

# Society for Immunotherapy of Cancer: updates and best practices for multiplex immunohistochemistry (IHC) and immunofluorescence (IF) image analysis and data sharing

Janis M Taube,<sup>1,2</sup> Joel C Sunshine,<sup>2</sup> Michael Angelo,<sup>3</sup> Guray Akturk,<sup>4</sup> Margaret Eminizer,<sup>5,6</sup> Logan L Engle ,<sup>2</sup> Cláudia S Ferreira,<sup>7</sup> Sacha Gnjatich,<sup>8</sup> Benjamin Green,<sup>2</sup> Shirley Greenbaum,<sup>3</sup> Noah F Greenwald,<sup>3</sup> Cyrus V Hedvat,<sup>9</sup> Travis J Hollmann,<sup>10</sup> Daniel Jiménez-Sánchez,<sup>2</sup> Konstanty Korski,<sup>11</sup> Ana Lako,<sup>12</sup> Edwin R Parra,<sup>13</sup> Marlon C Rebelatto,<sup>14</sup> David L Rimm ,<sup>15</sup> Scott J Rodig,<sup>12</sup> Jamie Rodriguez-Canales,<sup>16</sup> Jeffrey S Roskes,<sup>2,6</sup> Kurt A Schalper ,<sup>15</sup> Emanuel Schenck,<sup>14</sup> Keith E Steele,<sup>17</sup> Michael J Surace,<sup>14</sup> Alexander S Szalay,<sup>1,2,6</sup> Michael T Tetzlaff,<sup>18</sup> Ignacio I Wistuba,<sup>13</sup> Jennifer H Yearley,<sup>4</sup> Carlo B Bifulco<sup>19</sup>

**To cite:** Taube JM, Sunshine JC, Angelo M, *et al.* Society for Immunotherapy of Cancer: updates and best practices for multiplex immunohistochemistry (IHC) and immunofluorescence (IF) image analysis and data sharing. *Journal for ImmunoTherapy of Cancer* 2025;**13**:e008875. doi:10.1136/jitc-2024-008875

JMT and JCS contributed equally.

Accepted 12 November 2024

## ABSTRACT

**Objectives** Multiplex immunohistochemistry and immunofluorescence (mIHC/IF) are emerging technologies that can be used to help define complex immunophenotypes in tissue, quantify immune cell subsets, and assess the spatial arrangement of marker expression. mIHC/IF assays require concerted efforts to optimize and validate the multiplex staining protocols prior to their application on slides. The best practice guidelines for staining and validation of mIHC/IF assays across platforms were previously published by this task force. The current effort represents a complementary manuscript for mIHC/IF analysis focused on the associated image analysis and data management.

**Methods** The Society for Immunotherapy of Cancer convened a task force of pathologists and laboratory leaders from academic centers as well as experts from pharmaceutical and diagnostic companies to develop best practice guidelines for the quantitative image analysis of mIHC/IF output and data management considerations.

**Results** Best-practice approaches for image acquisition, color deconvolution and spectral unmixing, tissue and cell segmentation, phenotyping, and algorithm verification are reviewed. Additional quality control (QC) measures such as batch-to-batch correction and QC for assembled images are also discussed. Recommendations for sharing raw outputs, processed results, key analysis programs and source code, and representative photomicrographs from mIHC/IF assays are included. Lastly, multi-institutional harmonization efforts are described.

**Conclusions** mIHC/IF technologies are maturing and are routinely included in research studies and moving towards clinical use. Guidelines for how to perform and standardize image analysis on mIHC/IF-stained slides will likely contribute to more comparable results across laboratories and pave the way for clinical implementation. A checklist

encompassing these two-part guidelines for the generation of robust data from quantitative mIHC/IF assays will be provided in a third publication from this task force. While the current effort is mainly focused on best practices for characterizing the tumor microenvironment, these principles are broadly applicable to any mIHC/IF assay and associated image analysis.

## INTRODUCTION

The need to better understand the immune response to cancer is a key challenge in the field of immuno-oncology (IO). Closely related is the need to define predictive tissue-based biomarkers of response to available immunotherapies or new drugs under development.<sup>1–3</sup> Some of the most promising, emerging biomarkers for IO use emerging multiplex immunohistochemistry/immunofluorescence (mIHC/IF) technologies to characterize the spatial relationship between multiple cell types, the location of cells within compartments of the tumor microenvironment (TME), for example, tumor versus stroma,<sup>4</sup> and complex coexpression-defined immunophenotypes, including immunoactive marker expression. Specifically, a meta-analysis that compared mIF/IHC assays to programmed death-ligand 1 (PD-L1) IHC, interferon-gamma-related gene signatures, and mutational density for predicting response to anti-PD-(L)1 therapies, showed that mIF/IHC assays had an area under the summary receiver operating characteristic curve on the order of 0.8, while the



© Author(s) (or their employer(s)) 2025. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

## Correspondence to

Dr Janis M Taube;  
jtaube1@jhmi.edu

other modalities had an area under the curves (AUCs) of ~0.65–0.7.

While there are no established guidelines, a validated AUC of 0.8 or above is in keeping with a potential companion diagnostic and may warrant consideration for a biomarker-driven clinical trial.<sup>5</sup> Some examples of mIF/IHC biomarkers that have been shown to predict therapeutic responses with AUCs in this range include: Quantifying the proportion of intratumoral CD8+CD39+ cells<sup>6</sup> or the density of CD8+FoxP3+ T cells in patients with non-small cell lung carcinoma;<sup>7</sup> assessing the density of programmed cell death protein-1 (PD-1)+to PD-L1+ cells within a certain proximity in patients with Merkel cell carcinoma;<sup>8</sup> and a combinatorial biomarker using CD8+FoxP3+PD-1<sup>low/mid</sup>+and CD163+PD-L1– cell densities in patients with advanced melanoma.<sup>9</sup> These technologies have also been used to describe and categorize the TME of different tumor types into geographic immune contexts, also known as “immunotypes” that may be used to further inform our understanding of immune escape mechanisms and predict therapeutic responses.<sup>4,10</sup>

Many analysis strategies for mIHC/IF provide single-cell resolution, while others are analyzed by pixel-based approaches that rely on regional similarities to infer coexpression.<sup>11</sup> Regardless, the resultant rich and complex mIHC/IF slide data typically require specialized computer algorithms for exploration. A comprehensive overview of different mIF/IHC technologies is provided in “The Society for Immunotherapy of Cancer statement on best practices for multiplex immunohistochemistry (IHC) and immunofluorescence (IF) staining and validation”,

and a smaller table for quick reference is also provided here in [table 1](#).

That manuscript also includes a discussion of the optimization and validation of immunolabeling mIHC/IF assay conditions (eg, antibody clone, antigen retrieval) that are critical to achieving the required performance criteria of individual markers have been discussed. Importantly, the digital image processing pipeline for mIHC/IF assays must also be validated and optimized, with quality assurance (QA) and quality controls (QC) applied to all steps from image acquisition and processing through final data output.

This current, companion manuscript describes a general framework for typical image acquisition and analysis workflows, associated QA/QC considerations, inter-institutional harmonization strategies, and data-sharing models. The aim is to define a basic understanding of “responsible use” by investigators to facilitate robust data generation and associated minimum reporting elements. Specifically, we believe the first step is to ensure that investigators are actually reporting key details of how they designed and performed their analytic steps (eg, cell segmentation, phenotyping, batch correction). Broad, detailed reporting of methodology will enable anticipated future analyses of which approaches provide the most robust results for a given research question.

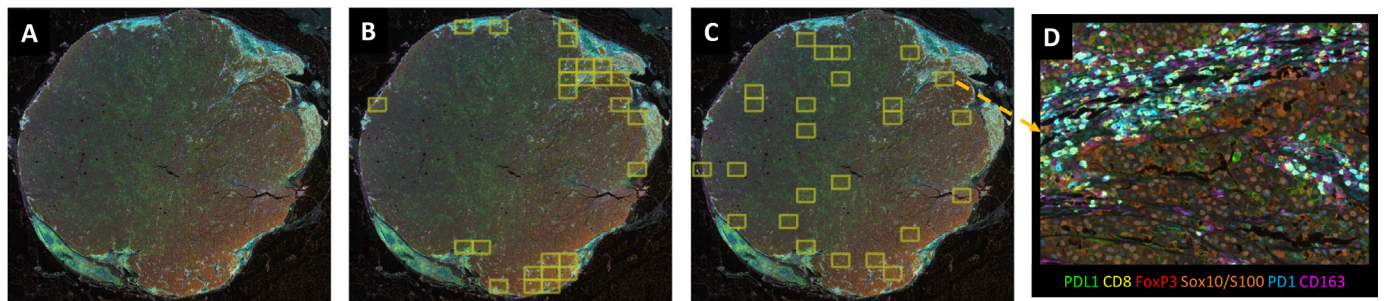
Image acquisition

The first step in digital pathology analysis is to acquire an image of part or all of a slide for analysis. [figure 1](#). Successful image analysis and reliable resultant outputs

**Table 1** Summary of current multiplex IHC/IF technologies (abbreviated from the Society for Immunotherapy of Cancer statement on best practices for multiplex immunohistochemistry (IHC) and immunofluorescence (IF) staining and validation)<sup>11</sup>

|                       | Light microscopy                                                                               |                                                                                                                               | Fluorescence microscopy                                                                                |                                                                                                                                         | Tissue-based mass spectrometry                                                        |
|-----------------------|------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|
|                       | Multiplex IHC                                                                                  | MICSSS                                                                                                                        | Multiplex IF                                                                                           | DSP                                                                                                                                     |                                                                                       |
| Basic description     | Simultaneous/ sequential application of immunostaining without removal of the previous marker. | Iterative cycles of immunostaining, scanning, removal of chromogenic enzyme substrate and blocking previous primary antibody. | Iterative cycles of immunostaining using cyclical stain/stripping, TSA amplification, or DNA barcodes. | 1° antibodies bound to UV cleavable fluorescent DNA tags. A numerical value is generated that corresponds to the # of antibodies bound. | Mass spectrometry imaging of primary antibodies tagged with elemental mass reporters. |
| # markers per section | 3–5                                                                                            | 10+                                                                                                                           | 5–8 for TSA-based, 30–60 for non-TSA-based, cyclical staining approaches.                              | 40–50                                                                                                                                   | 40                                                                                    |
| Imaging area          | Whole slide                                                                                    | Whole slide                                                                                                                   | Up to the whole slide                                                                                  | ROI=0.28 mm <sup>2</sup> (larger areas may be imaged by tiling ROIs).                                                                   | ROI=1.0 mm <sup>2</sup> (larger areas may be imaged by tiling ROIs).                  |

DSP, Digital Spatial Profiling; MICSSS, multiplexed immunohistochemical consecutive staining on single slide; ROI, region of interest; TSA, tyramide signal amplification; UV, ultraviolet.



**Figure 1** Overview slide scan of an mIF stained slide showing different HPF/ROI sampling strategies for image acquisition and subsequent analysis. (A) A 10× resolution overview slide scan from an advanced stage melanoma specimen taken by a Vectra multispectral microscope. The background lymph node and TIL appears as light blue–green surrounding the periphery of the tumor nodule. (B) Representative photomicrograph generated in Python showing HPF/ROI selection targeting “hot spots” with high CD8+T cell density, located here at the tumor–stromal interface. These HPFs can then be acquired (often at higher resolution) and studied in detail. (C) Representative photomicrograph showing sampling of the TME with fields deliberately chosen across both the central and more peripheral regions of the tumor nodule as well as areas with and without a prominent immune cell infiltrate. The aim of this sampling strategy is to capture the potential heterogeneity of the TME. (D) An example of a high resolution HPF taken with a Vectra multispectral microscope that can be used for further downstream analysis. The displayed HPF was generated in Python using an unmixed component image exported from inForm. HPF, high-power field; PD-1, programmed cell death protein-1; PD-L1, programmed death-ligand 1; ROI, region of interest; TIL, tumor-infiltrating lymphocyte; TME, tumor microenvironment.

are significantly dependent on a priori conditions, such as separable color combinations used and scanning procedures (eg, calibrated scanners, well-focused images, reduced tiling effects). For image acquisition, a protocol is typically designed in the application’s software, detailing the microscope objective and exposure time for each filter set. The appropriate per-pixel resolution for acquisition, as well as whether a whole slide (ie, the entire tissue section on a slide) or a specific region of interest (ROI) is imaged, depends on the specific research question, method, and technology used. For example, the multiplexed immunohistochemical consecutive staining on single slide (MICSSS) method typically acquires whole slide images using a bright field scanner.<sup>12</sup> In contrast, for some of the mIF technologies, specific ROIs rather than the entire slide are sometimes acquired.

To date, ROI selection has varied significantly between studies and is typically user-dependent. It is also dependent on the size of the tissue being scanned, for example, core biopsies versus large resection specimens. Previous studies have generally sampled a minimum of five high-power fields (HPFs), ranging from around 0.33–0.64 mm<sup>2</sup> each.<sup>13</sup> Some studies have deliberately sampled regions based on morphological features or immune cell densities (ie, “hotspots” and “coldspots”),<sup>14–17</sup> while others have assessed immune cells in both the tumor core and tumor invasive margin.<sup>18</sup> If a marker or phenotype is rare or extremely heterogeneous, then extended sampling strategies may be required.<sup>9</sup> In these instances, the diagnostic value of ROI selection compared with whole slide acquisition should first be preliminarily explored before deployment on the complete test cohort, including ideally using test, validation, and external validation cohorts.<sup>9 19 20</sup> To ensure reproducibility across studies, it is essential that investigators describe their approach to ROI selection—the number of ROI analyzed per specimen, how the ROI

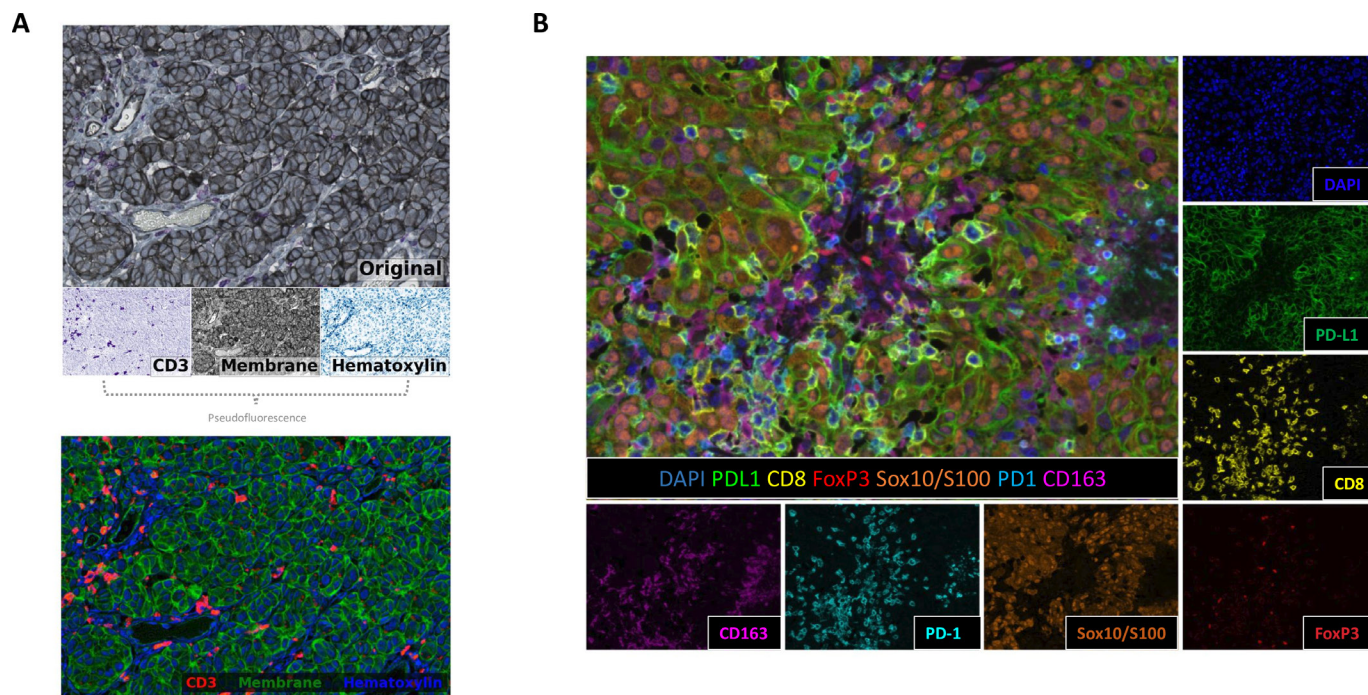
was chosen, and whether there are criteria for ROI inclusion/exclusion at the analysis phase.

Whole slide imaging, although more demanding in terms of time and computation, can be valuable where marker or tissue heterogeneity is high or when complex cell populations are being assessed across multiple large regions. With technological advances, the analysis of whole-slide tissue is becoming a common practice, as opposed to selected HPFs. Further, whole slide imaging can facilitate semiautomated or automated hotspot detection.<sup>9 21–23</sup> Whole-slide imaging followed by automated ROI detection has also been shown to improve the signal-to-noise ratio for certain mIF assays, resulting in improved predictive value.<sup>9 24</sup> This approach also reduces potential HPF/ROI selection bias, which will be of benefit in standardizing outputs across studies and institutions as these technologies move towards clinical use.

### Color deconvolution and spectral unmixing

For both mIHC and mIF, color deconvolution and spectral unmixing, respectively, are essential for an accurate assignment of marker expression. This process has a pronounced impact on the downstream steps of cell segmentation, phenotyping and scoring.

Singleplex chromogenic IHC uses a single primary antibody to detect and visualize the presence of a target protein in tissues or cells, producing a visible color reaction. When multiple markers are used, color deconvolution of the resultant color vectors from red, green, blue (RGB) images can be used to identify the specific contribution of individual stains.<sup>25</sup> For example, color deconvolution algorithms can extract hematoxylin and separate chromogen channels as 8-bit images, figure 2A.<sup>25 26</sup> This results in a separate chromogenic channel for each immunostain, where each chromogen is assigned a unique color code that can be used to generate pseudocolored



**Figure 2** mIHC/IF images contain multiple color vectors that can be separated. (A) mIHC image separated into its three individual color vectors using color deconvolution. Most deconvolution methods are derived from Ruifrok and Johnston,<sup>25</sup> and several open-source tools are available for this purpose. In this case, the image was deconvolved with the automatic color vector estimation tool in the QuPath image analysis platform. Additionally, mIHC images may also be transformed into pseudofluorescent representations, as demonstrated in the bottom panel, which was generated from the images in the top panel. (B) mIF image from a Vectra multispectral microscope of a 6-marker panel plus DAPI. The individual signals are separated by a combination of band-pass filters and unmixing algorithms that capitalize on the emission spectra of each individual fluorophore for isolation. In this example, inForm was used to unmix the individual signals and generate the displayed composite and single channel images. DAPI, 4'6-diamidino-2-phenylindole; IF, immunofluorescence; mIHC, multiplex immunohistochemistry; PD-1, programmed cell death protein-1; PD-L1, programmed death-ligand 1.

images of the tissue, [figure 2B](#). It is important to note that color deconvolution techniques assume a linear relation between stain concentration and absorbance.<sup>27</sup> This is not always the case—for example, after enzymatic chromogen signal saturation, the ratio of target to signal is no longer linear.

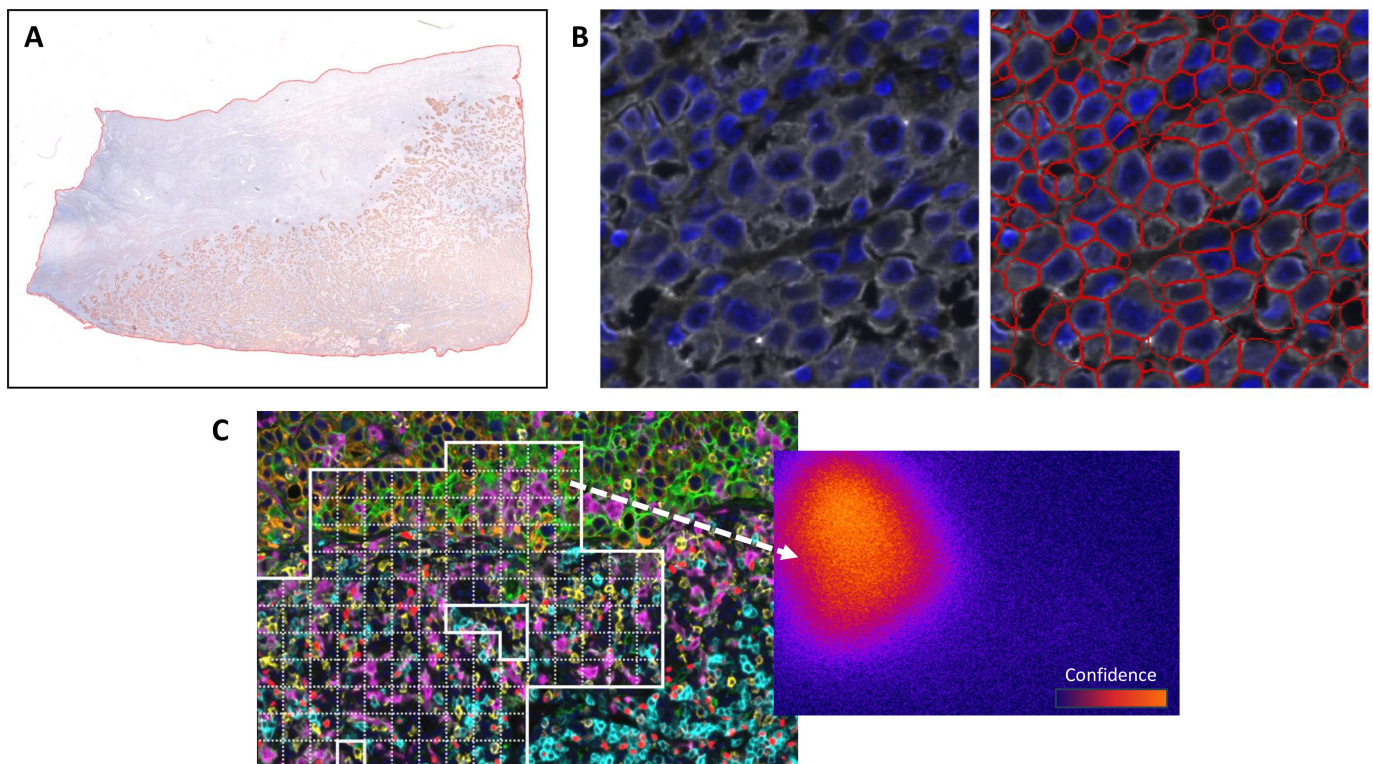
In IF, tissue samples or cell culture preparations are stained with fluorescently labeled antibodies that target specific proteins, with each antibody labeled with a different fluorophore emitting a unique wavelength of light when excited appropriately. The multiple signals are separated with optical filters that allow the selective transmission of light at certain wavelengths while blocking or attenuating light at other wavelengths, [figure 2C](#). Their spectral size can vary from dozens to hundreds of nanometers and is typically selected based on factors such as the spectral properties of the fluorophores, the degree of spectral overlap between different fluorophores, the protein target expression levels, and the desired sensitivity and specificity of detection. If two or more fluorophore signals are captured by the same filter(s), it becomes necessary to unmix the signals.

Spectral unmixing is a mathematical algorithm that separates the multiple fluorophore signatures into individual channels. It is typically facilitated by the generation

of spectral libraries. These libraries are the emission spectra of each individual fluorophore, captured using singleplex-stained tissue samples. The spectral libraries are used as inputs to an unmixing algorithm that determines the contributions of individual fluorophores from composite signals. The most commonly used spectral unmixing algorithm is non-negative matrix factorization, which finds the non-negative abundance coefficients for each fluorophore to reconstruct the acquired signal spectra.<sup>28</sup> Some systems also remove the spectral contribution from tissue autofluorescence, which is the intrinsic natural fluorescence emitted by endogenous molecules within tissues when they are excited by light. Autofluorescence can lead to increased background noise and can be removed using control samples that lack specific fluorophores but contain the same tissue background. Some more recent efforts include strategies for unsupervised spectral unmixing or unmixing without reference spectra of RGB, mIF, and image cytometry signals.<sup>29–31</sup>

### Tissue and cell segmentation

Separated colors and spectra are used as input into image analysis software for the detection of candidate objects, which include tissue, [figure 3A](#), tissue compartments, (tumor vs stroma) and cells or cell compartments,



**Figure 3** Tissue segmentation, cell segmentation, and pixel-based analysis. (A) Representative photomicrograph illustrating tissue segmentation, that is, selection, of the entire tissue area on the slide (red line). This example of tissue segmentation was performed using the Simple Detection tool in the open-source QuPath image analysis platform. (B) An image from a melanoma biopsy stained with DAPI (blue) and a cocktail of CD44/CD45/ATPase to identify cell membranes (white stain).<sup>34</sup> The image was taken using a Vectra multispectral microscope, unmixed with inForm, and the image for display was rendered using Python. The component layers were then passed to the pretrained Mesmer algorithm,<sup>35</sup> and the resultant cell segmentation is shown on the right (red lines indicate boundaries of individual cell membranes). (C) Non-segmented mIF image divided into patches of pixels that can be studied for expression patterns across numerous pixels within the patches. The arrow points to a representative patch, displayed as a heat map of pixel confidence values generated by passing the image patches into the NaroNet algorithm.<sup>44</sup> Confidence reflects the impact of each individual pixel on the machine learning model's prediction of patient outcome. DAPI, 4'6-diamidino-2-phenylindole; mIF, multiplex immunofluorescence.

figure 3B, among other things. The identification of tissue compartments and individual cell boundaries is often referred to as tissue segmentation and cell segmentation, respectively. This detection can be based on image segmentation methods (eg, intensity thresholding, edge detection) and morphology-related image processing operations (may take into account features such as signal intensity, texture, or morphology-related parameters). For example, algorithms could use staining features, for example, a pan-cytokeratin immunostain, to highlight epithelium and distinguish it from the stroma.

Morphological features should be confirmed by a pathologist to prevent potential tissue segmentation errors that can have a significant impact on further analysis, since even small mistakes during this process could result in downstream effects for subsequent steps (see Algorithm verification section below). For example, the quantification of the expression of a given marker by an individual cell or specific subcellular compartment is highly dependent on accurate cell segmentation. Another use of tissue classifiers is the removal of artifacts, such as tissue folds or tears, dust on the slide,

and surgical margin colors or pigments that obscure tissues, by classifying them as a separate category with this methodology and ultimately removing them from the analyzed data.

Single-cell segmentation offers a number of specific advantages. Cell segmentation allows for detailed assessment of the cartesian positioning of individual cells, enabling calculation of cell densities in user-defined samples or tissue areas/compartments and spatial cell-cell interactions, as well as subcellular localizations of protein and/or mRNA expression patterns (nuclear vs cytoplasm vs membrane). Unimarker or multimarker colocalization can be used to identify cell phenotypes and measure the levels of functional markers within specific cell populations (eg, Ki-67, GZB).<sup>32 33</sup> Cell segmentation algorithms typically rely heavily on the detection of the individual cell nuclei. Hematoxylin and 4'6-diamidino-2-phenylindole (DAPI) are commonly used for detecting nuclei in IHC and IF, respectively. Following nuclear detection, basic algorithms will expand the nuclear borders by a selected diameter or until they reach the border of an adjacent cell, in order to mimic the plasma membrane. In doing

so, this also allows for the provision of a cytoplasmic compartment for analysis.

Cell segmentation algorithms based solely on nuclear detection are heavily dependent on both cellular and extracellular uniformity. However, many tissues are composed of cells of diverse sizes and lineages, and cells may be in different planes as tissue is sectioned for slide generation. Ongoing efforts to improve accurate cell membrane detection include the development of robust pan-membrane cell stains to aid image analysis algorithms (as shown in [figure 2A](#)), as well as the utilization of new AI-based algorithms.<sup>34–37</sup>

Cell segmentation algorithms may segment all cell types in the image at the same time, for example, tumors, macrophages, and lymphocytes, or may be trained to segment a single cell morphology at a time. One recent study showed that the larger cells or those that are irregular or dendritic in shape (tumors and macrophages) were often significantly over-segmented when all cell types were processed simultaneously. Improved cell segmentation accuracy was achieved when each cell type was segmented independently.<sup>9</sup> Such a process can be labor intensive and does not always readily extrapolate across data sets without additional training. Cell segmentation of tissues with high lymphocyte densities, for example, lymph nodes, where cells are packed very tightly together, can also be error-prone, and diligent testing of different settings might be required for adequate cell segmentation. Cells where the nucleus is not in the plane of the section also present a segmentation challenge.

Due to some of the challenges with segmenting individual cells as well as some specific morphologic or diagnostic scenarios that benefit from pixel-based strategies, some investigators are forgoing cell segmentation and instead are using a pixel-based approach for analysis, [table 2](#).

These approaches treat each pixel as a unit of analysis and extract information based on the presence and intensity of chromogenic/fluorescent staining, [figure 3C](#).<sup>38–39</sup> This strategy has also been applied to high-parameter tissue imaging using imaging mass cytometry output.<sup>40</sup> Clustering, colocalization, or spatial relationships between different markers can then be assessed for geographical patterns and associations with other clinical and pathologic parameters. A version of this approach is used for “molecular compartmentalization” methods such as automated quantitative analysis and NanoString GeoMx.<sup>41–42</sup> In these approaches, a tissue mask is first defined by one marker, and quantitative analysis of protein expression within that mask (rather than feature extraction algorithms and/or cell segmentation) is used to ascertain the coexpression of other markers. This latter approach may be thought of as a data selection strategy, for example, the aforementioned use of a pan-cytokeratin pixel mask could be used to help delineate tumor versus stroma. Unsupervised deep learning approaches have also been applied to individual pixel-based imaging data.<sup>43</sup> Using technologies such as NaroNet, it is possible to study patches of pixels in specific arrangements to help identify discrete, repeating geographies that associated with patient outcomes,<sup>44</sup> [figure 3D](#). These pixel arrangements then typically need to be characterized using cell segmentation strategies capable of single-cell resolution to help define the underlying identified biology.

Many commonly used image analysis software packages such as HALO, inForm, or plugins for QuPath, among others, are capable of performing cell segmentation as well as determining the number of pixels that are positive above a given threshold in an annotated area. An outstanding question in the field is whether diagnostic or predictive biomarker performance differs between pixel-based and cell segmentation-based strategies. It is also

**Table 2** Comparison between cell segmentation versus pixel-based approaches to signal quantification

| Cell segmentation                                                                                                                                                                                                                                                                                                                                                                                                                                                                                             | Pixel-based analyses                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Advantage: characterization of biology in situ at single cell level<br>Disadvantage: can be labor-intensive to achieve accurate single-cell segmentation.                                                                                                                                                                                                                                                                                                                                                     | Advantage: does not require time-intensive and potentially error-prone process of cell segmentation.<br>Disadvantage: by definition lacks single-cell resolution and thus potential biological insights are not as specific.                                                                                                                                                                                                                                                                                                                                                        |
| May be of specific benefit when addressing: <ul style="list-style-type: none"><li>▶ Expression of a marker by an individual cell, that is, determining phenotypes.</li><li>▶ Expression detection in a subcellular compartment, for example, cytoplasm, nucleus, membrane, for mIF/IHC protein and/or mRNA in situ hybridization technologies.</li><li>▶ Characterizing spatial cell–cell interactions.</li><li>▶ Assessment of cell densities.</li><li>▶ Generating H-scores of marker expression.</li></ul> | May be of specific benefit when addressing: <ul style="list-style-type: none"><li>▶ Marker-based tissue classification or “molecular compartmentalization”.</li><li>▶ Unusually shaped cells providing segmentation challenges—dendritic or spindle cells, or those where nuclei are not in plane of section.</li><li>▶ Assays that rely on a surface area measurement of marker expression.</li><li>▶ Patches of pixels are often used as input for AI approaches.</li><li>▶ Tissue regions lacking cellular detail, for example, tears or folds in tissue, or necrosis.</li></ul> |
| AI, artificial intelligence; IHC, immunohistochemistry; mIF, multiplex immunofluorescence; mRNA, messenger RNA.                                                                                                                                                                                                                                                                                                                                                                                               |                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                     |

important to note that these approaches are not necessarily mutually exclusive. Cell segmentation algorithms are typically trained on an individual project basis and can require extensive manual training to achieve accurate machine-learning algorithm development. In contrast, pixel-based approaches are typically faster to implement since they do not require extended manual machine learning algorithm training. However, the interpretability of the underlying biology that manifests that observed spatial pixel pattern is limited. As such, a combination of pixel-based (for fast discovery and/or exclusion of tissue artifacts) and cell-based (for interpretability) methods could be of benefit. Another example of combining these approaches is to use a pixel-based strategy overlying segmented cells to detect membrane marker polarity to aid in inferring the functional engagement of immune cells.<sup>45</sup> In the future, deep learning strategies may also help alleviate the manual burden currently associated with training accurate cell segmentation algorithms.<sup>46</sup>

### Phenotyping approaches

Phenotyping refers to the classification of cell types as a function of features such as color or spectral intensity, cell shape, or a combination thereof, and typically follows cell segmentation. The determination of whether cells are “positive” for a given feature may be achieved using either thresholding approaches or machine learning classifiers. It is recommended that a description of the phenotyping method used (a threshold value, machine learning training, other) is detailed in the methods of studies reporting mIHC/IF data. The expression intensity of a given marker beyond simply determining whether a cell is positive versus negative for a given marker, may also be of value when determining a phenotype, for example, PD-1<sup>neg</sup>, PD-1<sup>low</sup>, PD-1<sup>mid</sup>, and PD-1<sup>high</sup> phenotypes indicate different functional states of lymphocytes and have also been shown to have biomarker value.<sup>9 47 48</sup>

### Conventional thresholding

Thresholding is the simplest and fastest method to classify cells based on a single feature. Each segmented cell bears various intensity features, for example, color or spectral expression intensity for each subcellular compartment (membranous, cytoplasmic, nuclear, and total cell area), and thresholding uses a cut-off value for feature data to identify a cell population that is considered “positive”. It may also be used in instances where cell segmentation is not performed and individual pixels are quantified instead.<sup>38</sup> The threshold for positivity is typically determined manually by the analyst using visual verification within the software package. Although visual thresholding is widely used, intuitive and commonly accessible, it is subjective and can vary across samples and users.

### Machine learning classifiers

There are two main types of machine learning methods—supervised and unsupervised—both of which may be used to aid in phenotyping cells in digital pathology images.

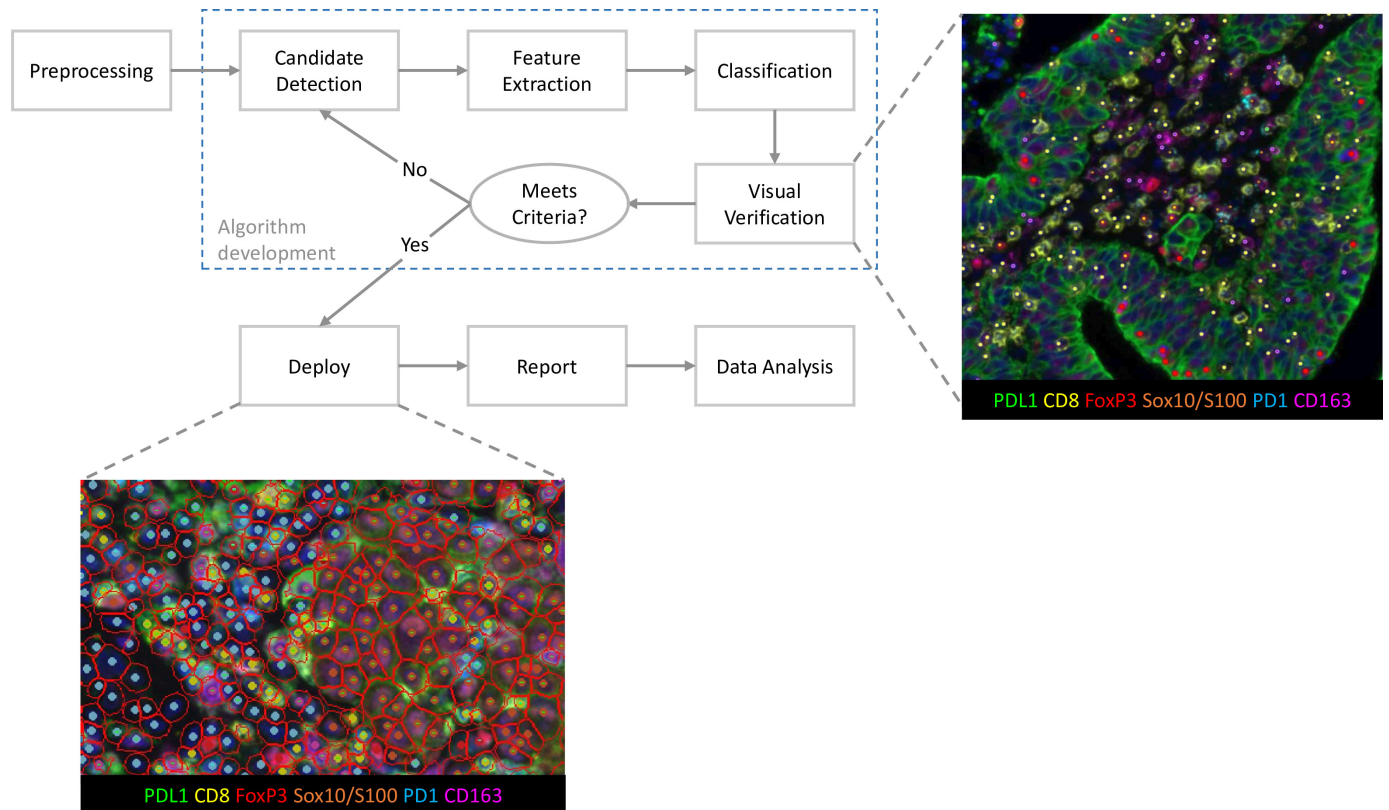
The most common approach to phenotyping involves supervised learning, which is a machine learning process where the algorithm is trained on labeled data, [figure 4](#). In this case, images are manually annotated for a given phenotype, using visual inputs of cell morphology and marker expression. The representative image gallery is then used to generate labels on new cells or cases that the machine had not seen before. Examples of mathematical methods that are employed by supervised algorithms include random forests, boosted decision trees, multidimensional regression, k-nearest neighbor, and support vector machines.

Unsupervised learning, on the other hand, involves using unlabeled data. The objective is for the algorithm to discover patterns, structures, or relationships within the pathology data without any explicit guidance on what to look for. In the context of computational pathology, it can be used to identify subtypes or clusters of similar tissue samples, cell types, or morphologies without prior knowledge of their diagnoses. This can be achieved with k-Means and hidden Markov models. Neural network/deep learning algorithms using weak or minimal supervision have also been described.<sup>49</sup>

Machine learning classifiers and thresholding methods can also be used in combination. For example, PD-L1+tumor cells could be identified by combining morphologic features (irregularly shaped, large, and hyperchromatic nuclei) detected using machine learning, and thresholding (PD-L1 expression above a given intensity). In the future deep-learning approaches may provide additional solutions, including potentially simultaneous segmentation and phenotyping.<sup>46</sup>

### Tissue segmentation, cell segmentation, and phenotyping algorithm verification

Digital pathology image analysis tools applied to multiplexed assays can provide quantitative results, once the algorithm has been optimized for the application of interest.<sup>50</sup> The tissue segmentation, cell segmentation, and phenotyping algorithms all require strong QC of the output data, and best practice guidelines are currently scarce. The pathologist/trained specialist remains the gold standard, and as such, visual verification of the algorithm’s performance is critical at each of these steps to ensure consistent and accurate performance across samples, [figure 4](#). Typically, the user is presented with a selection of fields-of-view (FOVs) where the objects detected by the algorithm are already marked with overlays. The user then adds any missing objects and/or edits false detections to the algorithm’s detection result. The algorithm is processed again to create a new corrected model, which can then be further iterated on, depending on performance. It is important to include enough input data during training to represent the anticipated variation across cases. This typically requires training and algorithm verification across a number of distinct specimens. It is recommended that key choices regarding image analysis algorithm parameters, for example, whether a



**Figure 4** Flowchart illustrating the iterative approach typical for development of a phenotyping algorithm. Representative photomicrographs from a Vectra multispectral microscope showing visualizations of algorithm performance. Typically phenotyping or cell classification algorithms are initially developed on a small subset of data using interactive software, for example, the algorithm for the displayed images was generated in InForm. Once trained, the algorithm can be applied to a larger data set, and visual verification should be performed on this larger subset. In the visual verification example on the right, from a non-small cell lung biopsy and generated in MATLAB using the QC module from Merge a Single Sample (MaSS):<sup>9</sup> the pathologist is presented with a selection of HPFs where algorithm-detected objects are already marked with overlays. The different colored points represent each identified phenotype. The pathologist then evaluates whether the phenotype classification meets an acceptable criteria. If the algorithm fails, adjustments are made to address any missed objects and/or false detections. The revised algorithm is then reprocessed to generate a corrected score which is compared with the original algorithm output. Following final algorithm deployment, identified cells and structures, for example, boundaries, are described by Cartesian coordinates, which form the basis of the output report for data analysis. The photomicrograph in the bottom left, from an advanced melanoma and generated in MATLAB, displays a representation of this final output with red lines showing cell membranes and small dots overlaid to represent the phenotype algorithm output. PD-1, programmed cell death protein-1; PD-L1, programmed death-ligand 1; QC, quality control.

nuclear-detection algorithm was used, and details such as maximum cell size detected, are reported for the machine learning-based approaches. This could even take the form of a screenshot showing the fundamental choices made during image analysis algorithm development, which could be included in supplemental materials. In the case of data sets where the raw images themselves are released, it is suggested that the thresholds that the investigators used for determining positive versus negative signals for each marker are also shared, so others may readily replicate the results.

Algorithm performance may be quantitatively assessed in a number of different ways throughout the image analysis process. For example, with regard to phenotyping, the pathologist/trained specialist can count cells of interest in selected FOV and compare them to results provided by the algorithm. The measurement of

concordance between the user's score(s) and the algorithm score(s) can be assessed using both the slope and the correlation coefficients, such as Pearson's correlation and Lin's concordance correlation coefficient. Together with sensitivity and specificity, this information provides guidance on the possible improvement range for the algorithm, ultimately meeting predetermined acceptance criteria. Ideally, an algorithm performance verification task should be performed by more than one pathologist/trained specialist to account for inter-user variation. In one report describing a novel cell segmentation algorithm, performance was assessed by showing algorithmic and human-determined results to trained pathologists and determining quantitatively whether there was a visually appreciated difference in performance.<sup>35</sup> Metrics such as the F1-score/Dice coefficient for pixel/cell/object detection can be used to quantify an overlap between

the human's reference annotations and the algorithm's output.<sup>34 51</sup> Such visual validation strategies for algorithm performance are not readily applicable to technologies such as GeoMx, and other orthogonal analytic validation approaches are of value.<sup>42</sup>

In theory, it would be ideal to train an image analysis algorithm and use it in a completely automated fashion for all slides in a project or even multiple different projects, that is, a stable-over-time, "locked" algorithm. However, when there are multiple slides in a project with different preanalytical variables and/or slides stained in multiple batches, there may be differences in staining color and fluorophore intensity between slides. In the absence of effective preprocessing steps for color/intensity normalization, using the same threshold(s) across these conditions might decrease accuracy. Well-trained machine learning-based supervised classifiers are sometimes able to overcome this limitation, since they use additional features beyond thresholding for scoring. In either approach, the algorithm performance may "fail" visual inspection or the otherwise defined acceptance criteria for a subset of specimens, thus, the question of how to best handle these cases remains. One option is to count those specimens in another way: either manually by a pathologist/trained specialist or by using a computer-assisted approach with more human involvement. If such an alternate counting approach is taken, it is important to disclose the number of cases this was used for and to show that the specimens are not skewed between study groups to guard against any potential bias, for example, the specimens that failed and had to be recounted using an alternate approach included those from both "responders" and "non-responders" to therapy. Another option is to simply exclude the cases where the algorithm failed from the final analysis. This is the most conservative approach, and while it may suffice for research studies, it is suboptimal as mIHC/IF technologies move toward clinical care. Going forward, algorithms that account and adjust for preanalytical sample variations as well as approaches that normalize for batch-to-batch variation are anticipated to improve and standardize data output quality.<sup>52–54</sup>

## Annotations

Once images are processed, it is possible to select certain regions of the images for inclusion in the analysis and/or for exclusion, such as tissue tears, [figure 5](#). This is typically achieved through manual annotation tools, which facilitate the drawing of different tissue regions, topographic boundaries, [figure 5](#), etc, or help highlight a specific anatomic compartment, for example, portal tracts or tertiary lymphoid structures (TLS). The strategy for annotation should be predefined and consistent across all specimens in a given cohort. Having a single user perform the annotations or at least inspect them all for consistency is recommended. Algorithms that automatically mask tissue tears and folds or other artifacts, for example, formalin deposition in tissue, or which can select features

of interest, such as the aforementioned TLS, have been described and may facilitate throughput.<sup>55</sup>

## Batch-to-batch correction

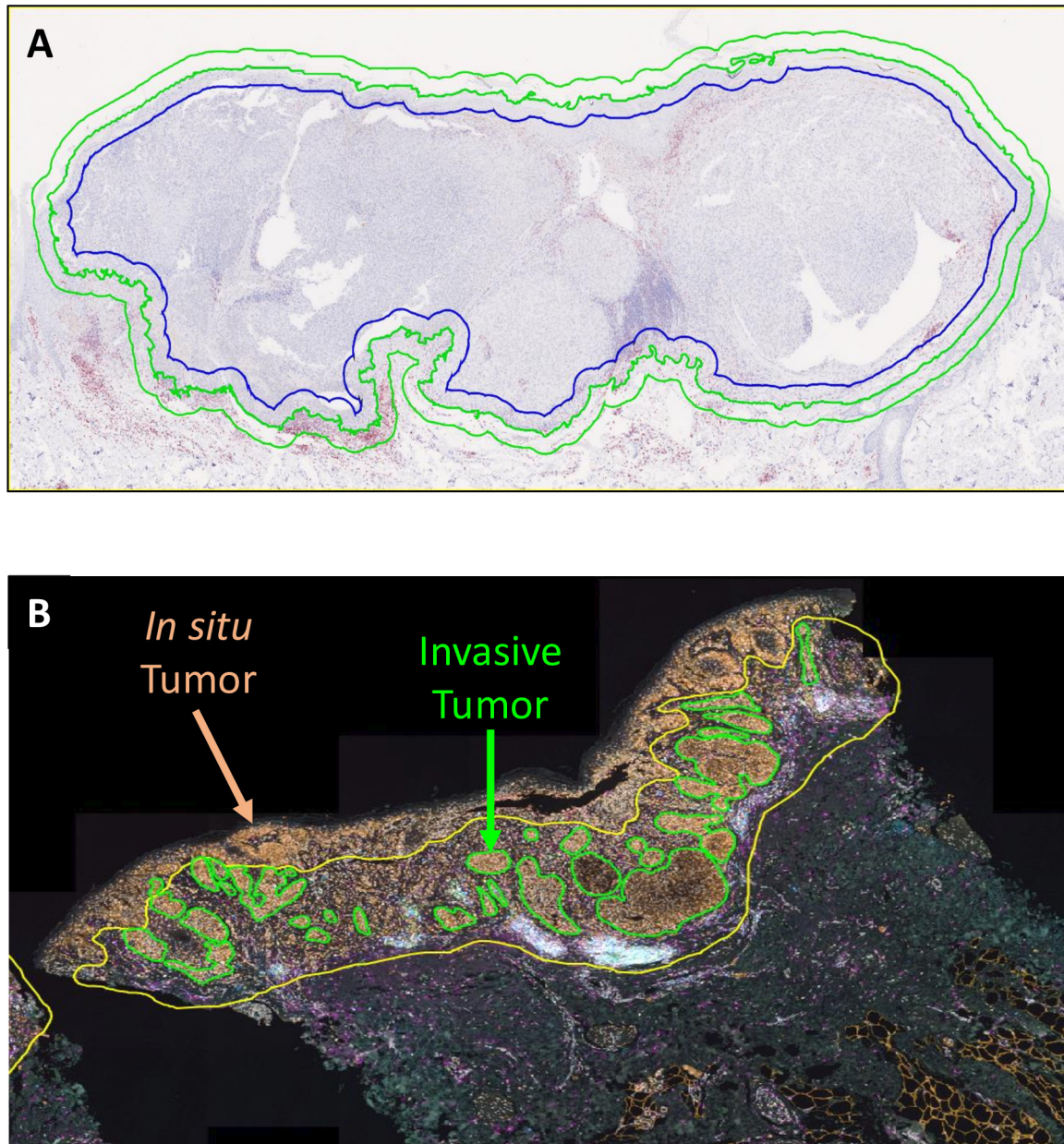
As previously discussed, measurements of individual marker expressions vary from batch to batch, necessitating normalization across batches. This may be due to variable staining performance of instruments between runs, changes in reagents over time, recalibration of instruments, etc.<sup>56</sup> It is important that new image libraries are generated after scanner maintenance to ensure comparative performance before versus after maintenance. Additionally, bulb change information should be collected and included in the meta-data for any study. An additional consideration is the comparison of results across multiple imaging devices, and an approach to aligning illumination intensity and spectral sensitivity between multispectral microscopes has now been described.<sup>57</sup>

Batch-to-batch corrections to address staining and instrument variations may be performed at multiple points in the workflow, depending on the image analysis strategy used. Ideally, the imagery would be corrected for stain intensity variation prior to cell segmentation and subsequent phenotyping; however, it is recognized that for those using off-the-shelf image analysis suites, it is often not possible to perform batch corrections until after segmentation and phenotyping have been performed.

It is possible to use whole slide sections of control tissues, for example, tonsil, for biomarker-positive and biomarker-negative controls. A minimum of 10 different control tissues have been recommended for robust batch effect correction.<sup>58</sup> Tissue microarrays (TMAs), including both tissues and engineered control preparations such as cell line transfectants or beads, are especially well suited to this task as they provide the ability to use multiple different types of tissue, allowing for averaging over the variations expected between different tissue types. Serial slices of TMAs provide near-identical groups of control tissues that may be stained and imaged along with each batch of slides analyzed. Normalization may then be achieved by quantifying differences in marker expression intensity between batches, facilitating more accurate assessments of intensity as an output metric. One study showed that across nine mIF staining batches, the coefficient of variation of expression intensity for PD-1 and PD-L1 was ~20%. When the signals across batches were normalized using the TMA controls, this variance was reduced by 50%.<sup>9</sup>

## Whole slide image QC

As previously noted, when most mIHC/IF images are acquired, they are acquired in smaller HPF regions that must be tiled together, especially when performing whole slide analysis. Some algorithms use microscope-stamped image coordinates to arrange the images together. However, small mechanical deviations, typically caused by an internal motor, can lead to "jitter" in the HPF locations. Such "jitter" should be accounted for in the form



**Figure 5** Annotation tools may be used to help select regions of the TME for either inclusion or exclusion in subsequent analysis. (A) The blue line outlines a tumor nodule, in this example the annotation was generated using HALO. Image analysis tools enable the expansion of this annotation at predetermined increments, for example, 50  $\mu$ m, facilitating a standardized and reproducible approach to characterizing the peritumoral zone. (B) Regions may also be excluded from analysis. In this representative photomicrograph of a skin biopsy stained with a mIF assay and imaged with a Vectra multispectral microscope, invasive melanoma tumor nodules have been annotated (green line) using HALO. The yellow line delineates the TME regions included in the final analysis, deliberately excluding any in situ melanoma in the overlying epidermis, which was the a priori study design. mIF, multiplex immunofluorescence; TME, tumor microenvironment.

of image alignment. Most algorithms for registration use a form of cross-correlation in overlapping regions to find the optimal image coordinates.<sup>9 59 60</sup>

Other forms of image correction facilitating the generation of seamless, whole-slide images from tiled HPFs with consistent illumination include flat-fielding, lens correction, and exposure time corrections. Additional descriptions of these phenomena, their potential impact on mIF data, and the development of corrective factors can be found by Berry *et al.*<sup>9</sup> Some platforms do not already

include algorithms for image correction. In those cases, open-source modules for image correction are available.<sup>61</sup>

### Image registration

Image registration is a key component of MICSSS, tissue-based cyclic immunofluorescence (t-CyCIF), and other “stain and strip” cyclic approaches to mIHC/IF which require multiple rounds of staining and imaging a single slide.<sup>12 62–64</sup> It can also be used to facilitate the analysis of serial slide sections from a sample—with either different

mIHC/IF panels performed on sequential slides, or with the same mIHC/IF panel performed on sequential slides. These may facilitate increased “plex” of an assay, by performing different panels on two sequential slides or even three dimensional reconstruction of a tumor, respectively. Most image registration methods perform a grid search to maximize a similarity measurement. Some measurements are more robust for particular registration experiments and are detailed in the subsequent sections. Moreover, for some experiments regional registration (either manual or automatic) may be sufficient; while for others it may be necessary to use a more local registration, typically made up of a random grid of control points.

#### Repeatedly imaging the same slide stained with a consistent IHC/IF method

“Stain and strip” platforms require stripping and antigen retrieval before every cycle of immunostaining. Each slide is also scanned after immunostaining, and placing slides into scanner racks with small variations changes the coordinates of the tissue between images. This introduces tissue warping and shifts at the microscopic level on each marker image, potentially establishing slightly different total cell numbers for the same ROI. The alignment of images with minimal tissue shifts and warping can be achieved with algorithms like scale-invariant feature transform,<sup>65</sup> speeded-up robust features, Oriented FAST, and Rotated BRIEF, which detect large, local histologic features such as normal ducts or tumor profiles for registration. ROIs annotating the same tissue zone for multiple markers are exported as individual images from whole slide images, and they are either registered

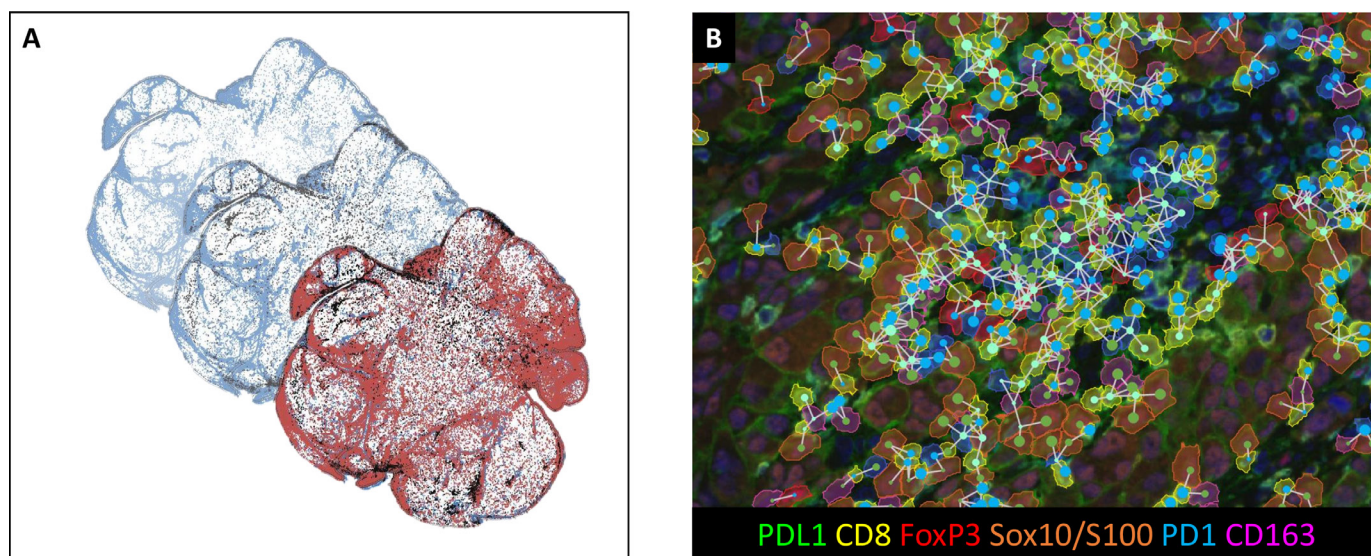
automatically or manually. The Fiji TrakEM2 plugin<sup>66</sup> is one example of a useful tool for this purpose, and can register images automatically or based on selected landmarks. Other more robust algorithms can be applied and may register images using multiple random control points across the entire slide.<sup>59</sup> Algorithms for this kind of registration, typically need less robust registration criteria such as the mean square difference or cross-correlation to fix all cells to single coordinates for different markers on a single slide.

#### Registering images from adjacent slides stained with a consistent IHC/IF method

Serial slide sections taken from a single block may also be stained to increase the “plex” of IHC/IF assays. Serial slide registration is typically performed by putting the two slides into a single coordinate system using an algorithm based on a cross-correlation function minimization, some of which also include deformable or elastic registration to account for image differences (figure 6A). Some examples of publicly available algorithms for serial section registration include Warpy and Ashlar, among others.<sup>59 67</sup> Composite images can be analyzed in image analysis software with similar principles to those used for individual images (figure 6B)<sup>38</sup>

#### Multimodal registration

Multimodal registration, such as combining mIF and mIHC images acquired on the same or adjacent slides, or incorporating spatial transcriptomics, poses additional challenges as compared with registration of the same types of images. The registration approach must account



**Figure 6** Serial section images can be registered and analyzed as though the stains were performed in mIHC/IF on a single slide. (A) IHC for three different markers was performed on sequential slides, which were scanned using a Hamamatsu NanoZoomer digital slide scanner. The images were then registered to each other in a Z-stack using HALO. Image courtesy of Dr Nicolas Giraldo. (B) The individual images can then be merged into a multicomponent image for further analysis using mIHC/IF tools. The MATLAB-generated photomicrograph overlay shows cell geometries colored by phenotype, with dots indicating PD-1 (blue) or PD-L1 (green). Dot size represents PD-1 and PD-L1 expression levels, and lines indicate PD-1+ and PD-L1+ cells within 20µm of each other in this proximity analysis. IF, immunofluorescence; mIHC, multiplex immunohistochemistry; PD-1, programmed cell death protein-1; PD-L1, programmed death-ligand 1.

for image rotation and deformation to register the images and must synthesize different image types into the same coordinate system. These challenges make algorithms based on cross-correlation less reliable; instead, registration algorithms based on mutual information, such as elastix,<sup>68</sup> Warpy,<sup>69</sup> or Whole-Slide Imaging, Mutual Information Registration<sup>34</sup> can be employed.

### Quality metrics for image registration

A lack of ground truth has hampered the generation of quality metrics for image registration. Currently, the reference for image alignment relies largely on a qualitative assessment by pathologists. One approach to allow for a direct comparison between registration methods involves the identification of tissue boundaries by a boundary identification tool, and then using those boundaries to evaluate the maximization of overlap generated by different registration approaches.<sup>34</sup> Another approach could be to evaluate the minimization of overlap of registered images on smaller image regions.<sup>59</sup> Common examples of overlap measurements used for assessment in this way include the Jaccard index, the Dice similarity coefficient, and the Hausdorff distance.

### Proportion, density, neighborhoods, and other spatial analyses

When assessing shifts or differences in cell populations, it is crucial to consider both the proportion and density of cells, as these metrics provide complementary insights into the underlying biology. The choice of analysis should reflect the most unbiased representation of the data. Additionally, spatial dependence is an important factor that can significantly impact results. For instance, while changes in average cell–cell distances can be informative and correlate with clinical outcomes, they often fail to account for variations in cell density or proportions within a sample. To accurately account for spatial dependence, data should be normalized against both cell density and proportion.

Spatial statistics is a well-established field with roots in disciplines like economics, public health, astronomy, and metrology, all of which offer valuable methods for image analysis. Summers, *et al.*, provided an insightful perspective of the diverse applications of spatial statistics in biological image analysis.<sup>70</sup> Broadly, spatial clustering and autocorrelation statistics, such as Moran's I and Geary's C, offer a more robust analysis than simple distance-based measures by inherently accounting for spatial dependence.<sup>71 72</sup>

Identifying spatially coherent regions or neighborhoods is a key focus in multiplex mIF and mIHC image analysis. The rapid growth of spatial transcriptomics has further driven the development of clustering techniques which are also applicable in this space. Most statistical methods start by identifying spatially variable genes or phenotypes using global versions of the spatial autocorrelation statistics,<sup>73 74</sup> with neighbors typically defined through N-nearest neighbor algorithms or a maximum distance threshold between cells. Additionally, these statistics can be used to estimate local autocorrelation for

specific phenotypes, providing more localized insights.<sup>75</sup> Some approaches apply a grid over the cells and compare these grids rather than individual cells directly.<sup>76</sup> Neighbors can then be compared across or between samples to estimate enrichment, clustering, or spatial variance.<sup>77–79</sup>

More advanced methods employ permutation testing to identify statistically significant spatial phenotypes, comparing neighboring regions to a null hypothesis of spatial independence or randomness.<sup>80 81</sup> Others, such as SPARK and SpatialDE, use generalized linear spatial models to model cell expressions—either directly or normalized—offering greater efficiency for larger data sets by comparing modeled distributions.<sup>82 83</sup> Recent efforts have focused on integrating spatially defined phenotypes with histological data and gene expression to cluster neighborhoods or spatial domains.<sup>84–86</sup> However, many of these methods are still computationally intensive and may require significant optimization for large-scale, whole-slide analyses.

### Data and code sharing

Once produced in a validated and reproducible way, the ability to share data is critical. Some consensus on the data structure and related data management requirements that permit effective data sharing is, therefore, obviously needed.

Readouts from mIHC/IF assays include tabulated results of two-dimensional (X, Y) coordinates of pixels and/or cells showing the spatial distribution and intensity of expression for each marker across the imaged portion of the slide. Additional features can also be extracted and reported, such as the shape and texture of cells/objects. For cells that are irregularly shaped, some reporting strategies may not fully describe cell contours, potentially impacting the computed location of the cell membrane when calculating distance measurements. However, various methods are suitable for different applications, so standardization is not indicated here, rather caution regarding clarity when communicating methods used. When image analysis tools are used to segment cells, the tables may also include the coordinates of the centroid of individual cells. This information may be used to calculate outputs such as cell density (eg, # lymphocytes/mm<sup>2</sup>), marker expression intensity (eg, low vs high expression), inter-object distances, percentages of object types in certain tumor areas (eg, % lymphocytes in tumor stroma, % tumor area with macrophage marker expression), etc.<sup>87 88</sup> These algorithms may be in-house coded via scripts (Python, Julia, R, MATLAB), notebooks (Jupyter, Google Colab) or analyzed with publicly available toolboxes like MCMICRO or CytoMAP.<sup>76 89</sup>

Best practices for data sharing include reporting of the raw data from the image files prior to image analysis. This is often in the form of a .csv tabular file or raw image data in czi, tiff, or qptiff formats. A data dictionary should also be provided that defines all abbreviations. This allows other research groups to then explore the data to not only verify results but to test different hypotheses. This

practice also aligns mIHC/IF assays with the new National Institutes of Health (NIH) guidelines for data reporting as of July 1, 2023. Additional key outputs after image analysis that are critical to the conclusion of the paper should also be reported, for example, if a paper is reporting PD-1 to PD-L1 distance metrics as a proposed biomarker, those distances should be displayed in data tables on a per-patient basis.

In addition to sharing data, it is important to share the code required to reproduce the analysis results. The code and its updates should be readable, with clear documentation of inputs and outputs. Ideally, anyone should be able to download the code and either run on the data provided to exactly reproduce the plots in the paper, or run on their own data, processed into appropriate input formats, to produce equivalent plots. The use of versioning software (Git and SVN) and digital repositories (GitHub, Code Ocean, and Zenodo, among others) can be used to streamline code and data sharing, including digital object identifier (DOI) issuance.

Representative photomicrographs are also recommended for inclusion to allow the reader the opportunity to visually verify critical results, for example, key or unusual phenotypes. To date, the resultant raw mIHC/IF images themselves have not been routinely shared in publications using this technology. One major barrier to this is the infrastructure required to host large files, which may be several gigabytes (GB) in size for each whole-slide image. Some individual laboratories have been hosting their own slide scans, for example, <https://datasets.deepcell.org/>. It may also be possible to host images on Synapse ([www.synapse.org](http://www.synapse.org))<sup>90</sup> or a similar public repository, such as the National Cancer Institute's (NCI's) imaging data commons (<https://datacommons.cancer.gov/repository/imaging-data-commons>) which facilitate archiving and DOI generation. Web-browser architectures for visualizing mIHC/IF data may also assist with this task.<sup>91</sup> Irrespective of whether they are posted publicly, we recommend maintaining the raw images as they come off the instrument in a permanent location along with any associated metadata. Going forward, it will be of benefit to have multiple public data repositories that can host large volumes of mIHC/IF images. While these data are more complex than raw sequencing data, mIHC/IF images and associated raw data output share similar needs with other fields of research that increasingly rely on the analysis of so-called "big data". A number of resources are available that describe key features of big data management, including the important principles of findability, accessibility, interoperability, and reusability.<sup>92</sup> When sharing data publicly, it is also important to ensure patient anonymization to ensure privacy.

### Current multi-institutional harmonization efforts

A number of multi-institutional efforts are underway to harmonize mIHC/IF. One harmonization effort performed across six sites on the same mIF platform showed strong inter-site and intrasite concordance of

a 6-plex assay performed on tumor tissue for measures requiring single-cell resolution, for example, density of specific immune cell subsets, coexpression metrics (%PD-L1 on immune cells, etc), and proximity (PD-1 to PD-L1 distance).<sup>56</sup> A second multi-institutional effort was conducted by the NCI-designated Cancer Immune Monitoring and Analysis Centers, which are tasked with providing standardized biomarker assays for NIH-sponsored clinical trials.<sup>93</sup> After harmonizing antibody clones used in a 5-plex assay, they found a strong agreement for immune cell densities characterized by each of the five individual markers in their panel. Notably, this effort compared cell density results generated from MCISSE and brightfield scanning to mIF staining and scanning with spectral unmixing technology. These two distinct multi-institutional efforts had some similar features and findings worth noting: (1) Image analysis was highly coordinated across sites or performed at one site, and did not use a prospectively-developed, locked-down algorithm; (2) The highest concordances were observed when quantifying parameters related to lymphocytes. Lower concordances were observed when quantifying macrophages and other cell types with irregular morphologies; (3) Both studies used either TMAs as input or select HPFs. Similar studies using whole slide images have yet to be completed.

The expectation of these processes is not to achieve perfect agreement between institutions, but rather to quantify inter-institutional variability following harmonization efforts. The results are then considered a benchmark for achievable precision for a specified biomarker assay that can, in turn, be used to establish rational guidelines for interpreting the results of the test when it is applied to clinical samples.

### FUTURE DIRECTIONS

The sections in this manuscript described best practices for processing multiplex images, with a focus on techniques for minimizing systematic errors. Crucial next steps for the field of mIHC/IF image analysis and data sharing include harmonization across platforms where possible, continued standardization, and an improved understanding and quantitation of residual error and its impact. This will help pave the way to broad clinical implementation of these promising technologies.

Both alignment across platforms and reproducibility across different platforms are critical but different goals for the field. Alignment on staining sensitivity can be aided by bridging results from different platforms using the same reference material/samples. Further, if validation is performed to the "gold-standard" of single-plex chromogenic IHC, it ensures a similar minimum threshold of detection across platforms and can facilitate clinical deployment.<sup>13</sup> Similar "minimum standards" of different steps in the image acquisition, processing, and analysis pipeline are the next step, and the development of a checklist, currently underway through the Society

for Immunotherapy of Cancer, may allow for improved interplatform alignment and more consistent comparisons and data generation. Efforts such as the NIH Human Biomolecular Atlas Program (<https://commonfund.nih.gov/hubmap>) will also undoubtedly help drive the continued development of a common, community standard for multiplex data generation and formatting.<sup>94</sup>

In addition, the field needs to go beyond the development of techniques for minimizing systematic errors and evaluate how much uncertainty remains after corrections have been made to better understand how those uncertainties can affect the ultimate analysis, for example, predictions of patient outcomes. If a large fraction of tumor cells are over-segmented, that would impact tumor-cell-dependent outputs, but may have relatively little effect on an analysis that focuses only on lymphocytes. Gaining an understanding of these uncertainties would allow results to be reported with detailed potential errors, strengthening scientific reporting. Additionally, identifying dominant uncertainties can help pinpoint calibrations that need further work to improve future analyses. This can inform methods development, both in the lab and in data analysis.

Guidelines such as those described here are a requisite step for moving multiplexed image analysis systems and technologies into a clinical environment. While the regulatory environment involves many considerations beyond the scope of this manuscript, the considerations of shareability and comparability are germane for diagnostic testing applied to patient samples in a College of American Pathologists accredited, Clinical Laboratory Improvement Amendments certified, and/or US Food and Drug Administration regulated setting. There are also non-trivial logistical and financial considerations when considering incorporating mIHC/IF assays into hospital workflows, including potential laboratory information management system/electronic medical record inclusion of images and questions surrounding payer reimbursement.<sup>95</sup>

## CONCLUSIONS

The tissue biomarker community faces a number of challenges in order to optimize the value of the rich, quantitative data that is rapidly being generated. The comparability of mIHC/IF data produced by individual labs and their reproducibility is a critical issue, since producing the type of rich data sets required to truly understand the TME and the immune response to cancer will require contributions by multiple institutions. Nonetheless, models such as Digital Imaging and Communications in Medicine and NCI Data Commons are available that illustrate the evolution of performance standards, validation requirements, and harmonization practices needed to ensure the quality and comparability of tissue biomarker data.

Here we describe guidelines for the acquisition and processing of digital images from mIHC/IF slides, including approaches to algorithm verification, image QC, quantification of registration accuracy, and batch-to-batch correction, among others. We also describe key considerations for what data should be shared, and how, for appropriate publication of this mIHC/IF assay results.

## Author affiliations

<sup>1</sup>Mark Foundation Center for Advanced Genomics and Imaging, Johns Hopkins University SOM, Baltimore, Maryland, USA

<sup>2</sup>Bloomberg-Kimmel Institute of Cancer Immunotherapy, Johns Hopkins University SOM, Baltimore, Maryland, USA

<sup>3</sup>Department of Pathology, Stanford University School of Medicine, Palo Alto, California, USA

<sup>4</sup>Merck & Co, Rahway, New Jersey, USA

<sup>5</sup>Johns Hopkins University, Baltimore, Maryland, USA

<sup>6</sup>Department of Astronomy and Physics, Johns Hopkins University, Baltimore, Maryland, USA

<sup>7</sup>Pharma Research and Early Development (pRED), Roche Innovation Center Munich, Penzberg, Germany

<sup>8</sup>Icahn School of Medicine at Mount Sinai Tisch Cancer Institute, New York, New York, USA

<sup>9</sup>Department of Pathology, Molecular and Cell-based Medicine, Icahn School of Medicine at Mount Sinai, New York, New York, USA

<sup>10</sup>Department of Pathology and Laboratory Medicine, Memorial Sloan Kettering Cancer Center, New York, New York, USA

<sup>11</sup>Department of Personalized Healthcare, Data, Analytics and Imaging Group, F Hoffmann-La Roche AG, Basel, Switzerland

<sup>12</sup>Dana Farber/Brigham and Women's Cancer Center, Harvard Medical School, Boston, Massachusetts, USA

<sup>13</sup>Department of Translational Molecular Pathology, University of Texas MD Anderson Cancer Center, Houston, Texas, USA

<sup>14</sup>AstraZeneca, Gaithersburg, Maryland, USA

<sup>15</sup>Department of Pathology, Yale University School of Medicine, New Haven, Connecticut, USA

<sup>16</sup>Daiichi Sankyo, Basking Ridge, New Jersey, USA

<sup>17</sup>SR Pathology LLC, Gaithersburg, Maryland, USA

<sup>18</sup>Departments of Pathology and Dermatology, University of California San Francisco, San Francisco, California, USA

<sup>19</sup>Providence Portland Medical Center, Portland, Oregon, USA

**Acknowledgements** The authors would like to acknowledge the work of the Society for Immunotherapy of Cancer (SITC) Pathology Task Force which was instrumental in discussions pertaining to the direction of this work. In addition, authorship would like to acknowledge Katherina von Loga for participation in prior manuscript authorship that pertains to this effort. Medical writing support for the development of this white paper was provided by Hiromi Sato, PhD, and Ben Labbe, PhD, and administrative support was provided by staff of the SITC. Drs. Taube and Szalay would also like to acknowledge support from The Mark Foundation Center for Cancer Research and the Bloomberg-Kimmel Institute for Cancer Immunotherapy at Johns Hopkins University.

**Contributors** All authors contributed to discussions pertaining to the content of this manuscript, as well as drafting and critical review. All authors approved of this manuscript prior to submission.

**Funding** The authors have not declared a specific grant for this research from any funding agency in the public, commercial or not-for-profit sectors.

**Competing interests** JT—Contracted research: Bristol Myers Squibb, Akoya Biosciences; Consulting fees: Bristol Myers Squibb, Roche/Genentech, Merck, AstraZeneca, Regeneron, Lunaphore, Akoya Biosciences, Compugen; Stock holdings: Akoya Biosciences; JCS—Contracted research: Palleon Pharmaceuticals; MA—Consulting fees: Ionpath; Royalties: Ionpath; GA—Salary and employment: Merck; LLE—Royalties: Akoya Biosciences; CSF—Salary and employment: Roche; Stock ownership: Roche; BG—Consulting fees: Rome Therapeutics, PMV Pharma, Merck; Contracted research: Bristol Myers Squibb; IP rights: Rome Therapeutics; SGN—Contracted research: Regeneron, Boehringer Ingelheim, Janssen R&D,

Genentech, Takeda, Bristol Myers Squibb, Celgene; Patents: Named coinventor on an issued patent for multiplex immunohistochemistry to characterize tumors and treatment responses, filed through Icahn School of Medicine at Mount Sinai (ISMMs) and currently unlicensed; CVH—Consulting fees: PathAI; TJH—Salary and employment (current): Bristol Myers Squibb; KK—Salary and employment: F. Hoffman-La Roche AG; AL—Salary and employment: Bristol Myers Squibb; Stock holdings: Bristol Myers Squibb; ERP—Consulting fees: Nucleai, iTeos Belgium SA; MCR—Salary and employment: AstraZeneca; DLR—Consulting fees: AstraZeneca, Cell Signaling Technology, Cepheid, Danaher, Daiichi Sankyo, NextCure, Paige. AI, Regeneron, Sanofi; Royalties: RareCyte; SJR—Consulting fees: Immunitas Therapeutics; Contracted research: Bristol Myers Squibb, Kite/Gilead; JR-C—Salary and employment: Daiichi Sankyo; KAS—Consulting fees: Shattuck Labs, EMD Serono, Clínica Alemana de Santiago, Genmab, Takeda, Merck Sharp & Dohme, Bristol Myers Squibb, AstraZeneca, Agenus, Repertoire; Fees for non-CE services: Bristol Myers Squibb, Fluidigm Corporation, Genmab, Merck, Takeda; Contracted research: Navigate BioPharma, Tesaro/GSK, Moderna, Pierre Fabre, Takeda, Surface Oncology, Merck Sharp & Dohme, Bristol Myers Squibb, AstraZeneca, Ribon Therapeutics, Akoya Biosciences, Boehringer Ingelheim, Eli Lilly; ES—Salary and employment: AstraZeneca; KES—Salary, employment, and ownership: SR Pathology LLC; MJS—Salary and employment: AstraZeneca; Stock holdings: AstraZeneca; IIW—Consulting fees: Genentech/Roche, Bayer, Bristol Myers Squibb, AstraZeneca, Pfizer, Merck, Guardant Health, Flame, Novartis, Sanofi, Daiichi Sankyo, Amgen, Jansen, Merus, AbbVie, Catalyst Therapeutics, Regeneron, Oncocyte; Contracted research: Genentech, Merck, Bristol Myers Squibb, MedImmune, Adaptive, Adaptimmune, EMD Serono, Pfizer, Takeda, Amgen, Karus, Johnson & Johnson, Bayer, Iovance, 4D, Novartis, Akoya; JHY—Salary and employment: Merck; CBB—Ownership: PrimeVax; Consulting fees: Sanofi, Agilent, Roche, Incendia, PrimeVax, BioAI, Lunaphore; Contracted research: Illumina; ME, SGR, NFG, DJ-S, JSR, MTT, ASS—Nothing to disclose; SITC Staff: BL, HS—Nothing to disclose.

**Patient consent for publication** Not applicable.

**Ethics approval** Not applicable.

**Provenance and peer review** Commissioned; externally peer reviewed.

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See <http://creativecommons.org/licenses/by-nc/4.0/>.

**Author note** JT and CBB served as leadership for the Society for Immunotherapy of Cancer (SITC) Pathology Task Force and serve as current leadership for the SITC Pathology Committee.

#### ORCID iDs

Logan L Engle <http://orcid.org/0000-0002-6375-9060>

David L Rimm <http://orcid.org/0000-0001-5820-4397>

Kurt A Schalper <http://orcid.org/0000-0001-5692-4833>

#### REFERENCES

- Yuan J, Hegde PS, Clynes R, *et al.* Novel technologies and emerging biomarkers for personalized cancer immunotherapy. *J Immunotherapy cancer* 2016;4:3.
- Masucci GV, Cesano A, Hawtin R, *et al.* Validation of biomarkers to predict response to immunotherapy in cancer: Volume I — pre-analytical and analytical validation. *J Immunotherapy cancer* 2016;4:76.
- Gnjatic S, Bronte V, Brunet LR, *et al.* Identifying baseline immune-related biomarkers to predict clinical outcome of immunotherapy. *J Immunotherapy cancer* 2017;5:44.
- Halse H, Colebatch AJ, Petrone P, *et al.* Multiplex immunohistochemistry accurately defines the immune context of metastatic melanoma. *Sci Rep* 2018;8:11158.
- Hu C, Dignam JJ. Biomarker-Driven Oncology Clinical Trials: Key Design Elements, Types, Features, and Practical Considerations. *JCO Precis Oncol* 2019;3:1–12.
- Yeong J, Suteja L, Simoni Y, *et al.* Intratumoral CD39<sup>+</sup>CD8<sup>+</sup> T Cells Predict Response to Programmed Cell Death Protein-1 or Programmed Death Ligand-1 Blockade in Patients With NSCLC. *J Thorac Oncol* 2021;16:1349–58.
- Cohen E, Wang D, Engle E, *et al.* 57 CD8<sup>+</sup>FoxP3<sup>+</sup> cells represent early, effector t-cells and predict outcomes in patients with resectable non-small cell lung carcinoma (nscl) receiving neoadjuvant anti-pd-1-based therapy. SITC 37th Annual Meeting (SITC 2022) Abstracts; November 2022
- Fowler EH, Feldman MK. Histochemical demonstration of hydroxysteroid dehydrogenase activity in the mature canine ovary. *Gen Comp Endocrinol* 1970;14:484–90.
- Berry S, Giraldo NA, Green BF, *et al.* Analysis of multispectral imaging with the AstroPath platform informs efficacy of PD-1 blockade. *Science* 2021;372:eaba2609.
- Elliott R. Nursing care study: treatment without surgery for carcinoma of the breast. *Nurs Times* 1980;76:958–60.
- Taube JM, Akturk G, Angelo M, *et al.* The Society for Immunotherapy of Cancer statement on best practices for multiplex immunohistochemistry (IHC) and immunofluorescence (IF) staining and validation. *J Immunother Cancer* 2020;8:e000155.
- Akturk G, Sweeney R, Remark R, *et al.* Multiplexed Immunohistochemical Consecutive Staining on Single Slide (MICSSS): Multiplexed Chromogenic IHC Assay for High-Dimensional Tissue Analysis. *Methods Mol Biol* 2020;2055:497–519.
- Parra ER, Uraoka N, Jiang M, *et al.* Validation of multiplex immunofluorescence panels using multispectral microscopy for immune-profiling of formalin-fixed and paraffin-embedded human tumor tissues. *Sci Rep* 2017;7:13380.
- Nawaz S, Heindl A, Koelble K, *et al.* Beyond immune density: critical role of spatial heterogeneity in estrogen receptor-negative breast cancer. *Mod Pathol* 2015;28:766–77.
- Heindl A, Sestak I, Naidoo K, *et al.* Relevance of Spatial Heterogeneity of Immune Infiltration for Predicting Risk of Recurrence After Endocrine Therapy of ER<sup>+</sup> Breast Cancer. *J Natl Cancer Inst* 2018;110.
- Feldmeyer L, Hudgens CW, Ray-Lyons G, *et al.* Density, Distribution, and Composition of Immune Infiltrates Correlate with Survival in Merkel Cell Carcinoma. *Clin Cancer Res* 2016;22:5553–63.
- AbdulJabbar K, Raza SEA, Rosenthal R, *et al.* Geospatial immune variability illuminates differential evolution of lung adenocarcinoma. *Nat Med* 2020;26:1054–62.
- Pagès F, Mlecnik B, Marliot F, *et al.* International validation of the consensus Immunoscore for the classification of colon cancer: a prognostic and accuracy study. *The Lancet* 2018;391:2128–39.
- Rodriguez-Canales J, Surace M, Cann J. Abstract 491: tumor heterogeneity and its impact on immunoprofiling data: whole tumor vs regions of interest (roi) analysis of multiplex immunofluorescence in four carcinoma types. American Association for Cancer Research (AACR) 2019 Annual Meeting; Atlanta GA, 2019
- Mani NL, Schalper KA, Hatzis C, *et al.* Quantitative assessment of the spatial heterogeneity of tumor-infiltrating lymphocytes in breast cancer. *Breast Cancer Res* 2016;18:78.
- Angell HK, Bruni D, Barrett JC, *et al.* The Immunoscore: Colon Cancer and Beyond. *Clin Cancer Res* 2020;26:332–9.
- Nielsen TO, Leung SCY, Rimm DL, *et al.* Assessment of Ki67 in Breast Cancer: Updated Recommendations From the International Ki67 in Breast Cancer Working Group. *J Natl Cancer Inst* 2021;113:808–19.
- Lu H, Papathomas TG, van Zessen D, *et al.* Automated Selection of Hotspots (ASH): enhanced automated segmentation and adaptive step finding for Ki67 hotspot detection in adrenal cortical cancer. *Diagn Pathol* 2014;9:216.
- Jang MH, Kim HJ, Chung YR, *et al.* A comparison of Ki-67 counting methods in luminal Breast Cancer: The Average Method vs. the Hot Spot Method. *PLoS One* 2017;12:e0172031.
- Ruifrok AC, Johnston DA. Quantification of histochemical staining by color deconvolution. *Anal Quant Cytol Histol* 2001;23:291–9.
- Bankhead P, Loughrey MB, Fernández JA, *et al.* QuPath: Open source software for digital pathology image analysis. *Sci Rep* 2017;7:16878.
- Haub P, Meckel T. A Model based Survey of Colour Deconvolution in Diagnostic Brightfield Microscopy: Error Estimation and Spectral Consideration. *Sci Rep* 2015;5:12096.
- Stack EC, Wang C, Roman KA, *et al.* Multiplexed immunohistochemistry, imaging, and quantitation: a review, with an assessment of Tyramide signal amplification, multispectral imaging and multiplex analysis. *Methods* 2014;70:46–58.
- Jiménez-Sánchez D, Ariz M, Morgado JM, *et al.* NMF-RI: blind spectral unmixing of highly mixed multispectral flow and image cytometry data. *Bioinformatics* 2020;36:1590–8.
- Landini G, Martinelli G, Piccinini F. Colour deconvolution: stain unmixing in histological imaging. *Bioinformatics* 2021;37:1485–7.
- Seo J, Sim Y, Kim J, *et al.* PICASSO allows ultra-multiplexed fluorescence imaging of spatially overlapping proteins without reference spectra measurements. *Nat Commun* 2022;13:2475.

- 32 Topalian SL, Bhatia S, Amin A, *et al.* Neoadjuvant Nivolumab for Patients With Resectable Merkel Cell Carcinoma in the CheckMate 358 Trial. *J Clin Oncol* 2020;38:2476–87.
- 33 Gettinger SN, Choi J, Mani N, *et al.* A dormant TIL phenotype defines non-small cell lung carcinomas sensitive to immune checkpoint blockers. *Nat Commun* 2018;9:3196.
- 34 Doyle J, Green BF, Eminizer M, *et al.* Whole-Slide Imaging, Mutual Information Registration for Multiplex Immunohistochemistry and Immunofluorescence. *Lab Invest* 2023;103:100175.
- 35 Greenwald NF, Miller G, Moen E, *et al.* Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. *Nat Biotechnol* 2022;40:555–65.
- 36 Stevens M, Nanou A, Terstappen LWMM, *et al.* StarDist Image Segmentation Improves Circulating Tumor Cell Detection. *Cancers (Basel)* 2022;14:2916.
- 37 Stringer C, Wang T, Michaelos M, *et al.* Cellpose: a generalist algorithm for cellular segmentation. *Nat Methods* 2021;18:100–6.
- 38 Giraldo NA, Nguyen P, Engle EL, *et al.* Multidimensional, quantitative assessment of PD-1/PD-L1 expression in patients with Merkel cell carcinoma and association with response to pembrolizumab. *J Immunother Cancer* 2018;6:99.
- 39 McCabe A, Dolled-Filhart M, Camp RL, *et al.* Automated quantitative analysis (AQUA) of in situ protein expression, antibody concentration, and prognosis. *J Natl Cancer Inst* 2005;97:1808–15.
- 40 Martinez-Morilla S, Villarroel-Espindola F, Wong PF, *et al.* Biomarker Discovery in Patients with Immunotherapy-Treated Melanoma with Imaging Mass Cytometry. *Clin Cancer Res* 2021;27:1987–96.
- 41 Rubin MA, Zerkowski MP, Camp RL, *et al.* Quantitative determination of expression of the prostate cancer protein alpha-methylacyl-CoA racemase using automated quantitative analysis (AQUA): a novel paradigm for automated and continuous biomarker measurements. *Am J Pathol* 2004;164:831–40.
- 42 Toki MI, Merritt CR, Wong PF, *et al.* High-Plex Predictive Marker Discovery for Melanoma Immunotherapy-Treated Patients Using Digital Spatial Profiling. *Clin Cancer Res* 2019;25:5503–12.
- 43 Jiménez-Sánchez D, López-Janeiro Á, Villalba-Esparza M, *et al.* Weakly supervised deep learning to predict recurrence in low-grade endometrial cancer from multiplexed immunofluorescence images. *NPJ Digit Med* 2023;6:48.
- 44 Jiménez-Sánchez D, Ariz M, Chang H, *et al.* NaroNet: Discovery of tumor microenvironment elements from highly multiplexed images. *Med Image Anal* 2022;78:102384.
- 45 Wu E, Wu Z, Mayer AT, *et al.* PEPsi: Polarity measurements from spatial proteomics imaging suggest immune cell engagement. *Pac Symp Biocomput* 2024;29:492–505.
- 46 Zhang H, AbdulJabbar K, Grunewald T, *et al.* Self-supervised deep learning for highly efficient spatial immunophenotyping. *EBioMedicine* 2023;95:104769.
- 47 Chen DS, Mellman I. Elements of cancer immunity and the cancer-immune set point. *Nature New Biol* 2017;541:321–30.
- 48 Mazzaschi G, Madeddu D, Falco A, *et al.* Low PD-1 Expression in Cytotoxic CD8<sup>+</sup> Tumor-Infiltrating Lymphocytes Confers an Immune-Privileged Tissue Microenvironment in NSCLC with a Prognostic and Predictive Value. *Clin Cancer Res* 2018;24:407–19.
- 49 Wu Z, Trevino AE, Wu E, *et al.* Graph deep learning for the characterization of tumour microenvironments from spatial protein profiles in tissue specimens. *Nat Biomed Eng* 2022;6:1435–48.
- 50 Parra ER, Francisco-Cruz A, Wistuba II. State-of-the-Art of Profiling Immune Contexture in the Era of Multiplexed Staining and Digital Analysis to Study Paraffin Tumor Tissues. *Cancers (Basel)* 2019;11:247.
- 51 Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Med Imaging* 2015;15:29.
- 52 Ahmed Raza SE, Langenkämper D, Sirinukunwattana K, *et al.* Robust normalization protocols for multiplexed fluorescence bioimage analysis. *BioData Min* 2016;9:11.
- 53 Harris CR, McKinley ET, Roland JT, *et al.* Quantifying and correcting slide-to-slide variation in multiplexed immunofluorescence images. *Bioinformatics* 2022;38:1700–7.
- 54 Moutafi M, Robbins CJ, Yaghoobi V, *et al.* Quantitative measurement of HER2 expression to subclassify ERBB2 unamplified breast cancer. *Lab Invest* 2022;102:1101–8.
- 55 Li Z, Jiang Y, Li B, *et al.* Development and Validation of a Machine Learning Model for Detection and Classification of Tertiary Lymphoid Structures in Gastrointestinal Cancers. *JAMA Netw Open* 2023;6:e2252553.
- 56 Taube JM, Roman K, Engle EL, *et al.* Multi-institutional TSA-amplified Multiplexed Immunofluorescence Reproducibility Evaluation (MITRE) Study. *J Immunother Cancer* 2021;9:e002197.
- 57 Eminizer M, Nagy M, Engle EL, *et al.* Comparing and Correcting Spectral Sensitivities between Multispectral Microscopes: A Prerequisite to Clinical Implementation. *Cancers (Basel)* 2023;15:3109.
- 58 Graf J, Cho S, McDonough E, *et al.* FLINO: a new method for immunofluorescence bioimage normalization. *Bioinformatics* 2022;38:520–6.
- 59 Muhlich JL, Chen Y-A, Yapp C, *et al.* Stitching and registering highly multiplexed whole-slide images of tissues and tumors using ASHLAR. *Bioinformatics* 2022;38:4613–21.
- 60 Chalfoun J, Majurski M, Blattner T, *et al.* MIST: Accurate and Scalable Microscopy Image Stitching Tool with Stage Modeling and Error Minimization. *Sci Rep* 2017;7:4988.
- 61 Peng T, Thorn K, Schroeder T, *et al.* A BaSiC tool for background and shading correction of optical microscopy images. *Nat Commun* 2017;8:14836.
- 62 Lin J-R, Izar B, Wang S, *et al.* Highly multiplexed immunofluorescence imaging of human tissues and tumors using t-CyCIF and conventional optical microscopes. *Elife* 2018;7.
- 63 Goltsev Y, Samusik N, Kennedy-Darling J, *et al.* Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed Imaging. *Cell* 2018;174:968–81.
- 64 Cappel G, Dupouy DG, Comino MA, *et al.* Ultra-fast and automated immunohistofluorescent multitasking using a microfluidic tissue processor. *Sci Rep* 2019;9:4489.
- 65 Lowe DG. Proceedings of the seventh IEEE international conference on computer vision 1150. Kerkyra, Greece.
- 66 Cardona A, Saalfeld S, Schindelin J, *et al.* TrakEM2 software for neural circuit reconstruction. *PLoS One* 2012;7:e38011.
- 67 Chiaruttini N, Burri O, Haub P, *et al.* An Open-Source Whole Slide Image Registration Workflow at Cellular Precision Using Fiji, QuPath and Elastix. *Front Comput Sci* 2022;3.
- 68 Klein S, Staring M, Murphy K, *et al.* elastix: a toolbox for intensity-based medical image registration. *IEEE Trans Med Imaging* 2010;29:196–205.
- 69 Shamonin DP, Bron EE, Lelieveldt BPF, *et al.* Fast parallel image registration on CPU and GPU for diagnostic classification of Alzheimer's disease. *Front Neuroinform* 2013;7:50.
- 70 Summers HD, Wills JW, Rees P. Spatial statistics is a comprehensive tool for quantifying cell neighbor relationships and biological processes via tissue image analysis. *Cell Rep Methods* 2022;2:100348.
- 71 MORAN PAP. Notes on continuous stochastic phenomena. *Biometrika* 1950;37:17–23.
- 72 Geary RC. The Contiguity Ratio and Statistical Mapping. *The Inc Stat* 1954;5:115.
- 73 Satija R, Farrell JA, Gennert D, *et al.* Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol* 2015;33:495–502.
- 74 Palla G, Spitzer H, Klein M, *et al.* Squidpy: a scalable framework for spatial omics analysis. *Nat Methods* 2022;19:171–8.
- 75 Anselin L. Local Indicators of Spatial Association—LISA. *Geogr Anal* 1995;27:93–115.
- 76 Stoltzfus CR, Filippek J, Gern BH, *et al.* CytoMAP: A Spatial Analysis Toolbox Reveals Features of Myeloid Cell Organization in Lymphoid Tissues. *Cell Rep* 2020;31:107523.
- 77 Schürch CM, Bhat SS, Barlow GL, *et al.* Coordinated Cellular Neighborhoods Orchestrate Antitumoral Immunity at the Colorectal Cancer Invasive Front. *Cell* 2020;182:1341–59.
- 78 Ester M, Kriegel HP, Sander J. A density-based algorithm for discovering clusters in large spatial databases with noise. Proceedings of the Second International Conference on Knowledge Discovery and Data Mining; Portland, Oregon: AAAI Press, 1996:226–31.
- 79 Campello RJGB, Moulavi D, Zimek A, *et al.* Hierarchical Density Estimates for Data Clustering, Visualization, and Outlier Detection. *ACM Trans Knowl Discov Data* 2015;10:1–51.
- 80 Zhu Q, Shah S, Dries R, *et al.* Identification of spatially associated subpopulations by combining scRNAseq and sequential fluorescence in situ hybridization data. *Nat Biotechnol* 2018.
- 81 Edsgård D, Johnsson P, Sandberg R. Identification of spatial expression trends in single-cell gene expression data. *Nat Methods* 2018;15:339–42.
- 82 Sun S, Zhu J, Zhou X. Statistical analysis of spatial expression patterns for spatially resolved transcriptomic studies. *Nat Methods* 2020;17:193–200.
- 83 Svensson V, Teichmann SA, Stegle O. SpatialDE: identification of spatially variable genes. *Nat Methods* 2018;15:343–6.
- 84 Dries R, Zhu Q, Dong R, *et al.* Giotto: a toolbox for integrative analysis and visualization of spatial expression data. *Genome Biol* 2021;22:78.

- 85 Yuan Z, Zhao F, Lin S, *et al.* Benchmarking spatial clustering methods with spatially resolved transcriptomics data. *Nat Methods* 2024;21:712–22.
- 86 Hu J, Li X, Coleman K, *et al.* SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. *Nat Methods* 2021;18:1342–51.
- 87 Nawaz S, Yuan Y. Computational pathology: Exploring the spatial dimension of tumor ecology. *Cancer Lett* 2016;380:296–303.
- 88 Carstens JL, Correa de Sampaio P, Yang D, *et al.* Spatial computation of intratumoral T cells correlates with survival of patients with pancreatic cancer. *Nat Commun* 2017;8:15095.
- 89 Schapiro D, Sokolov A, Yapp C, *et al.* MCMICRO: a scalable, modular image-processing pipeline for multiplexed tissue imaging. *Nat Methods* 2022;19:311–5.
- 90 Aleynick N, Li Y, Xie Y, *et al.* Cross-platform dataset of multiplex fluorescent cellular object image annotations. *Sci Data* 2023;10:193.
- 91 Manz T, Gold I, Patterson NH, *et al.* Viv: multiscale visualization of high-resolution multiplexed bioimaging data on the web. *Nat Methods* 2022;19:515–6.
- 92 Wilkinson MD, Dumontier M, Aalbersberg IJ, *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 2016;3:160018.
- 93 Akturk G, Parra ER, Gjini E, *et al.* Multiplex Tissue Imaging Harmonization: A Multicenter Experience from CIMAC-CIDC Immunology Biomarkers Network. *Clin Cancer Res* 2021;27:5072–83.
- 94 Snyder MP, Lin S, Posgai A, *et al.* The human body at cellular resolution: the NIH Human Biomolecular Atlas Program. *Nature New Biol* 2019;574:187–92.
- 95 Pourmaleki M, Socci ND, Hollmann TJ, *et al.* Moving Spatially Resolved Multiplexed Protein Profiling toward Clinical Oncology. *Cancer Discov* 2023;13:824–8.