STAR✦Protocols

CellPress

Quantification of immune cell distribution from the vasculature to the glioma microenvironment on sequential immunofluorescence multiplex

Hinda Najem[1,2^], Sebastian Pacheco[1,2], Joanna Kowal[3], Dan Winkowski[4], Jared K. Burks[5], and Amy B. Heimberger[1,2*]

[1]Department of Neurological Surgery, Northwestern University, Chicago, IL, USA
[2]Malnati Brain Tumor Institute of the Robert H. Lurie Comprehensive Cancer Center, Feinberg School of Medicine, Northwestern University, Chicago, IL, USA
[3]Lunaphore Technologies SA, Tolochenaz, Switzerland
[4]Visiopharm A/S, Horsholm, Denmark
[5]MD Anderson Cancer Center, Houston TX, USA
^Technical contact: hinda.najem@northwestern.edu
*Lead contact and Correspondence: Amy B Heimberger, M.D; Northwestern University; SQ6-516; 303 E. Superior Street Chicago, IL 60611; 312-503-3805 amy.heimberger@northwestern.edu

Summary

Except for microglia, immune surveillance and infiltration into the glioma tumor microenvironment (TME) arise from the periphery. Although myeloid-derived immune cell lineages can be dispersed throughout the TME, anti-tumor effector cells are confined to the perivascular space. Therapeutics that induce the dispersal of effector immune cells throughout the TME would be advantageous. This bioinformatic analysis of multiplex images quantifies immune cell distribution in the glioma TME as a function of vessel proximity and will be used in the analysis of prognosis and immunotherapy clinical trials.

Graphical abstract

Cf. Separate file titled "STAR_Graphical Abstract".

Highlights

- Deep-learning algorithm to analyze and quantify complex multiplex imaging.
- Quantifies the immune cell distribution relative to the vasculature within a complex and heterogeneous TME.
- Can be used to establish baseline differences in immune cell functional and phenotypic distribution.
- Will be integrated into the analysis of immune therapeutic clinical trials.

Before you begin

In solid tumors, there is not uniform distribution of immune cells throughout the TME because of chemokines gradients, osmotic pressures, and physical stromal barriers, to name a few. This lack of immune cell distribution is further confounded in the case of central nervous system tumors because of the variable presence of the blood-brain barrier (BBB). Innate immune cells such as macrophages constitute a frequent peripheral originating immune cell population that can distribute widely throughout the glioma TME[1,2]; however, T cell responses are typically confined to the perivascular niche[3]. With the development of strategies that now open the BBB in clinical trials[4] and a variety of immune therapeutics that may enhance distribution[5], quantification of immune cells in various TME regions will be increasingly needed for endpoint evaluations during window-of-opportunity clinical trials[6,7]. The protocol below describes the specific steps for spatial bioinformatic analysis of immune cell distribution from the tumor vasculature on sequential multiplex immunofluorescence images. Baseline immune cell distribution including markers of activation and immune suppression in gliomas, before initiating an immunotherapy clinical trial, were quantified based on the distribution in the TME as a function of the tumor vasculature.

Institutional permissions

Human glioma specimens were collected from consented patients who received surgery at Northwestern Memorial Hospital and Ann & Robert H. Lurie Children's Hospital of Chicago, according to the approved Institutional Review Board protocols STU00214485 and STU00217628, respectively. All experiments conformed to the relevant regulatory standards. Similar permissions are required for other investigators at their respective institutions.

Key resources table

Cf. Separate file titled "Key resources Table Immune Quantification protocol".

Sequential Multiplex Immunofluorescence (SeqIF™) Staining on the COMET™ Platform

Timing: 24-28 hours for 4x 20plex SeqIF™

This technique generates immunofluorescence (IF) multiplex images using the COMET™ platform from Lunaphore Technologies™ on which the subsequent spatial bioinformatic analysis is performed (**Fig. 1A**).

1. Formalin fixed paraffin embedded (FFPE) slides are prepped first.
    a. 4 microns thick tissue sections are mounted on positively charged slides and stored at room temperature until use. Store slides desiccated at room temperature if they will be sitting for more than 2 weeks.

Note: Antigen expression can be lost on cut sections over time, so it is advisable to section tissues within several weeks of the antigen retrieval step.

   b. Antigen retrieval and dewaxing of the FFPE slides is done through incubation in Dewax and Hier retrieval buffer pH9 (Lunaphore Technologies) for 1 hour at 102˚C using the PT module (Epredia).

2. All antibodies are validated using conventional immunohistochemistry and/or IF staining on positive control tissue in conjunction with the corresponding fluorophore and the spectral 4',6-diamidino-2-pheynlindole (DAPI; ThermoFisher Scientific) counterstain.
   a. For optimal concentration and the best signal/noise ratio, antibodies are tested at three different dilutions, starting with the manufacturer-recommended dilution (MRD), then MRD/2 and MRD/4.
   b. Secondary Alexa fluorophore 555 (ThermoFisher Scientific) and Alexa fluorophore 647 (ThermoFisher Scientific) are used at 1/200 and 1/400 dilutions, respectively.
3. The optimizations (characterization 2 and 3) and full protocol runs (Sequential IF) of the multiplex panel are executed using the sequential IF (seqIF™) methodology and protocols integrated into the Lunaphore COMET™ [8].
4. The staining is performed on up to 4 tissue slides simultaneously following automated cycles of staining with 2 antibodies at a time, imaging, and then elution in which no human intervention is required.
   c. All reagents are diluted in Multistaining Buffer (BU06, Lunaphore Technologies).
   d. The elution step lasts 2 min for each cycle and is performed with Elution Buffer (BU07-L, Lunaphore Technologies) at 37°C.
   e. Quenching lasts for 30sec and is performed with Quenching Buffer (BU08-L, Lunaphore Technologies).
   f. Staining step is performed with incubation times set at 4 min for all primary antibodies, and secondary antibodies at 2 min.
   g. Imaging is performed with Imaging Buffer (BU09, Lunaphore Technologies)
      a. Imaging is performed with an integrated epifluorescent microscope at 20x magnification.
      b. Image registration is performed immediately after concluding the staining and imaging procedures by COMET Control Software.
5. Each seqIF™ protocol results in a single multi-layer OME-TIFF file where the imaging outputs from each cycle are stitched and aligned.
   a. COMET™ OME-TIFF files contain a DAPI image, intrinsic tissue autofluorescence in TRITC and Cy5 channels, and a single fluorescent layer per marker.
6. Images generated are automatically saved in an online shared drive within the institution that can be accessed by the users from any computer.

Note: Markers can be analyzed and pseudo-colored for visualization in the Viewer from Lunaphore at any time after the completion of image registration.

Data Transfer to Visiopharm® Software

Timing: 5 min

7. Within the Visiopharm® software 'Database' tab, a new folder can be created according to each project and the respective raw generated multiplex images imported.
    a. By selecting the 'Add' icon from within the Database toolbar, images can be imported by choosing the 'New Images to Database (Import)' setting.

Note: For this protocol, all raw OME.TIFF files are imported and analyzed without any manipulation or background subtraction added. These latter features can be enabled and exported into the images using the Viewer from Lunaphore.

Critical: For the appropriate transfer of OME.TIFF file formats into the software: After the folder that includes the images is selected for transfer, the file type option to choose should be 'All (*)'.

Step-by-step Spatial Bioinformatic method details

Note: Default settings are selected if not otherwise specified.
All intensity ranges for each marker were empirically determined from the positive control staining (for the higher limit) and the signal to noise ratio (for the lower limit).

Step 1: Tissue Segmentation

Timing: 6 hours

This step identifies the tissue areas and eliminates background or non-tissue from subsequent analysis. It also compensates for missing tissue in certain regions of the image resulting from sample degradation during post-resection processing. It ultimately refines the blood vessel shape and improves the accuracy of their identification (Cf. Step 2).

8. Create a new APP for training the "deep learning" algorithm. Two image classes should be used: Class 1 for 'Background' and Class 2 for 'Tissue.'
    a. For 'Input' settings, a magnification of 5X and a field of view of 1024 × 1024 pixels are set. In the 'Regions to Analyze' settings, the ROI chosen for outlining training areas should be used. In this instance, ROI 2.
    b. For 'Classification' settings, the Deep Learning method with the DeepLabv3+ architecture is set. In the 'AI Architect' menu, the input size is set to 512 × 512 pixels, the learning rate adjusted to 1.0E-5, and the minibatch size to 4. Training should be conducted in this step using only the DAPI channel, restricting pixel intensities to a range between 3000-65000 to reduce noise because data is in 16-bit format.

Note: The DAPI channel signal might vary depending on cellular density between tissues. As such, the intensity interval should be adjusted accordingly for the best visualization of the nuclear morphology across tissues.

9. For the Classifier, training areas are outlined using ROI 2. In each image, at least one training area should be generated to enhance the robustness of the classifier and the view of GFAP, DAPI, and CD31 channels should be enabled to help in labeling training areas.
    a. Areas with high GFAP, DAPI, and CD31 signal should be labeled as 'Tissue' under Class 2 (**Fig. 1B**).
    b. Areas without tissue should be assigned to the 'Background' label under Class 1. The software will then automatically label unselected areas within the ROI as either Class 1 or 2 accordingly.

Note: Only a small set of examples are needed in the Class 1 category for the software to identify regions accurately.

10. Training the algorithm of the "tissue segmentation App" on all the images simultaneously increases its accuracy. Therefore, all images should be loaded into the slide tray.
    a. The training is initiated within the 'Classification' section of the App, where monitoring of the "loss function" graph can be done.
    b. Training should be maintained until the loss function converges and plateaus to a final minimal value of <0.5 (the closer to zero the better, indicating a low final loss value). For instance, our APP is trained for approximately 22000 iterations until a plateau at the lowest value of 0.3 is achieved.
11. At this point the "tissue segmentation App" is finalized. To add subsequent steps and move to the next APP, removal of all training areas by clearing ROIs and labels is needed. Editing the 'Post Processing' settings is therefore necessary.
    a. The first step is to convert areas labeled as 'Background' to the 'Clear' label.
    b. Next, it is necessary to surround areas labeled as 'Tissue' with a specific ROI. In this instance, ROI 2 is used for this example.
    c. Finally, a step to change the 'Tissue' label to the 'Clear' label should be included.

Step 2: Vessel Identification and Segmentation

Time: 3 hours

This step identifies all vessels within the tissue area and segments the tumor areas according to the distance away from the vessels. A threshold algorithm is employed to accurately label regions exhibiting the presence of CD31, an endothelial marker of choice in this project, thereby establishing the vessel area.

Note: Alternative markers for vessel labeling include collagen 4A, podoplanin, or smooth muscle actin alpha 2.

12. A new App should be created for this thresholding algorithm and one image class should be used: Class 5 for 'Vessel'.
    a. For 'Input' settings, a magnification of 5X and a field of view of 1024 × 1024 pixels in 'Advanced Sampling' settings are set.

b. In the 'Regions to Analyze' settings, the ROI to be chosen is the one that would surround areas identified as tissue in the previous App 'Tissue Segmentation'. For this instance, this would be ROI 2.
c. For 'Classification' settings, the 'Threshold' method is selected. The 'Vessel' label is assigned to areas displaying the CD31 marker, with pixel intensities ranging from 6000 to 30000. This range was visually chosen across multiple vessels within the tissue to reduce noise and most effectively identify all the vessels within the tissue but may be different for others (**Fig. 1C**).

Note: Depending on the intensity of the CD31 signal between projects and staining techniques, the pixel intensity range assigned may change. By visual assessments, testing multiple ranges throughout the tissue and between images is necessary for each project to allow the software to find the most appropriate range to identify potentially all vessels present.

13. Refinement of vessel identification and segmentation:
    d. A new step can be added within the 'Post-Processing' section to exclude vessel-labeled areas smaller than 4 μm^2. This filters out isolated pixels expressing CD31, ensuring the inclusion of regions with vessel-like morphology.
    e. Another step can be used to expand ('dilate') vessel-labeled areas by 15 pixels. This mirrors the visual signal given by the CD31 marker after thresholding, taking into consideration the vessel architecture and thickness from the lumen to the outer membrane.

Critical: Images can display areas of nonspecific binding indicated visually by high levels of secondary antibodies signal (Cy5 and TRITC in this setting). These areas may inadvertently receive a 'Vessel' label. To correct this and refine your labeling, the eraser tool for both the ROIs and labels can be used repetitively to remove these areas and refine vessel segmentation. The refinements can be done manually on one of the images, but the algorithm is being trained on multiple images at the same time loaded within the slide tray.

14. To refine the labeling of the perivascular areas within the tissue, a new step can be added within the 'Post-Processing' section.
    a. A new ROI can be assigned for each area of tissue chosen at a certain distance surrounding a vessel. This ROI would be different than the ROI assigned previously for all the tissue areas (ROI 2).
        i. In this instance, ROI 3 was assigned to the proper vessel area consisting of the 15 pixels around the CD31 marker (**Fig. 1D**).
        ii. Multiple ROIs can be then generated at a certain assigned distance to create segmented areas of the tumor microenvironment (e.g., ROI 4 and ROI 5, at 50 and 100 microns from the vessel (ROI 3), respectively).
        iii. Perivascular regions can be consolidated into a designated ROI to account for heterogeneity in vascular density (**Fig. 1E**).

Step 3: Nuclei Detection

Time: 3 hours

Following vessel labeling in Step 2, this step focuses on detecting nuclei via the detection of the DAPI staining to facilitate subsequent cell phenotyping (Cf. Step 5 and 6). This process uses a pre-trained deep-learning algorithm commercially available within the Visiopharm® Software.

15. Within the new App button, select the 'Nuclei Detection, AI (Fluorescence)' App. Three image classes to identify a cell are used: Class 1 for 'Background', Class 2 for 'Nucleus', and Class 3 for 'Border.' The 'Border' label is precisely drawn around objects in the image labeled as 'Nucleus', representing areas with strong circular DAPI signals, ensuring differentiation between cells. Everything else is assigned a 'Background' label.

Note: As the App is pre-trained, additional training areas like those mentioned in "Step 1: tissue segmentation" are not necessary. If desired, more training data can be added.

    a. For 'Input' settings, a magnification of 10X and a field of view of 512 × 512 pixels is set.
    b. In 'Advanced Sampling' settings, a 25µm counting frame is applied and the option to discard labels should be deselected.
    c. For 'Regions to Analyze', ROI 2 should be selected, including ROIs for tissue areas with labeled vessels identified in the previous App (Cf. Step 2: Vessel Segmentation).
    d. For 'Classification' settings, the Deep Learning method with the U-Net architecture is used.
    e. In the 'AI Architect' menu, the input size is adjusted to 512 × 512 pixels, the learning rate to 1.0E-5, and the minibatch size to 2. The algorithm is then trained using only the DAPI channel and limited pixel intensities to the empirically determined range of 6500-65000 to minimize noise.

16. Refining steps within the 'Post-Processing' section of the App:

Note: For better visualization of the subsequent steps the 'Border' and 'Background' labels can be shifted to the 'Clear' label.

    a. A 'Separate Objects' step can be added to refine each nucleus/cell identification separately from its surroundings.
        i. The object label should be the 'Nucleus'.
        ii. The separator label should be the 'Separator image class' created within the App in the 'Image Classes for Output' section.
        iii. The background label should be set to 'All except object'.
        iv. The object's diameter should be 15µm and the option to separate the object's surroundings should be a surrounding width of 3µm.

Note: it is preferable not to select the option that assumes that objects are 'elliptic'. Rather, it is better to select the 'F(3): Boundary' heatmap option with a value of 0.7.

    b. To take into consideration the cytoplasm and membrane of the cell, a step to dilate the 'Nucleus' label for 3 pixels is standardized and can be implemented. In this step

the option to avoid merging objects is disabled and a 'Separator' is chosen as the exclusion label.

c. A 'Change Surrounded' step, converting the 'Clear' label surrounded by 'Nucleus' to the 'Nucleus' label, with a 0.1 coverage. The image border should be included as well.
d. Next, the 'Separator' label should be changed to the 'Clear' label.
e. Adding a 'Change by Shape' step can be beneficial, taking into consideration all nuclei and cell shapes, especially within central nervous system tissue. The 'Area' label is used in this setting and the 'Nucleus' label is replaced by the 'Clear' label with a maximum chosen value of 10 μm^2.
f. Lastly, the counting frame to all labels can then be applied.

17. To collect relevant data from each cell identified, editing the 'Output Variable' section of the App is an important option to consider while avoiding the use of ROIs.
    a. A new variable named 'Total Nuclei' is created using the 'Classified image' as the input source. It is configured to calculate the 'Count' of the occurrences of the 'Nucleus' label per image. The output level is set to the image avoiding the use of ROI.
    b. The variable named 'Object' is created next also using the 'Classified image' as input. This should calculate the 'Object Info' for all labels on a per-object basis. The output level is set per object this time, without using the ROI.
        i. Create another variable 'Multiplex', using the 'Original image' as input. This variable calculates the 'Multiplexing' for all labels on a per-object basis. The output level is set per object, without using the ROI.
        ii. Create an 'Area' variable, again using the 'Classified image' as input. This variable calculates the area of the 'Nucleus' label on a per-object basis for the output level.
        iii. Finally, a variable named 'Distance' using the 'Classified image' as input can be used instead.
            1. The 'Object Distance' from the 'Nucleus' label to the 'Vessel' label (also referred to as 'Label 005') can be calculated on a per-object basis. A maximum distance limit of 1mm can be set and it can be increased according to the computer's processing power and any other time constraints.

Step 4: Data Extraction

Time: 12 Hours

Note: This step is optional.

In this step, the data already generated from the previous steps can be extracted and used in R Studio to normalize the data.

18. Within the 'Database', select all images for batch analysis.
    a. The Apps should be selected to run sequentially beginning with tissue segmentation followed by vessel identification and segmentation, and finally nuclei detection.
    b. Initiate batch analysis and wait until completion.

19. After completion, for each image, the data generated should be saved. Within the 'Database', select an image and click on 'Export selected results (.tsv)' to extract its data. This action will generate a TSV file for each image.

Note: All TSV files should be saved in the same folder before moving to the next step.

20. Download and install the latest version of R Studio.
    a. A new R script is created and saved in the folder containing the TSV files.
    b. In R Studio, the working directory should be set up to this folder.
    c. Load the tidy verse library and convert each TSV file into CSV files (cf. Code 1).

Code 1

    d. Load all the newly created CSV files into a list.
    e. Create a function to normalize columns with multiplex data, by dividing the signal intensity of the marker of interest by the signal intensity of its associated fluorophore (in this instance Cy5 or TRITC) (cf. Code 2).
        i. The resultant column shows the relative intensity of the marker against autofluorescence, controlling for artifacts like nonspecific antibody binding due to red blood cells.
        ii. Rows with 'NA' values, indicating cells outside the distance limit, are removed.
        iii. Only the columns showing the normalized signal intensity, area, and distance are preserved.
        iv. This function is applied to each CSV file, generating a new processed CSV file.

Code 2

    f. The folder corresponding to the working directory should now contain all generated CSV files. Each file contains data for all cells identified by the nuclei detection App contained within an image. When exported into Excel, each row in a file corresponds to a specific cell.

Note: The column names in the TSV files, used to calculate normalized multiplex values, can vary significantly. The code provided is just an example of how those column names may look. These names depend on the markers used in a specific project and how the markers are associated with the fluorophore channels. In this instance, the column name for each marker intensity value is divided by the column name for TRITC or Cy5 intensity values accordingly.

Critical: Selecting the appropriate fluorophore, TRITC, or Cy5, for association with your marker of interest is important for accurate normalization. This choice varies depending on the specifics of the staining protocol.

Step 5: Machine Learning Labeling

Time: 24 hours

Note: This step is optional.

It allows the user to train the Visiopharm machine learning algorithm while integrating the data into other modalities of analysis like R Studio and subsequently Spyder for more accurate results (cf. step 6).

21. Within the 'Database' in the Visiopharm software, all columns containing data from all images are removed, and ROIs and labels are deleted.

Note: This data can be preserved by creating a copy of the folder containing these images and performing machine-learning labeling in the duplicated folder.

22. To start, a marker of interest is selected to generate datasets for training the machine learning models.
    a. In the Visiopharm viewer, the channel corresponding to this marker is enabled along with other channels that could help identify cells expressing this marker of interest. These later include DAPI and known cell lineage markers. The 'Min/Max %' intensity settings are chosen accordingly for an effective visualization of these markers.
    b. Next, all areas where cells are positive for the chosen marker are identified in all images. ROI 2 is used to outline these areas.

23. Run the nuclei detection App (created in step 3) on these areas and save the data generated.

24. Create a new App for labeling. Four image classes are used in this setting: Class 1 for 'Background', Class 2 for 'Cell', Class 3 for '1', and Class 4 for '0'.
    a. In the 'Input' settings, a magnification of 10X and a field of view of 512 × 512 pixels are selected under 'Advanced Sampling' and default settings for all other options.
    b. In the 'Regions to Analyze' settings ROI 2 is chosen.
    c. For 'Classification' settings 'No Classification' is selected.

25. Edit the 'Output Variable' section of the App to collect relevant data from each cell.
    a. The variable 'Multiplex' is created using the 'Original image' as input. This variable calculates the 'Multiplexing' inputs for all labels on a per-object basis. The output level is set per object, without using the ROI.
    b. An 'Area' variable is created next, again using the 'Classified image' as input. This variable calculates the area of all labels on a per-object basis for the output level.
    c. The variable named ObjectInfo is created, also using the 'Classified image' as input. This should calculate the 'Object Info' for all labels on a per-object basis.

26. By accessing the Wheel for labels, objects classified as 'Cell', generated by the nuclei detection App (Step 23), can be tagged.

a. In 'more settings', by accessing the 'Draw' section, select 'Edit drawn objects(s)'.
b. Hover over cells and label them '1' if you assess them to be positive for the marker of interest, or '0' if negative.
c. After labeling all cells within the ROIs, the App is ready to be run.

Note: It is advisable to label all positive cells first, then change all other objects marked as 'Cell' to '0' to streamline the process. In addition, post-processing steps can be used to change 'Cell' objects based on intensity, assigning them the label '1' or '0' depending on their marker average intensity. Once the App is run, you can use the wheel to adjust the labels as necessary. When all cells have been labeled, disable these post-processing steps before running the App to extract the output variables.

27. The data generated by the Apps should be saved for each image.
    a. Within the 'Database' select the folder holding all the images and select 'Batch Export.'
    b. Choose to export the results as 'Numerical variables (.tsv).'
    c. Select 'Export a single file for all images' to combine all labeled data from every image into one file, this is important for subsequent training.

28. After clearing all labels and ROIs from every image, steps 22-27 can be repeated to phenotype each additional marker of interest.

Critical: All these TSV files generated should be stored in a separate folder, distinct from the group of TSVs generated in the previous step 4.

Note: The percent composition of cells labeled '1' to '0' should be relatively similar for each marker. It's important to avoid a disproportionate representation of cells labeled '0' while training the algorithm to make sure you increase the chances of it identifying '1' labeled cells. Under-sampling the '0' class might be necessary so that cases labeled '1' make up at least 10% of all labeled cells.

29. The next steps are conducted in R Studio to process each TSV file.
    a. Create a new R script and save it in the folder containing these TSV files.
    b. Set your working directory to this folder.
    c. Load the tidy verse library and put these TSV files in a list to be processed (cf. Code 3).

Code 3

d. A function can be created to first rename the 'ObjectInfo - LabelName' columns to 'LabelName' only for better clarity (cf. Code 4). Each of these columns, containing 0s or 1s, indicates whether a cell (represented by a single row) is positive or negative for a specific marker. It will then normalize columns similarly.

e. Rows with the 'Cell' value located in the 'LabelName' column are removed as these represent unlabeled cells. Then only columns with the rest of labels and normalized signal intensities are kept (cf. Code 4). This function is applied to each TSV file within the folder, generating new processed TSV files respectively.

Code 4

30. The working directory folder should now contain these new TSV files. Each file contains data for all labeled cells in a sample identified by the nuclei detection App. When exported into Excel, each row in a file corresponds to a specific cell.

Step 6: Random Forest Classifier Creation using Spyder

Time: 12 hours

Note: This step is optional.

It allows the use of the generated data files in step 4 and 5 and integrate them into Spyder to generate scores indicating the accuracy of the results and cell phenotypes.

31. Download and install the latest version of the Spyder IDE. Once installed, launch Spyder.
    a. Create a new script file.
        i. On your computer a new folder should be created, where the new script file is saved.
        ii. The processed TSV files generated in steps 4 and 5 are also added to this folder.
        iii. Within Spyder settings, set the working directory to this newly created folder.
    b. Load necessary libraries for machine learning functions (cf. Code 5).

Code 5

32. First, load into the script the processed TSV file, generated in step 5, of the marker of interest.
    a. Divide the dataset based on values in the 'LabelName' column, separating groups with labels '1' and '0' (cf. Code 6).

Code 6

b. This labeled dataset is divided into training and testing sets (cf. Code 7). The standard split is 80% for training and 20% for testing. However, this ratio can be adjusted according to the users' specific needs.

Code 7

c. A random forest classifier is created next with a max depth of 4 to prevent overfitting (cf. Code 8). You may add other settings to enhance the classifier's robustness.

Code 8

d. Predicting the positivity of the marker on the test set and calculating key evaluation metrics: precision, recall, F1 score, accuracy, and area under the curve (AUC) (cf. Code 9).

Code 9

Critical: The metrics help determine the effectiveness of the random forest classifier and the accuracy of the data generated. If these scores are not satisfactory, the protocol in step 5 for that specific marker can be repeated. This may involve adding new examples of positive cells '1' or under-sampling negatively labeled cells '0' due to potential class imbalances in the dataset. Then codes 5-9 should be run again all at the same time.

Note: 'Precision' quantifies the accuracy of positive prediction. 'Recall' measures the accuracy of how many of the actual positive cases the model has been able to predict. The 'F1 score' is the harmonic mean value of the 'Precision' and 'Recall'. 'Accuracy' measures the proportion of true positives and true negatives identified by the model. Precision, recall, and F1 scores are more reliable of an effective model than accuracy alone. 'AUC' (area under the curve) evaluates the model's ability to differentiate between classes. For AUC, any score above 0.5 indicates performance better than random guessing. Scores ranging from 0.9-1 are excellent, 0.8-0.89 are good, 0.7-0.79 are fair, and anything below is poor.

33. Load one of the processed TSV files from step 4, containing the unlabeled dataset of all cells (cf. Code 10). Ensure that the features in this unlabeled set match the features used in the training set for step 5 for that particular marker.

Code 10

a. Use the classifier to add a new column to the unlabeled dataset, while choosing its positivity for the marker of interest (cf. Code 11).

Code 11

34. Steps 32-33 should be repeated for every marker analyzed in each sample. Once all the markers have been processed, save the sample with the newly added columns as a new TSV file (Cf. Code 12).

Code 12

Note: Remember to modify column names in codes 10-11 to add a new positivity column specific to each marker.

35. Implement step 34 for each sample in your study and update the sample names in codes 10-11 accordingly. This ensures the creation of a separate TSV file for each sample.

Step 7: Cell Phenotype Counting using R Studio

Time: 2 hours

Note: This step can follow either step 6 or 8.

It allows the counting and graphical representation of all cell phenotypes and data generated in the previous steps.

36. Using R Studio, the TSV files generated in step 6 and the Excel datasheets generated in step 8 can be processed. Within R Studio:
    a. Create a new R script and place it in a newly created folder.
    b. Add either the TSV files or Excel datasheets generated in previous steps into this folder.
    c. Set your working directory to this folder.
    d. Load the tidy verse library and put these files in a list to be processed (cf. Code 13).

Code 13

37. To hold all phenotypes, different vectors are created for different analyses, and count matrices to count these phenotypes in each sample are initialized (cf. Code 14).

Code 14

38. A function should be created to process each data file (cf. Code 15). Especially for the TSV file generated in step 6, it allows the addition of phenotype columns based on the values in the positivity columns, and thus, multiple phenotype columns can be generated. It is essential to ensure that all phenotypes listed in the vectors are accounted for with conditional statements to be added to a phenotype column.
    Note: Make sure 'Other' is a phenotype in both the vectors and conditional statements.
    a. A phenotype name and conditional statement for 'Other' to categorize all other unlabeled cells must be included.
    b. The count matrices must be updated to reflect the number of each unique phenotype in its respective column for each sample.
    c. This function will create a new processed TSV file containing the distance, phenotype columns, and count matrices.

Code 15

Note: To assign a phenotype as a string value in its respective phenotype column based on the corresponding positivity values in the same row, use multiple conditionals. These conditionals should be based on whether the value from the random forest classifiers in Step 6, in the positivity column, is either 1 or 0. Different phenotype columns can be generated by different ways of looking at the data. As an example, one phenotype column may count the number of cells positive for one marker and another phenotype column may highlight selected phenotypes of interest based on multiple conditionals.

39. With the processed TSV files and count matrices, you can perform different analyses on your data.
    a. For easier subsequent analysis, all these files can be bound together for data analysis (cf. Code 16).

Code 16

b. As an example, the counts of CD11c+ cells away from a vessel in the Phenotype1 column are generated and binned every 50 micrometers, using the combined processed data (cf. Code 17).

Code 17

c. Another example is to compare how the composition of CD11c+ versus CD11c- cells changes based on distance away from a vessel (cf. Code 18).
   i. For each phenotype, a bar is created with two segments stacked on each other. Each segment's height represents the proportion of that phenotype in the corresponding distance bin, relative to the total count of cells in that bin.

Code 18

Step 8: Cell Phenotyping using Visiopharm Phenoplex Feature

Timing: 4 hours

Note: This step can replace steps 4-6, and be implemented directly after step 3, for faster data generation.

In this step, the 'Phenoplex' feature within the Visiopharm software is implemented. The phenotypes of interest are predetermined and encompass various categories including those positive for a single biomarker (e.g., TMEM119+, CD4+, or CD11c+), double positive (e.g., CD163+CD11c+, CD68CD163+, or CD205+CD11c+), and functional markers (e.g., CD8+LCK+, CD11c+pSTAT+, or CD205+CD11c+PDL1+).

Note: New updates and refinements are regularly added to this feature software that should be implemented and trained for using the tools offered by the Visiopharm® company. The analysis for this protocol was conducted using version 12.2022 of the software.

40. To enable phenotyping, the initial step involves the detection of nuclei (cf. step 3). This is done by using the pre-trained deep-learning algorithm in the Visiopharm software. For training purposes, three image classes are used: Background, Nuclei, and Border corresponding to classes 1, 2, and 3, respectively.
    a. The Border label is precisely drawn around objects labeled as Nuclei, representing areas with strong DAPI signals, thereby ensuring differentiation between cells. All other regions are assigned a Background label.

b. To apply the classifier effectively, optimal inputs include a 20X magnification to capture intricate details, a field of view measuring 512 × 512 pixels, and a counting frame of 10 µm within ROIs 10-255.
c. The classifier (U-Net architecture; input size = 512 × 512 pixels; learning rate = 1.0E-5; mini-batch size = 2; 100000 iterations) was exclusively trained on the DAPI channel with pixel intensities ranging from 5000 to 65000.
d. During post-processing, both the Border and Background labels are cleared.
e. Nuclei objects are subsequently separated using a Separator label corresponding to class 4 with a surroundings width of 2 µm assuming the smallest objects to be separated are 15 µm in diameter.
f. The resulting Nuclei label is then dilated by 3 pixels to accurately resemble cell boundaries.
g. The Separator label is subsequently cleared, and Nuclei objects with an area of 5 μm^2 or less are also cleared.

41. For each cell phenotype and depending on the markers used in the multiplex panel, image classes for output are generated and named based on the respective biomarkers they represent (in this instance 45 image classes for output were highlighted).
    h. Nuclei objects are classified into the appropriate phenotype class based on the mean pixel intensity of various biomarker signals.
    i. To ensure accuracy, pixel threshold values are selected to minimize false positive and false negative results, and these thresholds are applied consistently across all images.
    j. Notably, 100% of pixels are considered in calculating the mean intensities, as previous applications had already removed areas that would introduce noise.
    k. Following the application of the algorithm, an object info output variable is generated, associating each cell with an image class and the corresponding ROI.

42. For quantification, an Excel table was downloaded, and R programming was employed to generate counts and percentages of all cell phenotypes per perivascular and tumor area similar to step 7 (Cf. Codes 13-18).

Note: Pause Points and breaks can be taken at any step of this protocol for any length of time. Only make sure to save data at the end of each step.

Quantification and statistical analysis

Data derived from the images are analyzed in R to create tables detailing the counts of each of the 45 phenotypes in each region of the tissue. Before any comparisons are made, regions with less than 100 cells are removed from the dataset. 214 perivascular regions and 351 tumor regions are left after filtering across pilocytic astrocytoma samples (n=4). To compare immune cell counts, relative immune cell composition, and immune cell phenotype counts, a Mann-Whitney test is performed between perivascular and tumor regions, and p-values are calculated. The same test was done to compare positive and negative phenotypes within each region.

Expected outcomes

Immune cell populations that originate from the periphery including CD4+ T cells, CD8+ T cells, CD163+ macrophages, CD205+ myeloid blood dendritic cells, CD11c+ dendritic cells, and CD68+ monocytes are all preferentially enriched in the perivascular space relative to the tumor (**Fig. 2A, B**). Only TMEM119+ microglia, which are intrinsic to the CNS, demonstrate no difference in the frequency in the perivascular region relative to the tumor. Similar findings were also observed when various immune cell populations were characterized based on more than one marker (**Fig. 2C**). When immune activation markers such as Lck or immune suppression such as p-STAT3 were assessed, these were also found to be significantly enriched in the perivascular region (**Fig. 2A, 3**). The bioinformatics analysis extraction was sufficiently robust that it could discern that the Lck expression in the CD4+ T cells was localized to the perivascular region (**Fig. 3C**). Notably if a cell population reaches at or below 1% within a given region, it is difficult to ascertain statistical differences. Cumulatively, these data indicate that most immune reactivity is localized to the perivascular regions of pilocytic astrocytomas.

This spatial analysis provides a strategy that successfully identifies regions that quantify cells based on phenotyping as a function of location relative to the tumor vasculature. The subsequent detection of nuclei, performed through a pre-trained deep-learning algorithm, allows for the classification of cells into specific image classes. Phenotypes of interest, including single positive, double positive, and functional marker positive cells are determined based on the mean pixel intensity of relevant biomarker signals. Pixel threshold values are selected to minimize false positive and false negative outcomes. Subsequent data analysis involved the generation of object-level information and the utilization of R programming to extract information such as phenotype counts, counts within specific regions, and the percentage composition of phenotypes.

Limitations

CD31, an endothelial cell marker, may not show a homogeneous consistent vessel border because it depends on the cell orientation. In addition, Pixels representing a long vessel are sometimes smaller in count than the number of pixels representing a DAPI-positive cell. This may not appear to be the case in visualized images depending on the thresholding. This made it difficult to apply the deep-learning algorithm to identify these vessels. Ultimately, other vessel markers may be more useful for training on vessel structures.

Rather than looking at positive or negative counts of a phenotype, pixel intensities could be used instead. This makes it possible to create a gradient representation of a marker to create a line graph showing the pixel intensity of a marker as a function of distance. This removes subjective decisions in deciding what is a “perivascular area” or “tumor area” or what threshold defines a particular cell lineage.

Non-specific binding for some targets such as PD-1 and PD-L1 will require a higher threshold. Adding another algorithm to remove these regions may increase accuracy.

Troubleshooting

Problem 1:

Based on the processing power of the computer available and the number of images within a project, the accuracy of your training may vary.

Potential solutions:

- The more images that are loaded into the slide tray and the App is trained on, the better the outcome.
- If the computer capacity is limited, load different amounts of images onto the slide tray. Train on one set of images for a portion of time and then train on the other sets of images for equal amounts of time.

Problem 2

Sometimes the 'loss function' won't converge based on the nature of the parameters.

Potential solution:

- Lower the learning rate and increase the mini-batch size. However, training time will be longer.

Problem 3:

Different staining techniques and different tissue types can generate different signal intensities for each of the markers.

Potential solution:

- For each tissue type, visually select the best threshold range appropriate to the marker's architecture.
    - To determine the best low and high intensity, start with a wider range interval and narrow it down until you have the best visualization of your tissue architecture (e.g., our ultimate range for CD31 was selected as 6000-30000, but we first started with the empirically determined range of 2000-50000 and narrowed it down accordingly).

Problem 4:

The DAPI channel signal may vary based on cellular density.

Potential solution:

- The input channel for DAPI can be adjusted based on the strength and range of your DAPI signal for the best detection and segmentation of the cells.

Resource availability

Lead contact

Amy.heimberger@northwestern.edu.

Materials availability
This study did not generate new unique reagents.

Data and code availability
The published article includes all [datasets/code] generated or analyzed during this study.

Acknowledgments

This study was funded by the NIH grants CA120813, NS120547, P30 CA016672, R50 CA243707 and P50CA221747. The Lunaphore COMET™ multiplex immunofluorescence system was enabled by a gracious gift from the Stephen M. Coffman Trust to the Northwestern Medicine Malnati Brain Tumor Institute.

Author contributions

HN performed the multiplex analysis and HN and SP devised the spatial bioinformatic analysis. ABH oversaw all aspects of the protocol and provided financial support. All authors contributed to writing the protocol.

Declaration of interests

ABH serves on the advisory board of Caris Life Sciences.

Figure Legends

Figure 1: Representative image of a pilocytic astrocytoma showing key steps in the bioinformatic analysis. A) The COMET™ platform from Lunaphore Technologies™ performs a fully automated protocol of seqIF™, where at each cycle 2 antibodies' staining, imaging, and elution are performed, with the ability to detect 40 biomarkers in a single experiment. The images are then transferred into Visopharm software for data analysis. B) Tissue Segmentation representative image showing nuclear DAPI (dark blue), glioma GFAP (orange), and vessel CD31 (cyan blue). Signal capture in the Cy5 (yellow) and TRITC (green) channels is initially turned on to remove the non-specific signals present within the image from subsequent analyses. Regions of the tissue that do not have either GFAP, CD31, or DAPI signal are labeled as background and removed from the analysis. C) During Vessel Identification, the CD31 signal is optimized through the selection of the threshold. D) During Vessel Segmentation, the distance from the vessel is then set at 55 µm (red) and then a further 55 µm (yellow) to define the tumor. E) Area of confluence are merged into one ROI. All images are 1.25x magnification.

Figure 2: A) Representative image of a pilocytic astrocytoma showing the tumor segmentation of the perivascular region (dashed blue line) and the tumor (dashed purple line). Cell populations are then identified based on the presence of DAPI+ nuclei and are assigned a phenotype based on the marker defined in the figure (10 x magnification). B) Quantification of immune cell populations based on the expression of a single marker as a function of regions (perivascular versus tumor). % of cells is calculated from the total number of DAPI+ cells. C) Quantification of immune cell populations based

on the expression of two markers as a function of regions (perivascular versus tumor). % of cells is calculated from the total number of DAPI+ cells. ns p-value >0.05; * p-value <0.05; ** p-value <0.01; *** value-value <0.001; **** p-value <0.0001

Figure 3: A) Representative multiplex image showing the nuclear staining of immune suppressive p-STAT3+ cells (purple) as a function of tumor region and relative to total DAPI+ cell count (grey) (10x magnification). B) Representative multiplex image analyzing LCK expression, as a marker for T cell activation, as a function of immune cell lineage, and as a function of regions (10x magnification). C) Quantification of immune effector functions based on regional location of perivascular versus tumor. % of cells is calculated from the total number of total DAPI+ cells. ns p-value >0.05; * p-value <0.05; ** p-value <0.01; *** p-value <0.001; **** p-value <0.0001.

References

1. Najem H, Ott M, Kassab C, et al. Central nervous system immune interactome is a function of cancer lineage, tumor microenvironment, and STAT3 expression. *JCI Insight*. May 9 2022;7(9)doi:10.1172/jci.insight.157612
2. Gabrusiewicz K, Rodriguez B, Wei J, et al. Glioblastoma-infiltrated innate immune cells resemble M0 macrophage phenotype. *JCI Insight*. 2016;1(2)doi:10.1172/jci.insight.85841
3. Weathers SP, Penas-Prado M, Pei BL, et al. Glioblastoma-mediated Immune Dysfunction Limits CMV-specific T Cells and Therapeutic Responses: Results from a Phase I/II Trial. *Clin Cancer Res*. Jul 15 2020;26(14):3565-3577. doi:10.1158/1078-0432.Ccr-20-0176
4. Sonabend AM, Gould A, Amidei C, et al. Repeated blood-brain barrier opening with an implantable ultrasound device for delivery of albumin-bound paclitaxel in patients with recurrent glioblastoma: a phase 1 trial. *Lancet Oncol*. May 2023;24(5):509-522. doi:10.1016/s1470-2045(23)00112-2
5. Sabbagh A, Beccaria K, Ling X, et al. Opening of the Blood-Brain Barrier Using Low-Intensity Pulsed Ultrasound Enhances Responses to Immunotherapy in Preclinical Glioma Models. *Clin Cancer Res*. Aug 1 2021;27(15):4325-4337. doi:10.1158/1078-0432.Ccr-20-3760
6. Vogelbaum MA, Li G, Heimberger AB, et al. A Window of Opportunity to Overcome Therapeutic Failure in Neuro-Oncology. *Am Soc Clin Oncol Educ Book*. Apr 2022;42:1-8. doi:10.1200/edbk_349175
7. de Groot J, Penas-Prado M, Alfaro-Munoz K, et al. Window-of-opportunity clinical trial of pembrolizumab in patients with recurrent glioblastoma reveals predominance of immune-suppressive macrophages. *Neuro Oncol*. Apr 15 2020;22(4):539-549. doi:10.1093/neuonc/noz185
8. Rivest F, Eroglu D, Pelz B, et al. Fully automated sequential immunofluorescence (seqIF) for hyperplex spatial proteomics. *Sci Rep*. Oct 9 2023;13(1):16994. doi:10.1038/s41598-023-43435-w