. Rashid R, Gaglia G, Chen Y-A, et al (2019) Highly multiplexed immunofluorescence images and single-cell data of immune markers in tonsil and lung cancer. *Sci Data* 6:323. <https://doi.org/10.1038/s41597-019-0332-y>
49. Dao D, Fraser AN, Hung J, et al (2016) CellProfiler Analyst: interactive data exploration, analysis and classification of large biological image sets. *Bioinformatics* 32:3210–3212
50. Jones TR, Kang IH, Wheeler DB, et al (2008) CellProfiler Analyst: data exploration and analysis software for complex image-based screens. *BMC bioinformatics* 9:1–16
51. Stirling DR, Carpenter AE, Cimini BA (2021) CellProfiler Analyst 3.0: accessible data exploration and machine learning for image analysis. *Bioinformatics* 37:3992–3994
52. Schapiro D, Jackson HW, Raghuraman S, et al (2017) histoCAT: analysis of cell phenotypes and interactions in multiplex image cytometry data. *Nat Methods* 14:873–876. <https://doi.org/10.1038/nmeth.4391>
53. Baker G (2021) CyLinter
54. Piccinini F, Balassa T, Szkalitsy A, et al (2017) Advanced Cell Classifier: User-Friendly Machine-Learning-Based Software for Discovering Phenotypes in High-Content Imaging Data. *Cell Systems* 4:651–655.e5. <https://doi.org/10.1016/j.cels.2017.05.012>
55. Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. *J Stat Mech* 2008:P10008. <https://doi.org/10.1088/1742-5468/2008/10/P10008>
56. Kiselev VY, Andrews TS, Hemberg M (2019) Challenges in unsupervised clustering of single-cell RNA-seq data. *Nat Rev Genet* 20:273–282. <https://doi.org/10.1038/s41576-018-0088-9>

57. Levine JH, Simonds EF, Bendall SC, et al (2015) Data-driven phenotypic dissection of AML reveals progenitor-like cells that correlate with prognosis. *Cell* 162:184–197
58. Traag VA, Waltman L, van Eck NJ (2019) From Louvain to Leiden: guaranteeing well-connected communities. *Sci Rep* 9:5233. <https://doi.org/10.1038/s41598-019-41695-z>
59. Dries R, Zhu Q, Dong R, et al (2021) Giotto: a toolbox for integrative analysis and visualization of spatial expression data. *Genome Biology* 22:78. <https://doi.org/10.1186/s13059-021-02286-2>
60. Hao Y, Hao S, Andersen-Nissen E, et al (2021) Integrated analysis of multimodal single-cell data. *Cell* 184:3573–3587.e29. <https://doi.org/10.1016/j.cell.2021.04.048>
61. Grün D, Lyubimova A, Kester L, et al (2015) Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature* 525:251–255. <https://doi.org/10.1038/nature14966>
62. Bodenheimer T, Halappanavar M, Jefferys S, et al (2020) FastPG: Fast clustering of millions of single cells. 2020.06.19.159749
63. Samusik N, Good Z, Spitzer MH, et al (2016) Automated mapping of phenotype space with single-cell data. *Nat Methods* 13:493–496. <https://doi.org/10.1038/nmeth.3863>
64. Stassen SV, Siu DMD, Lee KCM, et al (2020) PARC: ultrafast and accurate clustering of phenotypic data of millions of single cells. *Bioinformatics* 36:2778–2786. <https://doi.org/10.1093/bioinformatics/btaa042>
65. Bortolomeazzi M, Montorsi L, Temelkovski D, et al (2022) A SIMPLI (Single-cell Identification from MultiPLexed Images) approach for spatially-resolved tissue phenotyping at single-cell resolution. *Nature Communications* 13:1–14
66. Seal S, Wrobel J, Johnson AM, et al (2021) On Clustering for Cell Phenotyping in Multiplex Immunohistochemistry (mIHC) and Multiplexed Ion Beam Imaging (MIBI) Data
67. Pratapa A, Doron M, Caicedo JC (2021) Image-based cell phenotyping with deep learning. *Current opinion in chemical biology* 65:9–17
68. Alexandrov T, Chernyavsky I, Becker M, et al (2013) Analysis and Interpretation of Imaging Mass Spectrometry Data by Clustering Mass-to-Charge Images According to Their Spatial Similarity. *Anal Chem* 85:11189–11195. <https://doi.org/10.1021/ac401420z>
69. Alexandrov T, Kobarg JH (2011) Efficient spatial segmentation of large imaging mass spectrometry datasets with spatially aware clustering. *Bioinformatics* 27:i230–i238. <https://doi.org/10.1093/bioinformatics/btr246>
70. Krueger R, Beyer J, Jang W-D, et al (2020) Facetto: Combining Unsupervised and Supervised Learning for Hierarchical Phenotype Analysis in Multi-Channel Image Data. *IEEE Transactions on Visualization and Computer Graphics* 26:227–237. <https://doi.org/10.1109/TVCG.2019.2934547>
71. Rajaram S, Pavie B, Wu LF, Altschuler SJ (2012) PhenoRipper: software for rapidly profiling microscopy images. *Nat Methods* 9:635–637. <https://doi.org/10.1038/nmeth.2097>
72. Uhlmann V, Singh S, Carpenter AE (2016) CP-CHARM: Segmentation-free image classification made accessible. *BMC Bioinformatics* 17:. <https://doi.org/10.1186/s12859-016-0895-y>
73. Diem K, Magaret A, Klock A, et al (2015) Image analysis for accurately counting CD4+ and CD8+ T cells in human tissue. *Journal of virological methods* 222:117–121
74. Dobson AJ, Barnett AG (2018) An introduction to generalized linear models. Chapman and Hall/CRC

75. MacKinnon JG, White H (1985) Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties. *Journal of econometrics* 29:305–325
76. Zeileis A (2006) Object-oriented Computation of Sandwich Estimators. *Journal of Statistical Software* 16:1–16. <https://doi.org/10.18637/jss.v016.i09>
77. Hu T, Gallins P, Zhou Y-H (2018) A zero-inflated beta-binomial model for microbiome data analysis. *Stat* 7:e185
78. Martin BD, Witten D, Willis AD (2020) Modeling microbial abundances and dysbiosis with beta-binomial regression. *The annals of applied statistics* 14:94
79. Jakaitiene A, Avino M, Guarracino MR (2017) Beta-binomial model for the detection of rare mutations in pooled next-generation sequencing experiments. *Journal of Computational Biology* 24:357–367
80. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* 15:1–21
81. Robinson MD, Smyth GK (2008) Small-sample estimation of negative binomial dispersion, with applications to SAGE data. *Biostatistics* 9:321–332
82. Liang K-Y, Zeger SL (1986) Longitudinal data analysis using generalized linear models. *Biometrika* 73:13–22
83. Zeger SL, Liang K-Y, Albert PS (1988) Models for longitudinal data: a generalized estimating equation approach. *Biometrics* 1049–1060
84. Lee AH, Wang K, Scott JA, et al (2006) Multi-level zero-inflated Poisson regression modelling of correlated count data with excess zeros. *Statistical methods in medical research* 15:47–61
85. Jiang S, Xiao G, Koh AY, et al (2021) A Bayesian zero-inflated negative binomial regression model for the integrative analysis of microbiome data. *Biostatistics* 22:522–540
86. Zhang X, Yi N (2020) Fast zero-inflated negative binomial mixed modeling approach for analyzing longitudinal metagenomics data. *Bioinformatics* 36:2345–2351
87. Vuong QH (1989) Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica: Journal of the Econometric Society* 307–333
88. Desmarais BA, Harden JJ (2013) Testing for zero inflation in count models: Bias correction for the Vuong test. *The Stata Journal* 13:810–835
89. He H, Zhang H, Ye P, Tang W (2019) A test of inflated zeros for Poisson regression models. *Statistical methods in medical research* 28:1157–1169
90. Sanchez K, Kim I, Chun B, et al (2021) Multiplex immunofluorescence to measure dynamic changes in tumor-infiltrating lymphocytes and PD-L1 in early-stage breast cancer. *Breast Cancer Research* 23:1–15
91. Hammond MEH, Hayes DF, Dowsett M, et al (2010) American Society of Clinical Oncology/College of American Pathologists guideline recommendations for immunohistochemical testing of estrogen and progesterone receptors in breast cancer (unabridged version). *Archives of pathology & laboratory medicine* 134:e48–e72
92. Wilson C, Soupir AC, Thapa R, et al (2022) Tumor immune cell clustering and its association with survival in African American women with ovarian cancer. *PLoS computational biology* 18:e1009900

93. Altman DG, Lausen B, Sauerbrei W, Schumacher M (1994) Dangers of using “optimal” cutpoints in the evaluation of prognostic factors. *JNCI: Journal of the National Cancer Institute* 86:829–835
94. Mabikwa OV, Greenwood DC, Baxter PD, Fleming SJ (2017) Assessing the reporting of categorised quantitative variables in observational epidemiological studies. *BMC health services research* 17:1–8
95. Seal S, Vu T, Ghosh T, et al (2021) DenVar: Density-based Variation analysis of multiplex imaging data. *bioRxiv*
96. Harrell FE (2015) *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*. Springer
97. Dries R, Chen J, Del Rossi N, et al (2021) Advances in spatial transcriptomic data analysis. *Genome Research* 31:1706–1718
98. Stoltzfus CR, Filipek J, Gern BH, et al (2020) CytoMAP: a spatial analysis toolbox reveals features of myeloid cell organization in lymphoid tissues. *Cell reports* 31:107523
99. Palla G, Spitzer H, Klein M, et al (2022) Squidpy: a scalable framework for spatial omics analysis. *Nature methods* 19:171–178
100. Schürch CM, Bhate SS, Barlow GL, et al (2020) Coordinated Cellular Neighborhoods Orchestrate Antitumoral Immunity at the Colorectal Cancer Invasive Front. *Cell* 182:1341-1359.e19. <https://doi.org/10.1016/j.cell.2020.07.005>
101. Li M, Santpere G, Kawasawa YI, et al (2018) Integrative functional genomic analysis of human brain development and neuropsychiatric risks. *Science* 362:eaat7615. <https://doi.org/10.1126/science.aat7615>
102. Zhao E, Stone MR, Ren X, et al (2021) Spatial transcriptomics at subspot resolution with BayesSpace. *Nature biotechnology* 39:1375–1384
103. Wu S, Joseph A, Hammonds AS, et al (2016) Stability-driven nonnegative matrix factorization to interpret spatial gene expression and build local gene networks. *Proceedings of the National Academy of Sciences* 113:4290–4295
104. Chen Z, Soifer I, Hilton H, et al (2020) Modeling multiplexed images with spatial-LDA reveals novel tissue microenvironments. *Journal of Computational Biology* 27:1204–1218
105. Baddeley A, Rubak E, Turner R (2015) *Spatial point patterns: methodology and applications with R*. CRC press
106. Baddeley A, Turner R (2005) Spatstat: an R package for analyzing spatial point patterns. *Journal of statistical software* 12:1–42
107. Ripley BD (1976) The second-order analysis of stationary point processes. *Journal of applied probability* 13:255–266
108. Besag J, Diggle PJ (1977) Simple Monte Carlo tests for spatial pattern. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 26:327–333
109. Marcon E, Puech F, Traissac S (2012) Characterizing the relative spatial structure of point patterns. *International Journal of Ecology* 2012:

110. Tsakiroglou AM, Fergie M, Oguejiofor K, et al (2020) Spatial proximity between T and PD-L1 expressing cells as a prognostic biomarker for oropharyngeal squamous cell carcinoma. *British journal of cancer* 122:539–544
111. Rose CJ, Naidoo K, Clay V, et al (2013) A statistical framework for analyzing hypothesized interactions between cells imaged using multispectral microscopy and multiple immunohistochemical markers. *Journal of pathology informatics* 4:
112. Baddeley A, Gill RD (1997) Kaplan-Meier estimators of distance distributions for spatial point processes. *The Annals of Statistics* 263–292
113. Baddeley AJ, Gill RD (1994) *The empty space hazard of a spatial pattern*. University of Western Australia. Department of Mathematics Perth, Australia
114. Barua S, Fang P, Sharma A, et al (2018) Spatial interaction of tumor cells and regulatory T cells correlates with survival in non-small cell lung cancer. *Lung Cancer* 117:73–79
115. Bull JA, Macklin PS, Quaiser T, et al (2020) Combining multiple spatial statistics enhances the description of immune cell localisation within tumours. *Scientific reports* 10:1–12
116. Moran PA (1948) The interpretation of statistical maps. *Journal of the Royal Statistical Society Series B (Methodological)* 10:243–251
117. Anselin L (1995) Local indicators of spatial association—LISA. *Geographical analysis* 27:93–115
118. Lee S-I (2001) Developing a bivariate spatial association measure: an integration of Pearson's r and Moran's I . *Journal of geographical systems* 3:369–385
119. Chervoneva I, Peck AR, Yi M, et al (2021) Quantification of spatial tumor heterogeneity in immunohistochemistry staining images. *Bioinformatics* 37:1452–1460. <https://doi.org/10.1093/bioinformatics/btaa965>
120. Creed JH, Wilson CM, Soupir AC, et al (2021) spatialTIME and iTIME: R package and Shiny application for visualization and analysis of immunofluorescence data. *Bioinformatics* 37:4584–4586. <https://doi.org/10.1093/bioinformatics/btab757>
121. Maisel BA, Yi M, Peck AR, et al (2022) Spatial Metrics of Interaction between CD163-Positive Macrophages and Cancer Cells and Progression-Free Survival in Chemo-Treated Breast Cancer. *Cancers* 14:308
122. Vu T, Wrobel J, Bitler BG, et al (2021) SPF: A Spatial and Functional Data Analytic Approach to cell Imaging data. *bioRxiv*
123. Cui E, Crainiceanu CM, Leroux A (2021) Additive functional Cox model. *Journal of Computational and Graphical Statistics* 30:780–793
124. Laberiano-Fernández C, Hernández-Ruiz S, Rojas F, Parra ER (2021) Best Practices for Technical Reproducibility Assessment of Multiplex Immunofluorescence. *Frontiers in Molecular Biosciences* 8: