Journal Title Here, 2023, pp. 1–8

doi: DOI HERE
Advance Access Publication Date: Day Month Year
Paper

PAPER

Tribus: Semi-automated discovery of cell identities and phenotypes from multiplexed imaging and proteomic data

Ziqi Kang,[1] Angela Szabo,[1] Teodora Farago,[1] Fernando Perez-Villatoro,[1, 6] Inga-Maria Launonen,[1] Ella Anttila,[1] Kevin Elias,[2] Julia Casado,[1] Peter Sorger[3] and Anniina Färkkilä[1, 4, 5, 6, *]

[1]Research Program in Systems Oncology, University of Helsinki, Finland, [2]Division of Gynecologic Oncology, Brigham and Women's Hospital, Harvard Medical School, USA, [3]Laboratory of Systems Pharmacology, Harvard Medical School, USA, [4]Institute for Molecular Medicine Finland, Helsinki Institute of Life Sciences, University of Helsinki, Finland, [5]Department of Obstetrics and Gynecology, Department of Oncology Clinical Trials Unit, Helsinki University Hospital, Finland and [6]iCAN Digital Precision Cancer Medicine Flagship, Finland
*Corresponding author. anniina.farkkila@helsinki.fi

Abstract

Multiplexed imaging at single-cell resolution is becoming widely used to decipher the role of the cellular microenvironment in cancer and other complex diseases. To identify spatial patterns of single cells on a tissue, accurate cell-type phenotyping is a crucial step. This step is challenging due to (i) fluorescence noise and batch effects, (ii) segmentation artifacts, (iii) laborious annotation of ground truth, and (iv) difficulty in reproducing human-biased thresholding. Here we present Tribus, an interactive, knowledge-based classifier that avoids hard-set thresholds and manual labeling, is robust to noise, and takes fewer iterations from the user than current methods of labeling. Tribus has built-in visualization functions to gain insight into the input data and to evaluate the results. The Napari plug-in provides a user-friendly way to visualize the results and perform quality control. In this study, we validate the accuracy of Tribus by comparing its results to labels in public benchmarking datasets where manual cell type annotations are supported by the pathology community. We applied Tribus on a cyclic immunofluorescence (CyCIF) dataset, consisting of five matched ovarian cancer samples collected before and after neoadjuvant chemotherapy. Accurate cell-type phenotyping enabled a high-resolution analysis of cellular phenotypes, their spatial patterns, and their temporal dynamics during platinum-taxane chemotherapy. Tribus is provided as an easily embeddable open-source package and enables accurate phenotyping of single cells to facilitate biological discovery from highly multiplexed images.

Key words: Cell-type phenotyping, Clustering, Cyclic Immunofluorescence (CyCIF), Co-detection by indexing (CODEX)

Introduction

Multiplexed imaging techniques at single-cell resolution level, such as cyclic immunofluorescence (CyCIF) [1, 2], Co-detection by indexing (CODEX) [3], and multiplexed ion beam imaging by time of flight (MIBI-TOF) [4], have developed significant advantages in understanding tissue architecture. Such techniques allow researchers to measure dozens of proteins at single-cell resolution while retaining the spatial information of origin in the tissue section. Those experiment workflows generate datasets that include multiple features and provide novel insights to understand the diversity of cellular phenotypes and behaviors in tissues [5].

To obtain such research, cell type annotation becomes one of the most essential tasks and a bottleneck to gaining full potential from Multiplexed imaging technologies. The main obstacles to cell-type phenotyping on multiplexed imaging datasets are the unexpected combinations of markers, signal noise and the reproducibility limit causing by imperfect cell segmentation, pre-processing artifacts, or tissue processing artifacts like autofluorescence. It is hard to apply popular scRNA-seq cell-type callers [6] on multiplexed imaging datasets because they are not designed to overcome segmentation noise (which does not exist in scRNA-seq data). The order of magnitude difference in the number of features is also an indispensable problem. Several markers can dominate the clustering process in multiplexed imaging techniques, meanwhile, thousands of genes are provided to scRNA-seq callers to overcome this problem.

Gating [7] is a classical cell type annotation method, which requires manually identifying cell types based on the pattern of features visualized by histograms and bivariate plots. However,

© The Author 2023. Published by Oxford University Press. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com

along with the increasing size of the marker panel, manual gating becomes time-consuming and sensitive to the batch effect [8], which is no longer feasible. Unsupervised clustering tools such as PhenoGraph [9], Leiden [10] and DBscan [11] were applied to perform automatic data exploration [12], which indeed allows fast cell population discovery. However, manual checking is still required on the obtained clusters to assign meaningful cell types. To obtain deeper profiling of cell types, it is usually necessary to perform over-clustering and then merge clusters, which is both computationally and time-consuming. An alternative approach is to use supervised or semi-supervised machine learning methods [13], which were developed to perform accurate auto cell-type classification, but such methods require pre-training on manually gated datasets according to various experimental designs, thus require extra human work and lack extendable.

In this paper, we introduce the design of a novel cell-type caller named Tribus. Tribus could perform reproducible automatic cell-type annotation in different depths without previous gating. Tribus incorporates the widely used Self Organizing Map (SOM) [14] unsupervised neural network model and a scoring function for assigning cell types according to prior biological marker panel knowledge. A Napari plugin was integrated into Tribus, allowing convenient user interaction and result quality control. We validate the accuracy of Tribus on four public multiplexed imaging datasets generated by mass cytometry and CODEX methods. We compare the performance of Tribus with two similar knowledge-based approaches, ACDC [15] and Astir [16]. We also compared the performance of Tribus with a typical clustering method Phenograph as a baseline method. We show the usage of Tribus in the analysis of in-house tCyCIF datasets. Overall, Tribus could represent a new potential framework of automated cell type calling for multiplexed imaging techniques.

Methods

The structure of Tribus

To run Tribus, a marker expression table and a logic table are required as the input. The logic table could be defined as a data matrix $L(c_i, m_j)$ containing values $[-1, 0, 1]$, where c_i is cell type ith and m_j is marker jth. If a marker should be present in a specific cell type, the value in the logic table has to be assigned by 1. If a marker is supposed to be absent in a specific cell type, then the value in the logic table should be assigned -1. For neutral or unknown markers, 0 should be assigned.

Tribus consists of two main parts. First, it uses unsupervised learning to cluster the input data, then it assigns each cluster to one cell type based on the previously described cell type description table. The cell type assignment is performed hierarchically, which means Tribus first performs the assignment of higher-level cell types, then performs the cell type calling on each higher-level cell type, to create more precise categories.

Unsupervised learning in Tribus

Tribus uses Self-Organising Map (SOM) for data clustering, based on the MiniSOM package [17]. Quantization error (Q) was to evaluate the algorithm performance, calculated by determining the average distance of the sample vectors to the cluster centroids.

$$Q = \frac{1}{m}\sum_{q=1}^{m} \|\vec{\boldsymbol{x}}^{\boldsymbol{q}} - \vec{\boldsymbol{w}}_{\boldsymbol{i}}{}^{\boldsymbol{q}}\|^2 \tag{1}$$

Comparison of quantization error needs to be under the same grid size. [18]. To determine the optimal grid size we use the approach by Vesanto [19]: $G = 5 * sqrt(N)$, where N is the size of the input data. Optimizing sigma (spread of the neighborhood function) and learning rate by minimizing the quantization error. The Hyperparameter tuning was implemented based on package hyperopt [20], where the objective function is to minimize quantization error, and the search space for parameters includes sigma and learning rate from 0.001 to 5.

Cell type assignment by scoring functions

After clustering a node matrix $N(n_k, v_j)$ was generated, where v_j is calculated by the median expression of marker m_j in the cluster n_k. We designed a scoring function based on the idea of squared error implemented in the function QueryStarPlot of the package FlowSOM [21]. The function calculates the score s_i of certain cell type i for each node k:

$$f(n_k, s_i) = \frac{1}{k}\sum_{j}[1 - (v_{ij} - \hat{v_{ij}})^{2'}] \tag{2}$$

where,

$$\hat{v_{ij}} = \begin{cases} P_{99}(v_{ij}), & \text{if } L(c_i, m_j) = 1 \\ min(v_{ij}), & \text{if } L(c_i, m_j) = -1 \end{cases} \tag{3}$$

Instead of using the maximum value in 3, the 99 percentile was chosen to make Tribus more robust to outliers.

Assigning node n_k to cell type c_i is defined by:

$$c_i = argmax_{s_i} f(n_k, s_i) \tag{4}$$

Notice if the maximum score of a cluster is smaller than a certain threshold set by users, the cluster will be labeled as the "other" cell type. Similarly, if the difference between the maximum and second maximum score is smaller than a certain threshold, the cluster will be labeled as "undefined".

Napari plug-in

To efficiently assess and explore the cell type labeling results, we developed a custom plugin integrated with the Napari [22] framework. It allows the user to run Tribus over one sample at a time and display the results simultaneously, or load previously saved data. We incorporated the zarr format to overcome the computation limitations associated with large datasets. Key functionalities of the Napari plugin include:

(i) Cell-type mask visualization: The plugin sorts and displays the different cell-type labels as separate layers using visually distinct colors, allowing the user to overlay them with the imaging data for quality control.

(ii) Probability score visualization: The cell masks are represented as a color gradient of the probability score assigned by the algorithm. This allows the user to identify and check ambiguous cells.

(iii) Marker intensity visualization: The median expression levels of the desired markers are represented through gradient shading on the segmentation mask. It allows the user to visually assess the results and look for potential biases.

Benchmarking Datasets

We chose four public mass cytometry and multiplexed imaging datasets with ground truth labels to test the performance of Tribus.

AML

This dataset contains single-cell proteomic profiles of human bone marrow from both Acute myeloid leukemia (AML) patients and healthy adult donors [9]. AML dataset was collected by mass cytometry with a panel of 32 markers targeting 14 manually gated cell classes. When performing benchmarking, cell type "NotDebrisSinglets" was removed from the analysis (N = 104 184 cells). The cell typing logic table was based on the cell type-marker table provided in the ACDC research.

BMMC

BMMC dataset is from bone marrow mononuclear cells [23] profiled by mass cytometry. The publically available dataset contains 13 markers and 24 manually gated cell populations. Following the ACDC research, erythroblast, megakaryocyte platelet and myelocyte were merged as an unknown population and removed from the analysis. All "NotGated" cells were also removed (N = 61725 cells). The cell typing logic table was also based on the cell type-marker table provided in the ACDC research.

DCIS

Ductal carcinoma in situ (DCIS) dataset is collected by multiplexed ion beam imaging by time-of-flight (MIBI-TOF) with a 37-plex antibody staining panel. DCIS consists of 69672 cells, 23 cell types from tissues containing 79 clinically annotated surgical resections [24], including normal breast tissue (N = 9, reduction mammoplasty), primary DCIS (N = 58), and IBC (N = 12). The cell typing logic table for DCIS was also generated based on the panel design in the published study.

HuBMAP

HuBMAP is a published Co-detection by indexing (CODEX) imaging dataset collected from 9 donors with a total of 54 antibodies targeting 25 cell types [25]. HuBMAP dataset includes eight regions (in order of trajectory from the stomach): duodenum, proximal jejunum, mid-jejunum and the ileum from the small bowel, and ascending, transverse, descending and sigmoid regions of the large intestine (colon). In this research, we use the manually labeled donor 004 from the HuBMAP dataset (N = 110635 for colon regions, N = 137652 for small bowel regions). Cell typing logic tables were generated according to the panel design information provided in published studies.

In-house Dataset

To further validate Tribus's performance, we also applied Tribus to an in-house high-grade serous ovarian cancer dataset generated by tissue-based cyclic immunofluorescence (tCyCIF). The use of human tissue samples was approved by the Mass General Brigham Institutional Review Board. Informed consent was waived due to the use of archival samples and anonymization of the material.Two whole-slide images of tumor sections after neoadjuvant chemotherapy (NACT) and one treatment-naive biopsy sample were selected based on the 36-marker tCyCIF staining. The cell segmentation was carried out with StarDist and the mean fluorescence intensities were quantified by an in-house python script. The Hoechst cycles were used to filter out detached tissue areas and cells that moved between the cycles. Staining artifacts were removed via manual cropping. The quantified data was further pre-processed: log transformation, z-score normalization, 99.9 percentile outlier removal and finally co-factor 5 arcsinh transformation [**?**] were applied before phenotype assignment. A total of 976082 cells were annotated with a cell phenotyping logic table based on the panel design. The same logic table was applied to one of the NACT samples and the biopsy, while CD20 was excluded from the second NACT sample due to insufficient staining quality.

Methods for comparison and evaluation metrics

We compared the performance of Tribus with two other similar prior knowledge-based unsupervised cell-type calling tools: ACDC [15] and Astir [16].

We evaluated the overall clustering performance against all cell type classifications by comparing Rand Index [26], Accuracy, Precision score (weighted) and F1 score (weighted) [27]. All the above metrics were calculated by functions provided in Scikit-learn [28].

Rand Index was chosen to measure the similarity between two data clusterings and was defined by $RI = (a+b)/(\binom{N}{2})$. Accuracy was defined by $Accuracy = (TN+TP)/(TN+FN+TP+FP)$. F1-score can reflect the harmonic mean of precision (purity) and recall (yield),

$$F_i = 2 * \frac{(P_{ii}) * (R_{ii})}{P_{ii} + R_{ii}}, \tag{5}$$

$$P_{ij} = \frac{C_{ij}}{\sum_k C_{ik}}, R_{ij} = \frac{C_{ij}}{\sum_k C_{kj}}, \tag{6}$$

where C_{ij} is the number of events classified as population i that belongs to the manually annotated population j.

Results

Overview of Tribus

Tribus is a hierarchy framework for cell type assignment in multiplexed image datasets using user-defined prior knowledge (Fig 1).

The preprocessing of datasets for Tribus is performed using log and z-scale. The datasets should then be normalized via an arcsinh transformation with a co-factor of 5 [**?**](optional) and winsorized to [0%, 99.9%]. If the number of cells in the given dataset is larger than the threshold set by the user, Tribus will first perform unsupervised clustering using the self-organizing map on the input dataset, with tunable sigma and learning rate. Otherwise, Tribus will directly calculate the scoring function for each cell type on a single-cell level.

The labeling process of cell-type assignment using Tribus considers the ambiguous results, which means Tribus does not force naming. If the scores of a cluster are lower than a certain threshold for all cell types, it will be marked as 'unknown'. Similarly, if the cell type scores are closer than a threshold, the cluster will be marked as "undefined" to avoid incorrect assignments and facilitate the identification of new cell types.

Tribus recovers fine-grained cell types as accurately as human experts

To test whether Tribus can recover canonical cellular populations, we applied Tribus to the benchmarking dataset DCIS, which contains 23 cell types from a cohort of 79 patients. Tribus successfully identified every population highlighted in the study (Supplementary Figure X) and the median marker intensity of each cellular population showed virtually similar to human labels (Fig 2A).

To compare how cell type varied between manual labels and Tribus labels, we mapped cell masks back to original tiff images with a nine-color overlay of major makers and observed few differences (Fig 2B). The one significant exception was a population of MACS (macrophages) that showed a lower median intensity of CD14 in Tribus than with manually gating (Fig 2A). Considering CD14 was marked positive or negative for MACS identification in DCIS research, we didn't constrain CD14 to be positive for MACS cells in the logic table. We then validate the population of MACS by visualizing the original images with all six marker panels related. We observed that more cells were assigned to MACS in manual labels, but the corresponding MACS markers were absent in some cells (Fig 2B).

Compared among manual labels and Tribus labels, Supervised UMAP visualization of all cell types from DCIS datasets shows that Tribus can accurately annotate the majority of cells (Fig 2C). Using the manual labels as ground truth, Tribus achieved precision scores between 0.7 and 0.8, average recalls around 0.6, and F1 scores between 0.6 and 0.7 across all cell types (Supplementary Figure X).

Tribus can target ambiguous populations and discover phenotypically new subtypes

One challenge for prior knowledge-based cell-type calling methods is to discover ambiguous categories that are hard to predefined in the logic table. For example, groups of unknown cells that have low intensity among all markers. DCIS dataset is a good example of exploring the ambiguous population since it includes "IMMUNEOTHER" and "OTHER" cell types in manual gating. Among them, the "OTHER" cell type shows low marker intensity in all markers while "IMMUNEOTHER" is only highly expressed in CD44 and CD45.

We verified Tribus could successfully target ambiguous populations by setting the decision thresholds (Supplementary Figure X) in the scoring function. In the UMAP visualization, we could observe the same "IMMUNEOTHER" and "OTHER" cell clusters with the same local structures being kept (Fig 2C). Marker expression heatmap shows Tribus targeted ambiguous cell types have the same marker expression profile compared with manual gating cell types (Fig 2D).

Meanwhile, we also discover new clusters from UMAP plots highlighted with orange boxes (Fig 2C), including cell type of undefined tumor, fibroblast, and T-cell. Novel marker expression combinations were observed from the median marker expression heatmap (Fig 2D). We found CD36+CD31+ fibroblast and HER2- tumor luminal subtypes. The undefined T-cells have phenotypically higher marker expression compared with typical CD4T cells. We suggest that Tribus could be considered as a starting point to uncover novel cell states via further unsupervised clustering.

Tribus outperformed other similar methods

We benchmarked the performance of Tribus using several orthogonal approaches. We chose two prior knowledge-based unsupervised tools: ACDC and Astir.

The benchmarking experiments show that Tribus outperforms the other methods comparing efficiency and accuracy (Fig 3A, B). In the AML and BMMC datasets, Tribus reached a similar accuracy with ACDC but with a shorter running time in one order of magnitude. When it comes to high-complexity datasets such as DCIS or HubMAP (Fig 3C), Tribus maintains relatively high accuracy while still low time-consuming.

We also explored how the structure of the logic table affects the accuracy using the AML dataset as an example. We apply Tribus on the AML dataset with (1) a logic table with only one global level and (2) a logic table that includes major immune cell types at the global level, and further divides phenotypes (e.g. CD16- NK cells and CD16+ NK cells) in the second level. For each logic table, we repeat the Tribus analysis for five times, calculate average metrics, and visualize the result. When adjusting the logic table on the AML dataset (Supplementary Table X), using a hierarchical logic table can help improve the accuracy and increase the F1 score by 0.1 (Fig 3D).

Tribus yields rapid and accurate cell type calling in a large whole slide imaging dataset

We applied Tribus on the in-house dataset guided by a four-level logic table. The hierarchical logic table structure can help to improve phenotype assignment in case of unexpected staining patterns. We demonstrate this through the labeling of B cells: We observed that the CD20 antibody also stained the nuclei of the cancer cells which resulted in abundant false positive B cells when the phenotypes were assigned based on a one-level table. Creating a hierarchical order of phenotypes and separating cancer cells before assigning the immune sub-types improved our results significantly.

Tribus successfully identified major cell phenotypes as well as the subtypes which we were able to characterize based on the available marker panel. Tumor proliferation was characterized by Ki67 and myeloid cells were categorized into macrophage type 1 (M1) and type 2 (M2) by CD163 expression levels. We show that Tribus can successfully annotate dense tissue areas and create accurate phenotype labels for further analysis.

Discussion

Cell type calling and feature extraction are critical steps in high-dimensional image analysis. The growing complexity and increasing numbers of panels in high-dimensional data necessitate the development of reproducible, more automated cell-type calling approaches. In this study, we developed Tribus, a semi-supervised cell-type calling analysis framework for multiplexed imaging datasets. Tribus shows striking advantages in efficiency, accuracy, user-friendliness, and reproducibility.

Tribus was generated as an automatic cell-type caller for high-dimensional data by incorporating biological knowledge from panel design into the analysis. This was done through carefully designed scoring functions based on marker expression per grid generated by unsupervised machine learning, aiming to minimize bias. When the number of cells included in clusters is lower than the threshold set by the user, Tribus will skip generating the clusters and calculate the scores at the single-cell level. Such a flexible scoring function calculation strategy enables the discovery of rare cell types. Tribus was integrated

with Napari and a plugin was provided for importing the cell type identification result in one click, which significantly increased the simplicity of interactive quality control. Such integration enables convenient operations for users who are not familiar with programming and the interpretation of biological contexts. Tribus was specifically crafted for high-plex imaging and proteomic technologies, and it seamlessly operates across various imaging platforms.

However, Tribus is not without limitations. Tribus performance is highly related to the quality of the input dataset and the prior knowledge of expected cell types in the user-defined initial cell type logic table. To assign a uniform logic table the samples should have even staining patterns both within and across slides. Uneven staining patterns and antibodies with sub-optimal signal-to-noise ratio can heavily affect Tribus's results. In sub-optimal cases, the user can try to create a hierarchical logic table in such a way that major cell phenotypes with clear marker signals are separated first, on different levels. This can result in more accurate labels if distinct areas of the image are affected. Tribus cannot find cell types that do not exist in the input logic table, but it can return undefined cell types for further exploration with unsupervised clustering methods.

Moreover, cell-type annotation from bioimages has inherent challenges due to imperfect cell segmentation and collateral spillover. Expanding nuclei masks by a few pixels can enhance cytoplasmic marker visibility, as it allows to capture the signal better when cells express these markers. However, in dense tissue areas this might increase spillover which highlights the importance of the hierarchically structured logic table: Carefully designing the table can aid cell separation as only a subset of cells are taken into account on the lower levels during the phenotype assignment. Typically, mild spillover affects only part of a cell, while a true signal yields a more uniform expression pattern and higher mean fluorescence intensity. Tribus is designed to include both positive and negative components for expected marker expression. We could aid correct labeling as well in case of a mild spillover by a less strict logic table, where markers with false positive expressions are set as neutral.

Overall, we propose Tribus as a fast, accurate, and user-friendly cell-type identification method that could be integrated into multiplexed image analysis frameworks.

Key points

- A semi-supervised cell-type identification computational tool for imaging data was constructed, enabling fast and accurate phenotyping of single cells.
- The tool was integrated with the Napari plug-in to accomplish user-friendly assessing result quality before downstream analysis.
- Compared with classic and other similar methods, our method performed best on all benchmarking datasets.
- The method is capable of capturing the spatial patterns of biological functions in multiple and complex cell types.

Competing interests

No competing interest is declared.

Author contributions statement

Z.K, A.S, T.F, J.C developed the algorithm and the software. F.P participated in the discussion and provided feedback during the algorithm development. I-M.L, F.P, E.A, K.E generated the in-house dataset. Z.K, analyzed the public benchmarking datasets. A.S, analyzed the in-house dataset. Z.K and A.S wrote the manuscript. P.S provided resources. A.F conceived and supervised the work, provided resources and wrote the manuscript. All authors have reviewed and approved the manuscript.

Acknowledgments

This study was co-funded by the European Union (ERC, SPACE 101076096). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them. This study was also funded by the Sigrid Jusélius Foundation, Research Council of Finland (grant numbers 1339805, 350396), and Cancer Foundation Finland (A.F, F.P). The University of Helsinki Research Foundation (F.P.), Ida Montinin Säätiö (F.P.), and Biomedicum Helsinki Foundation (F.P.) also provide the funding. We also wish to thank the IT Center for Science (CSC) for the computational resources.

Data and code availability

Tribus package is available at https://github.com/farkkilab/tribus.

The results of this study can be replicated with the dataset and the jupyter notebook publicly available on Synapse: https://www.synapse.org/!Synapse:syn53754523/files/.

References

1. Jia-Ren Lin, Benjamin Izar, Shu Wang, Clarence Yapp, Shaolin Mei, Parin M Shah, Sandro Santagata, and Peter K Sorger. Highly multiplexed immunofluorescence imaging of human tissues and tumors using t-CyCIF and conventional optical microscopes. *eLife*, 7:e31657, July 2018. Publisher: eLife Sciences Publications, Ltd.
2. Anniina Färkkilä, Doga C. Gulhan, Julia Casado, Connor A. Jacobson, Huy Nguyen, Bose Kochupurakkal, Zoltan Maliga, Clarence Yapp, Yu-An Chen, Denis Schapiro, Yinghui Zhou, Julie R. Graham, Bruce J. Dezube, Pamela Munster, Sandro Santagata, Elizabeth Garcia, Scott Rodig, Ana Lako, Dipanjan Chowdhury, Geoffrey I. Shapiro, Ursula A. Matulonis, Peter J. Park, Sampsa Hautaniemi, Peter K. Sorger, Elizabeth M. Swisher, Alan D. D'Andrea, and Panagiotis A. Konstantinopoulos. Immunogenomic profiling determines responses to combined PARP and PD-1 inhibition in ovarian cancer. *Nature Communications*, 11(1):1459, March 2020. Publisher: Nature Publishing Group.
3. Sarah Black, Darci Phillips, John W. Hickey, Julia Kennedy-Darling, Vishal G. Venkataraaman, Nikolay Samusik, Yury Goltsev, Christian M. Schürch, and Garry P. Nolan. CODEX multiplexed tissue imaging with DNA-conjugated antibodies. *Nature Protocols*, 16(8):3802–3835, August 2021. Number: 8 Publisher: Nature Publishing Group.

4. Leeat Keren, Marc Bosse, Steve Thompson, Tyler Risom, Kausalia Vijayaragavan, Erin McCaffrey, Diana Marquez, Roshan Angoshtari, Noah F. Greenwald, Harris Fienberg, Jennifer Wang, Neeraja Kambham, David Kirkwood, Garry Nolan, Thomas J. Montine, Stephen J. Galli, Robert West, Sean C. Bendall, and Michael Angelo. MIBI-TOF: A multiplexed imaging platform relates cellular phenotypes and tissue structure. *Science Advances*, 5(10):eaax5851, October 2019. Publisher: American Association for the Advancement of Science.
5. Matthew H. Spitzer and Garry P. Nolan. Mass Cytometry: Single Cells, Many Features. *Cell*, 165(4):780–791, May 2016.
6. Bingbing Xie, Qin Jiang, Antonio Mora, and Xuri Li. Automatic cell type identification methods for single-cell RNA sequencing. *Computational and Structural Biotechnology Journal*, 19:5874–5887, January 2021.
7. Janet Staats, Anagha Divekar, J. Philip McCoy, and Holden T. Maecker. Guidelines for Gating Flow Cytometry Data for Immunological Assays. In Jr McCoy, J. Philip, editor, *Immunophenotyping: Methods and Protocols*, Methods in Molecular Biology, pages 81–104. Springer, New York, NY, 2019.
8. Chris P. Verschoor, Alina Lelic, Jonathan L. Bramson, and Dawn M. E. Bowdish. An Introduction to Automated Flow Cytometry Gating Tools and Their Implementation. *Frontiers in Immunology*, 6, 2015.
9. Jacob H. Levine, Erin F. Simonds, Sean C. Bendall, Kara L. Davis, El-ad D. Amir, Michelle D. Tadmor, Oren Litvin, Harris G. Fienberg, Astraea Jager, Eli R. Zunder, Rachel Finck, Amanda L. Gedman, Ina Radtke, James R. Downing, Dana Pe'er, and Garry P. Nolan. Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. *Cell*, 162(1):184–197, July 2015.
10. V. A. Traag, L. Waltman, and N. J. van Eck. From Louvain to Leiden: guaranteeing well-connected communities. *Scientific Reports*, 9(1):5233, March 2019. Number: 1 Publisher: Nature Publishing Group.
11. Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of the Second International Conference on Knowledge Discovery and Data Mining*, KDD'96, pages 226–231, Portland, Oregon, August 1996. AAAI Press.
12. Xiao Liu, Weichen Song, Brandon Y. Wong, Ting Zhang, Shunying Yu, Guan Ning Lin, and Xianting Ding. A comparison framework and guideline of clustering methods for mass cytometry data. *Genome Biology*, 20(1):297, December 2019.
13. Tamim Abdelaal, Vincent van Unen, Thomas Höllt, Frits Koning, Marcel J.T. Reinders, and Ahmed Mahfouz. Predicting Cell Populations in Single Cell Mass Cytometry Data. *Cytometry Part A*, 95(7):769–781, 2019. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cyto.a.23738.
14. Teuvo Kohonen. Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43(1):59–69, January 1982.
15. Hao-Chih Lee, Roman Kosoy, Christine E Becker, Joel T Dudley, and Brian A Kidd. Automated cell type discovery and classification through knowledge transfer. *Bioinformatics*, 33(11):1689–1695, June 2017.
16. Michael J. Geuenich, Jinyu Hou, Sunyun Lee, Shanza Ayub, Hartland W. Jackson, and Kieran R. Campbell. Automated assignment of cell identity from single-cell multiplexed imaging and proteomic data. *Cell Systems*, 12(12):1173–1186.e5, December 2021.
17. Giuseppe Vettigli. MiniSom, October 2023. original-date: 2013-07-03T10:10:06Z.
18. Georg Pölzlbauer. Survey and Comparison of Quality Measures for Self-Organizing Maps.
19. J. Vesanto and E. Alhoniemi. Clustering of the self-organizing map. *IEEE Transactions on Neural Networks*, 11(3):586–600, May 2000. Conference Name: IEEE Transactions on Neural Networks.
20. James Bergstra, Daniel Yamins, and David Cox. Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures. In *Proceedings of the 30th International Conference on Machine Learning*, pages 115–123. PMLR, February 2013. ISSN: 1938-7228.
21. Sofie Van Gassen, Britt Callebaut, Mary J. Van Helden, Bart N. Lambrecht, Piet Demeester, Tom Dhaene, and Yvan Saeys. FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data. *Cytometry Part A*, 87(7):636–645, 2015. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cyto.a.22625.
22. Ahlers, Jannis, Althviz Moré, Daniel, Amsalem, Oren, Anderson, Ashley, Bokota, Grzegorz, Boone, Peter, Bragantini, Jordão, Buckley, Genevieve, Burt, Alister, Bussonnier, Matthias, Can Solak, Ahmet, Caporal, Clément, Doncila Pop, Draga, Evans, Kira, Freeman, Jeremy, Gaifas, Lorenzo, Gohlke, Christoph, Gunalan, Kabilar, Har-Gil, Hagai, Harfouche, Mark, Harrington, Kyle I. S., Hilsenstein, Volker, Hutchings, Katherine, Lambert, Talley, Lauer, Jessy, Lichtner, Gregor, Liu, Ziyang, Liu, Lucy, Lowe, Alan, Marconato, Luca, Martin, Sean, McGovern, Abigail, Migas, Lukasz, Miller, Nadalyn, Muñoz, Hector, Müller, Jan-Hendrik, Nauroth-Kreß, Christopher, Nunez-Iglesias, Juan, Pape, Constantin, Pevey, Kim, Peña-Castellanos, Gonzalo, Pierré, Andrea, Rodríguez-Guerra, Jaime, Ross, David, Royer, Loic, Russell, Craig T., Selzer, Gabriel, Smith, Paul, Sobolewski, Peter, Sofiiuk, Konstantin, Sofroniew, Nicholas, Stansby, David, Sweet, Andrew, Vierdag, Wouter-Michiel, Wadhwa, Pam, Weber Mendonça, Melissa, Windhager, Jonas, Winston, Philip, and Yamauchi, Kevin. napari: a multi-dimensional image viewer for Python, July 2023.
23. Sean C. Bendall, Erin F. Simonds, Peng Qiu, El-ad D. Amir, Peter O. Krutzik, Rachel Finck, Robert V. Bruggner, Rachel Melamed, Angelica Trejo, Olga I. Ornatsky, Robert S. Balderas, Sylvia K. Plevritis, Karen Sachs, Dana Pe'er, Scott D. Tanner, and Garry P. Nolan. Single-Cell Mass Cytometry of Differential Immune and Drug Responses Across a Human Hematopoietic Continuum. *Science*, 332(6030):687–696, May 2011. Publisher: American Association for the Advancement of Science.
24. Tyler Risom, David R. Glass, Inna Averbukh, Candace C. Liu, Alex Baranski, Adam Kagel, Erin F. McCaffrey, Noah F. Greenwald, Belén Rivero-Gutiérrez, Siri H. Strand, Sushama Varma, Alex Kong, Leeat Keren, Sucheta Srivastava, Chunfang Zhu, Zumana Khair, Deborah J. Veis, Katherine Deschryver, Sujay Vennam, Carlo Maley, E. Shelley Hwang, Jeffrey R. Marks, Sean C. Bendall, Graham A. Colditz, Robert B. West, and Michael Angelo. Transition to invasive breast cancer is associated with progressive changes in the structure and composition of tumor stroma. *Cell*, 185(2):299–310.e18, January 2022.

25. John W. Hickey, Winston R. Becker, Stephanie A. Nevins, Aaron Horning, Almudena Espin Perez, Chenchen Zhu, Bokai Zhu, Bei Wei, Roxanne Chiu, Derek C. Chen, Daniel L. Cotter, Edward D. Esplin, Annika K. Weimer, Chiara Caraccio, Vishal Venkataraaman, Christian M. Schürch, Sarah Black, Maria Brbić, Kaidi Cao, Shuxiao Chen, Weiruo Zhang, Emma Monte, Nancy R. Zhang, Zongming Ma, Jure Leskovec, Zhengyan Zhang, Shin Lin, Teri Longacre, Sylvia K. Plevritis, Yiing Lin, Garry P. Nolan, William J. Greenleaf, and Michael Snyder. Organization of the human intestine at single-cell resolution. *Nature*, 619(7970):572–584, July 2023. Number: 7970 Publisher: Nature Publishing Group.
26. William M. Rand. Objective Criteria for the Evaluation of Clustering Methods. *Journal of the American Statistical Association*, 66(336):846–850, December 1971. Publisher: Taylor & Francis Online.
27. David M. W. Powers. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation, October 2020. arXiv:2010.16061 [cs, stat].
28. Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, and Édouard Duchesnay. Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12(85):2825–2830, 2011.

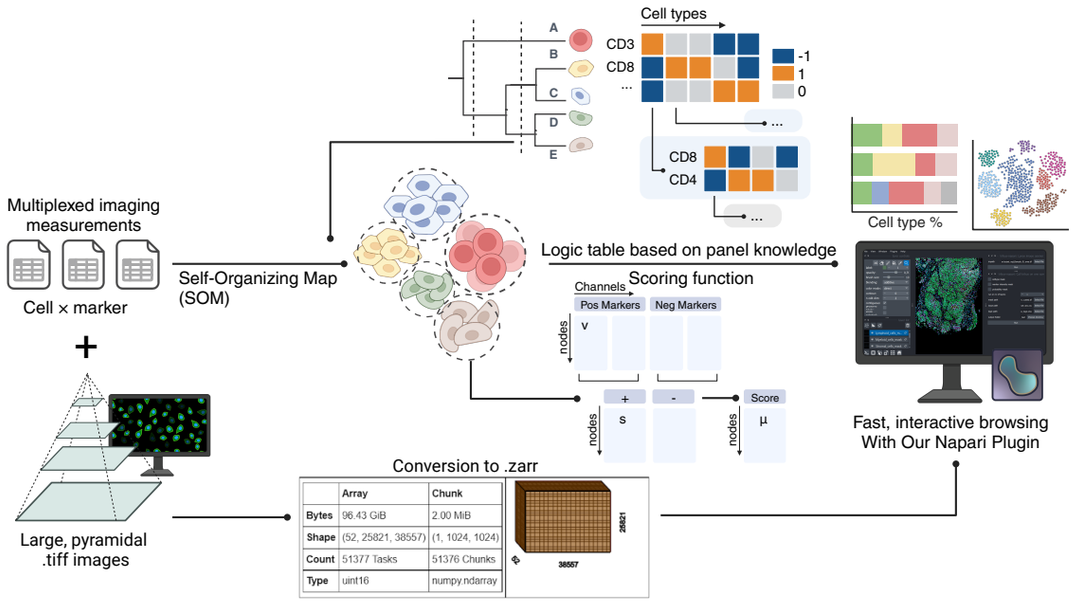

Fig. 1. Overview of Tribus architecture. Schematic diagram showing the workflow of Tribus and Napari plug-in.

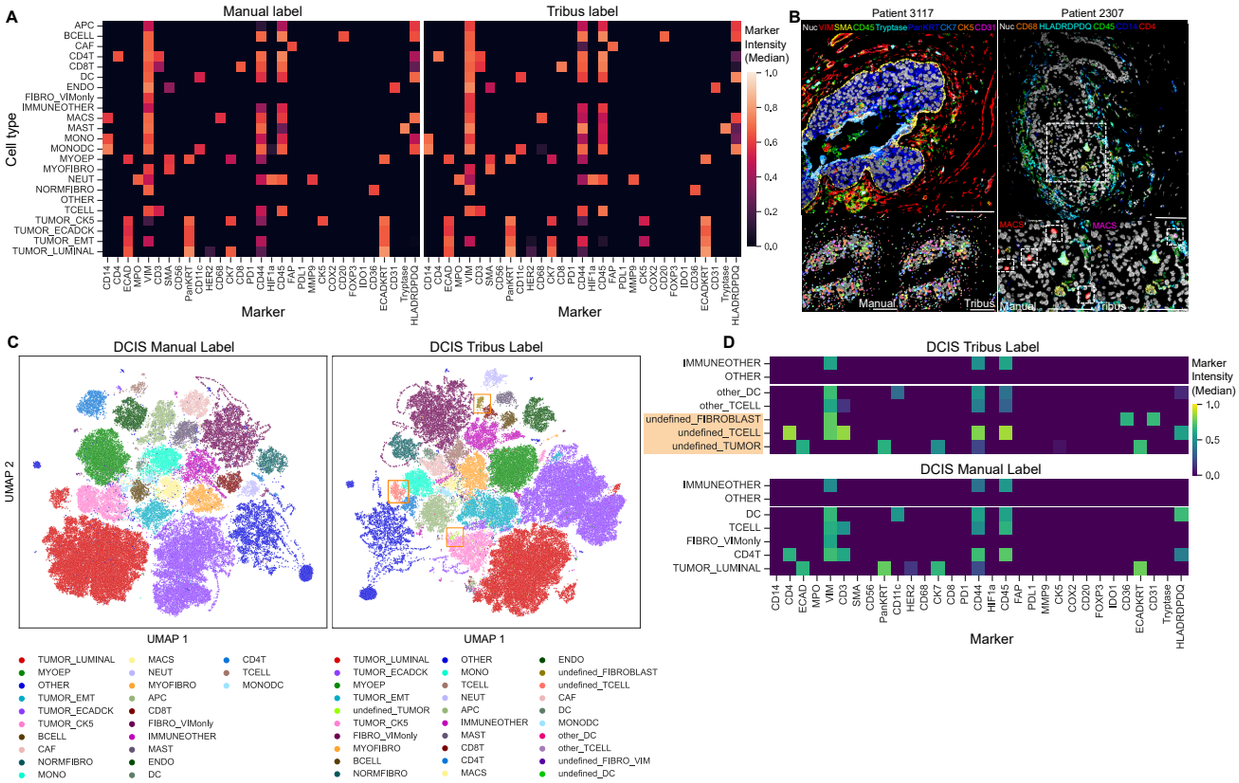

Fig. 2. Tribus applied to a published MIBI-TOF DCIS dataset. (A) Heatmaps showing the median marker intensity of manually gated cell populations and Tribus classification from the DCIS dataset are similar. (B) Representative MIBI images. The left image is from patient 3117 (DCIS tumor) with a nine-color overlay of markers related to major cell types. Cell-type masks from the manual label and Tribus show few differences. The right image is from patient 2307 (Normal tissue) with a six-color overlay of MACS-related markers. Compared to the MACS cell masks of manual and Tribus, some MACS cells from manual labels do not have related marker expression. (C) Supervised UMAP representation of manual and Tribus labels on the DCIS dataset. Cell types were color-coded based on original and Tribus annotation, the same cell type label was assigned the same color. (D) Compare undefined and other cell populations with ground truth labels.

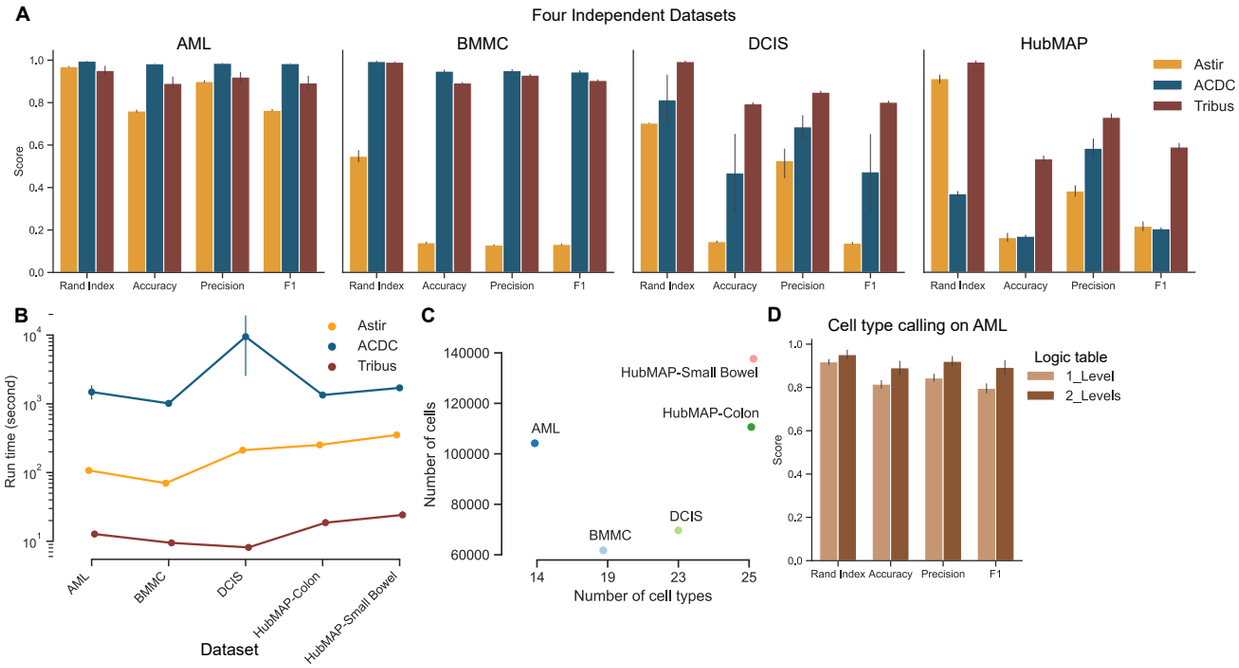

Fig. 3. (A) Performance comparison of Tribus and other similar methods on four datasets (AML, BMMC, DCIS, HubMAP) using four metrics for each. All analyses were repeated 5 times and the standard deviation for the error bar. (B) Models running time comparison. All analyses were repeated 5 times and used standard deviation as the error bar. (C) Data complexity comparison over multiple dataset sizes. (D) Compare Tribus performance under different logic tables for the AML dataset, using standard deviation as the error bar.

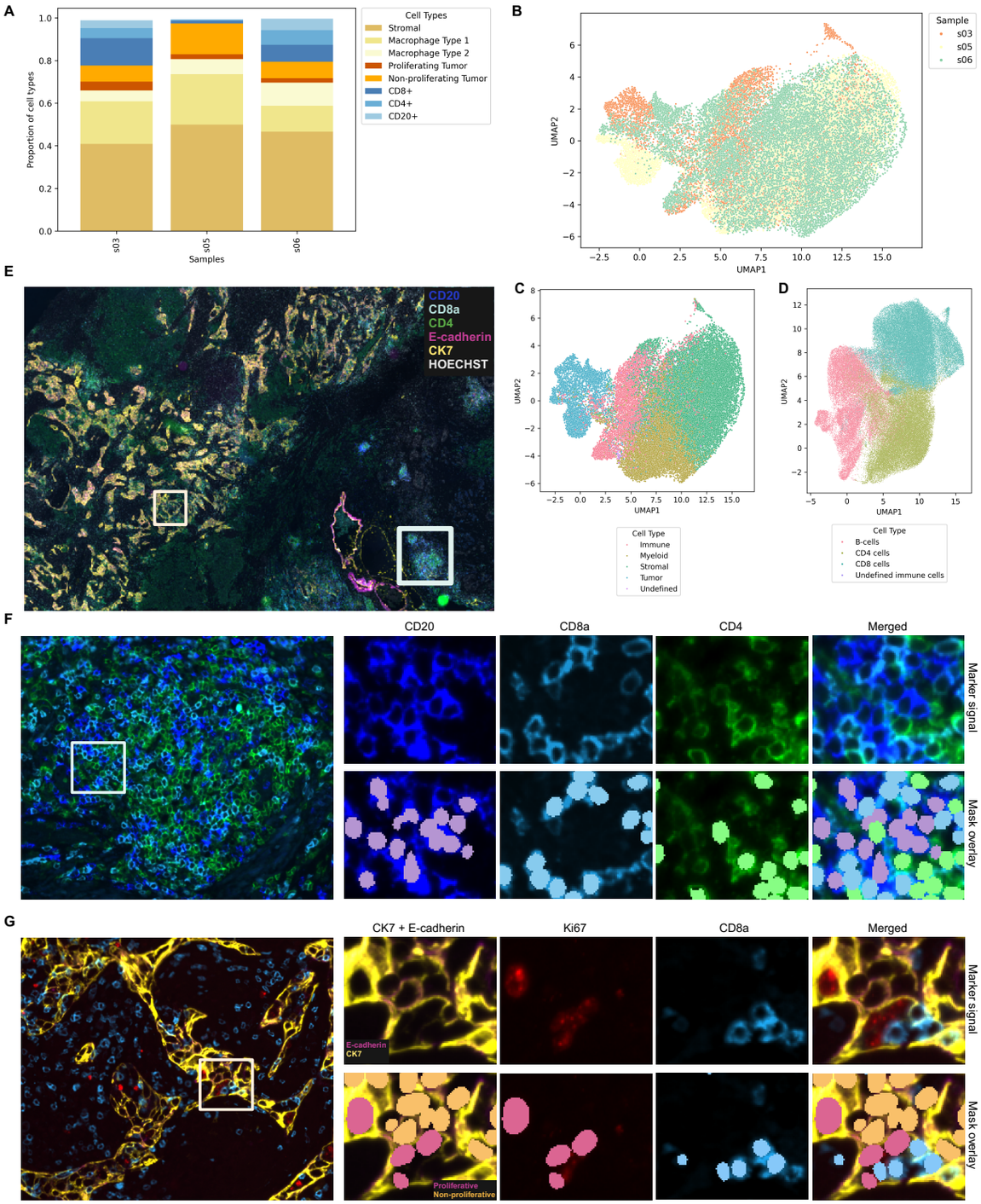

Fig. 4. (A) Annotated barplot of the cell phenotype compositions per sample. (B) UMAP visualizes cell populations colored by the patient of origin. (C) UMAP visualizes distinct cell populations, colored by the cell phenotypes. (D) UMAP visualizing the immune cell populations. (E) Representative tCyCIF image of the in-house dataset, showing the nuclei and representative tumor and immune markers. (F) Tribus accurately identified the CD20+, CD8a+ and CD4+ populations within a Tertiary Lymphoid Structure (TLS) despite the dense organization of these sub-phenotypes. (G) Representative image of a tumor area, showing Tribus annotation of proliferation in tumor cells and tumor-infiltrating CD8+ cells.