A Web-based Software Resource for Interactive Analysis of Multiplex Tissue Imaging Datasets

Allison L. Creason[1,*], Cameron Watson[1], Qiang Gu[1], Daniel Persson[1], Luke Sargent[1], Yu-An Chen[2,8], Jia-Ren Lin[2,8], Shamilene Sivagnanam[1], Florian Wünnemann[3], Ajit J. Nirmal[7], Koei Chin[1], Heidi S. Feiler[1,5], Lisa M. Coussens[1], Denis Schapiro[3,4], Björn Grüning[6], Peter K. Sorger[2,8], Artem Sokolov[2,8], and Jeremy Goecks[1,5,*]

[1] The Knight Cancer Institute, Oregon Health & Science University, Portland, OR USA
[2] Ludwig Center for Cancer Research at Harvard, Harvard Medical School, Boston, MA, USA
[3] Institute for Computational Biomedicine, Faculty of Medicine, Heidelberg University and Heidelberg University Hospital, Heidelberg, Germany
[4] Institute of Pathology, Heidelberg University Hospital, Heidelberg, Germany
[5] Department of Biomedical Engineering, School of Medicine, Oregon Health & Science University, Portland, OR USA
[6] Bioinformatics Group, Department of Computer Science, University of Freiburg, Freiburg, Germany
[7] Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA
[8] Laboratory of Systems Pharmacology, Harvard Medical School, Boston, MA, USA

[*] *Correspondence: creason@ohsu.edu, goecksj@ohsu.edu*

Abstract

Highly multiplexed tissue imaging (MTI) are powerful spatial proteomics technologies that enable *in situ* single-cell characterization of tissues. However, analysis and visualization of MTI datasets remains challenging, and we developed the Galaxy-ME software hub to address this challenge. Galaxy-ME is a web-based, interactive software hub that enables end-to-end analysis and visualization of MTI datasets and is accessible to everyone. To demonstrate its utility, Galaxy-ME was used to analyze datasets obtained from multiple MTI assays and evaluate assay concordance in both normal and cancerous tissues. Galaxy-ME is a publicly available web resource.

Main

Highly multiplexed tissue imaging (MTI) technologies, such as cyclic immunofluorescence (CycIF)[1], multiplex immunohistochemistry (mIHC)[2], Co-Detection by Indexing (CODEX)[3], imaging mass cytometry (IMC)[4], and multiplex ion beam imaging (MIBI)[5], are powerful *in situ* spatial proteomics technologies for characterizing tissues at single-cell, and potentially subcellular, resolution. With rapid adoption in basic and translational research, MTI has been used to identify differences in healthy and diseased tissue organization[3], quantify compositional and spatial features associated with cancer and other diseases, including disease development[6], patient survival[7–9], and response to therapy[10]. MTI is

also used extensively in several large tissue atlas consortia to create detailed 2D tissue maps for interrogating cellular and spatial relationships. Atlas consortia using MTI include the Human Cell Atlas (HCA)[11], the Human BioMolecular Atlas Program (HuBMAP)[12], and the Human Tumor Atlas Network (HTAN)[13].

With the substantial growth and use of MTI technologies in the biomedical research community, there is an acute need for robust analytical tools and visualizations of MTI datasets. A typical MTI dataset is tens or hundreds of gigabytes in size and includes images of a tissue sample from which each assayed marker protein is measured, yielding a stack of 30-100 images and tens or hundreds of thousands of cells analyzed per sample. A complete software analysis workflow for MTI often uses dozens of analysis tools to: (1) perform primary image processing to produce single-cell feature tables with marker intensity levels, morphological information, and spatial coordinates and (2) complete single-cell analysis to classify individual cells and quantify spatial relationships amongst cells. While several computational workflows that execute all tools in sequence for particular MTI assays, including CycIF[14], CODEX[15], mIHC[16], and IMC[17], have been developed[18,19] as well general purpose image analysis platforms, such as CellProfiler[20], QuPath[21], and Fiji[22], exist and provide some functionality for processing MTI datasets, there remains substantial challenges in accessibility, tool integration, and scalability that make it difficult to analyze MTI datasets. Analysis workflows that scale to many samples are often executed using a command line text interface, requiring significant computational expertise to use. Whereas some desktop applications integrate analysis and visualizations together, these applications cannot easily be deployed on disparate infrastructures at scale.

To address these challenges, we have developed Galaxy-MCMICRO Ecosystem (Galaxy-ME), a user-friendly and highly scalable, web-based software hub for interactive analysis of MTI datasets. Galaxy-ME provides a web-based software workbench for MTI analyses that is accessible to all scientists, a comprehensive tool and visualization suite for analysis of MTI datasets, and infrastructure to ensure that all analyses are scalable and reproducible (Figure 1). Galaxy-ME provides software analysis tools for (1) primary image processing to produce single-cell feature tables that include marker intensity levels, morphological information, and spatial coordinates; (2) single-cell analysis to classify individual cells and quantify spatial relationships amongst cells; and (3) interactive visualization of images and analysis results. Galaxy-ME is built on the Galaxy computational workbench (https://galaxyproject.org/)[23,24], an open-source platform for user-friendly, reproducible, and collaborative biomedical data analyses. Galaxy is among the most popular software analysis platforms in the world and used by thousands of scientists daily.

Galaxy-ME builds on and substantially advances our prior work developing primary image analysis workflows for multiplexed microscopy images via integration of the MCMICRO tool suite[19] into Galaxy. Galaxy-ME's advances include (1) expanding MCMICRO with many additional tools for analysis, visualization, and dashboarding of MTI datasets to create a tool suite of 17 tools that enables end-to-end analysis of MTI datasets, including both prior knowledge and data-driven based phenotyping and spatial analyses (Supplemental Table 1); (2) leveraging the full capabilities of Galaxy for data analysis, workflow editing and scalable execution, and interactive image viewers and dashboards for visualization of MTI datasets; and a (3) fully-featured and web-based platform for analysis of MTI datasets. The tools and visualizations in Galaxy-ME represent the current best-

practice analysis approaches and integrate analysis tools and visualizations from multiple tissue atlas consortia including HCA, HuBMAP, and HTAN. Using Galaxy-ME, scientists can analyze datasets from several MTI assays, including CycIF, mIHC, and CODEX.

Building on top of the Galaxy platform enables Galaxy-ME to take advantage of the accessibility, scalability, and reproducibility features that Galaxy offers. Galaxy-ME uses Galaxy's web-based graphical user interface (GUI), making analysis of MTI datasets widely accessible, regardless of computational expertise. The Galaxy-ME GUI makes it simple to move between selecting input datasets, running analysis tools/workflows, and visualizing imaging data or single cell analysis results. Galaxy-ME's analysis tools and visualizations are orchestrated and executed on remote computing resources by the Galaxy server. By using the Galaxy framework and sufficient remoting computing resources, Galaxy-ME can scale its analyses to process collections of imaging datasets that are hundreds of terabytes in size. Galaxy and Galaxy-ME are open-source and freely available. There are two public web services for using Galaxy-ME, https://cancer.usegalaxy.org/ and https://spatialomics.usegalaxy.eu/, and Galaxy-ME can also be downloaded and run locally as well.
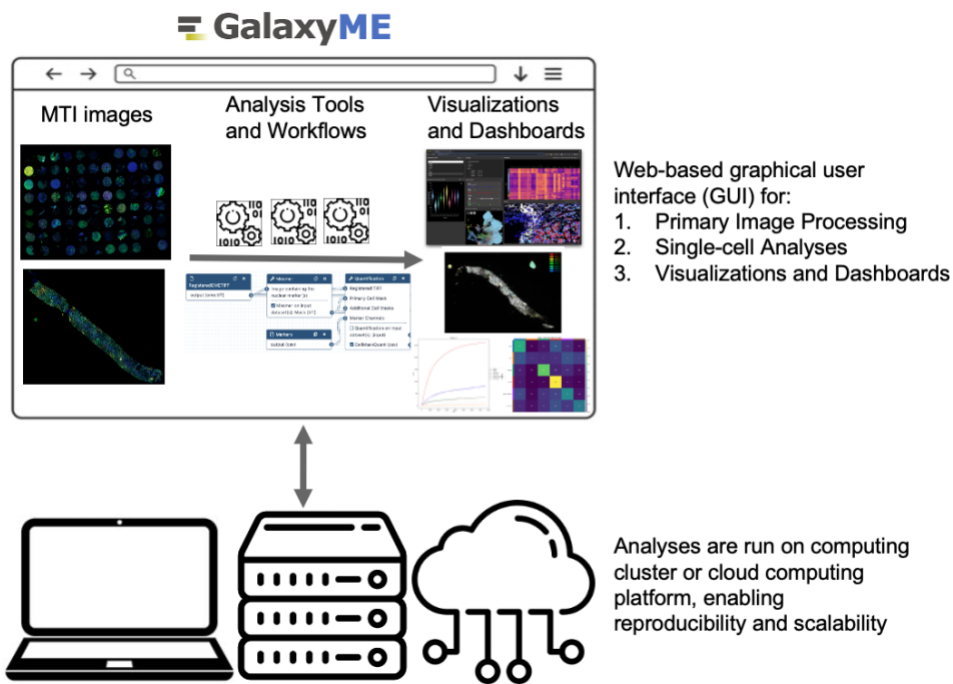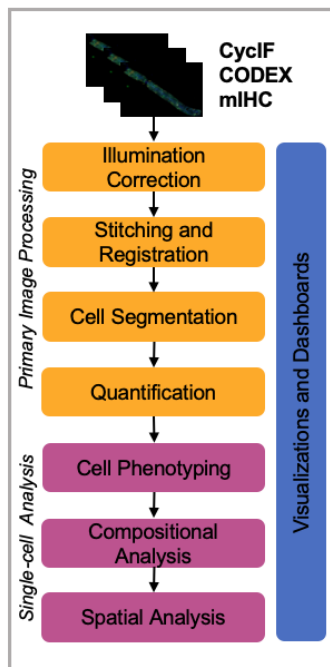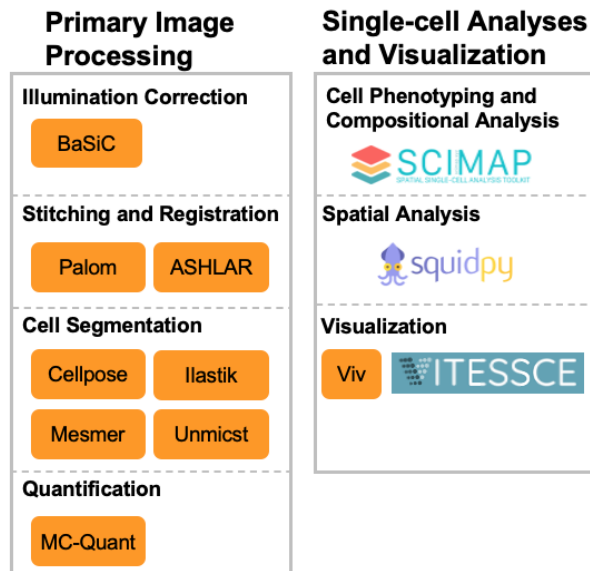
Figure 1: Galaxy-ME overview, Tools, and Visualizations. a, Galaxy-ME provides a web-based user interface to perform all analysis and visualization tasks for multiplexed tissue imaging datasets: primary image analysis and visualization, marker quantification, single-cell analyses, and integrated dashboards for visual exploration of both primary and single-cell data. Galaxy-ME analyses can be run on a laptop, computing cluster, or a cloud computing platform. All analyses are completely reproducible, and running Galaxy-ME on a computing cluster or cloud computing platform makes it possible to complete large-scale analyses. **b,** Galaxy-ME Tools and Visualizations overview of multiplexed tissue imaging analyses that can be performed in Galaxy-ME. These include primary image processing, single-cell analyses, and visualizations plus dashboards.

To demonstrate utility, Galaxy-ME was used to perform a fully automated analysis on both healthy and diseased tissue datasets from three distinct MTI assays (CycIF[1], mIHC[2], and CODEX[3]). The analysis included primary image processing and single cell analysis to quantify compositional and spatial features of the tissues and assess assay concordance. To the best of our knowledge, this is the first comparison of concordance across these MTI assays. A supplementary webpage linking to all Galaxy histories for this work is available at https://bit.ly/GalaxyME-Histories.

Galaxy-ME was used to perform an in-depth exploration of the compositional and spatial landscape using three MTI datasets generated by the HTAN[13] consortium where consecutive tissue sections from a healthy human tonsil tissue resection were profiled with CycIF, mIHC, and CODEX (Figure 2a). A Galaxy-ME workflow (Figure 2b) was implemented and applied to segment cells, quantify marker intensity, phenotype cells and quantify tissue cellular composition, compute spatial metrics between cells, and create Vitessce[25] dashboards for interactive analysis (Figure 2c). Mean cell count obtained from nuclear segmentation across the three assays was 69,329 cells with a standard deviation of 3,152 cells. The physical separation of the non-adjacent slides and assay-specific nuclear segmentation performance likely account for differences in total cell counts. The proportion of positive cells for the shared markers across the three MTI assays—Pan Cytokeratin, CD45, CD20, and CD8—was concordant across the 3 assays, with the exception of CD45, likely due to antibody performance issues and illumination artifacts in the CODEX data (Figure S1). Cell populations and tissue composition were highly concordant across the assays (Figure 2d). Spatial patterns across assays based on the normalized spatial interaction score[26] between cell populations were highly similar, demonstrating spatial organization is preserved across assays (Figure 2e).
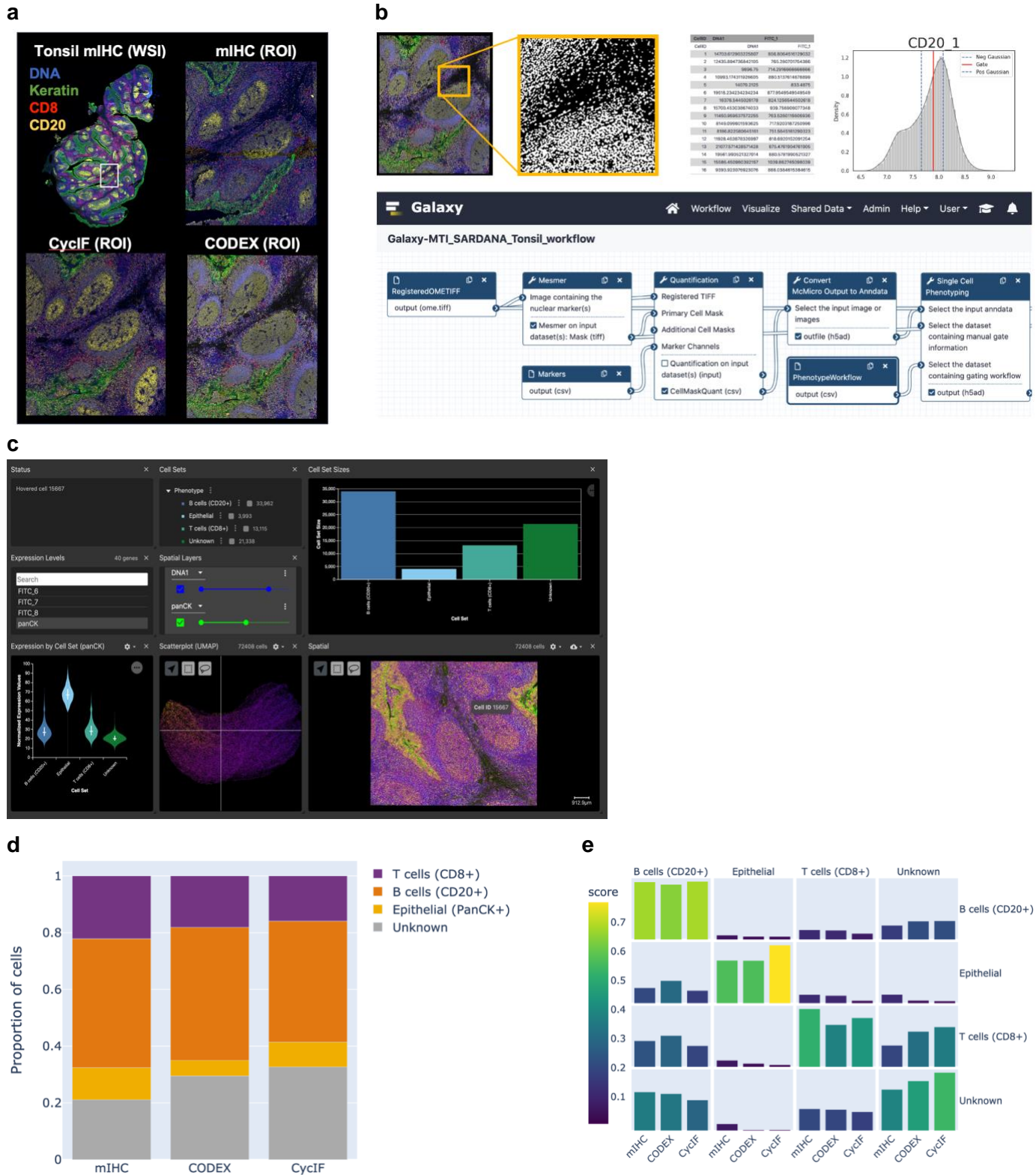
Figure 2: Comparison of three multiplex tissue imaging assays on a healthy tonsil specimen. **a**, Overview of the whole slide tonsil image profiled with mIHC (top left) and the region of interest using mIHC (top right), CycIF (bottom left), and CODEX (bottom right) pseudo-colored with the 5 shared markers for DNA (blue), Keratin/PanCK (green), CD8 (red), and CD20 (yellow). **b**, the Galaxy-ME workflow and history used for the CODEX tonsil image analysis. **c**, a Vitessce dashboard, launched out of the CODEX analysis history in Galaxy, includes views of the phenotype-labeled segmentation mask overlaid onto the registered image, compositional barplots, UMAP representations, violin plots of marker expression, and heatmaps (not shown). Individual cells can be highlighted using a cursor and data representing highlighted cells will be spotlighted in each of the plots. **d**, stacked barplots cell phenotype proportions in the tonsil ROI across each assay. **e**, barplot grid of pairwise spatial interaction scores across cell phenotypes.

Galaxy-ME was next applied to datasets generated using CycIF and mIHC on adjacent serial sections from an HTAN colorectal cancer (CRC) resection. Analysis with Galaxy-ME was performed on 7 regions of interest (ROIs) that showed a variety of tissue compositions (Figure 3a, S2, S3). Consistent with the tonsil analysis, total cell counts based on nuclear segmentation were concordant across the ROIs for both CycIF and mIHC. In CycIF, 46,510 mean cell counts per ROI were found for a total of 325,573 cells; in mIHC, 44,287 mean cell counts per ROI were found for a total of 310,006 cells (Figure 3b). Overall cell phenotype counts and tissue composition were highly correlated across CycIF and mIHC (r=0.89; Figure 3c, S4). To identify common cell type spatial patterns across assays, recurrent cellular neighborhoods (RCN)[27] were computed across all 7 ROIs using Latent Dirichlet Allocation (LDA) and K-means clustering (Figure S5). Across the two assays, 20 RCNs were identified with most ROIs displaying highly similar RCN composition (Cosine > 0.88), though ROI 14 and 15 had more variable composition (Cosine 0.67 and 0.76, respectively) due to lower cell density and higher stromal content (Figure 3d, S6). Common RCNs across assays had diverse cell type composition (Figure 3e) and were often enriched with epithelial cells (RCNs 1,5,8,14,16,18), M2 macrophages (RCN10), or CD8 T cells (RCN6) (Figure 3e, S7). Comparing RCNs across assays demonstrated that RCNs from both assays captured common spatial structures such as epithelial-stromal interface, and mixed populations of immune cells, and previously identified lymphoid follicles (Figure 3f).
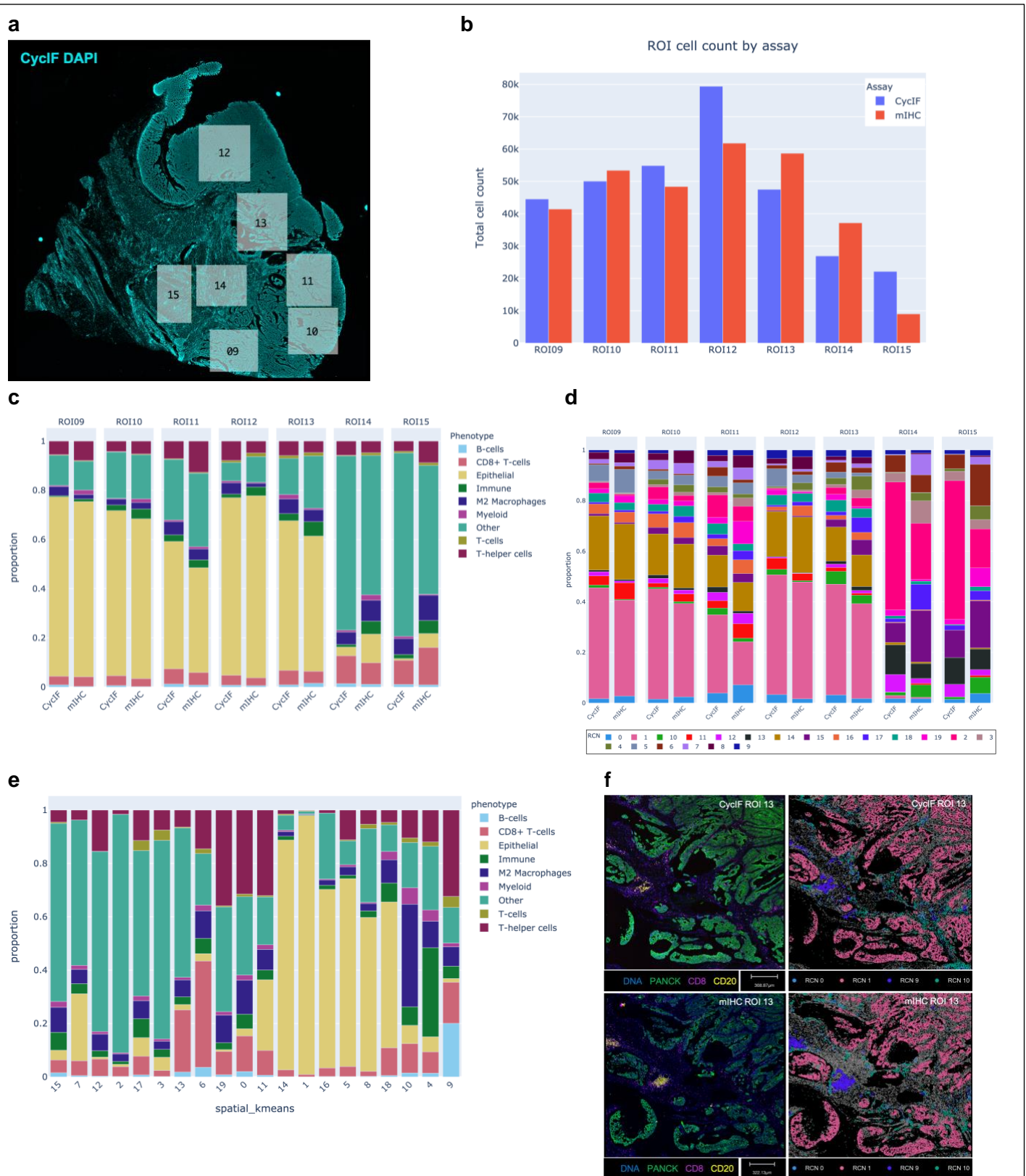
Figure 3: Compositional and spatial analysis of colorectal cancer (CRC) tissue assayed using CyclF and mIHC. a, Whole slide image of CRC tissue section with DAPI stain imaged by CyclF with regions of interest (ROIs) depicted in rectangular boxes. **b**, CRC total cell counts based on nuclear segmentation for each ROI across the CyclF (blue) and mIHC (red) assays. **c**, cell phenotype proportions for each ROI across the two assays. **d**, cellular composition of recurrent cellular neighborhoods (RCNs) for each ROI across both mIHC and CyclF assays. **e,** cell phenotype composition across the RCNs shared between assays. **f**, tissue image and computationally-identified RCNs shown side-by-side. Tissue image in ROI13 for CyclF (top left) and mIHC (bottom left) and selected RCNs in CyclF (top right) and mIHC (bottom right). Images are colored using four proteins: DNA (blue), PanCK (green), CD8 (purple), and CD20 (yellow). RCNs matching these proteins are colored based on ID: RNC0 in blue, RNC1 in pink, RNC9 in purple, RNC10 in teal (bottom right).

Analysis and visualization of multiplexed tissue imaging datasets remains a challenging problem as these datasets are very large and require using many analysis tools and visualizations together, in an iterative and interactive manner. Galaxy-ME is an interactive, web-based software hub, tool suite, and workflow engine that uniquely centralizes a broad collection of image processing and single-cell analytics methods for comprehensive analysis of MTI datasets. Galaxy-ME tools can be connected with more than 8,500 other tools available via the Galaxy Tool Shed[28], to create analyses that extend far beyond imaging, such as using machine learning tools[29] or integration with omics data for multimodal analysis. By streamlining analysis and visualization of MTI datasets into a web-based, graphical user interface, Galaxy-ME overcomes several barriers that other analysis approaches often encounter and democratizes access, analysis, and visualization of MTI datasets so that any scientist, regardless of their informatics expertise, can work with MTI datasets.

Th analysis of MTI datasets from healthy human tonsil and colorectal cancer tissue specimens with Galaxy-ME demonstrates the feasibility of automated analysis across multiple MTI assays with the ability to produce high-quality single-cell compositional and spatial results. The high concordance found between MTI assays in all facets of analysis—cell counts, tissue composition, and tissue spatial organization—provides confidence in both the imaging datasets produced by the multiplex tissue imaging assays as well as Galaxy-ME's analysis tools and workflows. Looking forward, there are opportunities to add additional tools and visualizations that extend Galaxy-ME beyond single-cell analysis to focus on quantification and organization of larger units of organization such as functional tissue units[30] or subcellular characteristics of cells.

Methods

Primary Image Processing

Primary image processing in Galaxy-ME is includes all of the MCMICRO[19] suite of tools, in addition to supplemental tools for key steps (Supplemental Table 1). Image processing steps include: (1) illumination correction between microscopy tiles (with the BaSiC[31] tool); (2) stitching of tiles into channel mosaics (ASHLAR[32]); (3) registration of channel mosaics into a multi-channel OME-TIFF pyramid (ASHLAR, PALOM); (4) single-cell segmentation (UnMICST[33] and S3segmenter[34], Ilastik[35], Cellpose[36], Mesmer[37]); and (5) quantification of protein marker intensities for every cell (MCQuant). Processing tissue microarrays requires an additional step to dearray sample cores into separate images (Unet Coreograph), which is performed after illumination correction and stitching. Single-cell segmentation—the process of creating an image mask that assigns pixels into individual indexed cells—is a challenging but critical image processing step that can vary dramatically in performance between assays and tissues. For this reason, several segmentation tools have been integrated into Galaxy-ME so that the best segmentation method can be applied to each dataset. Final outputs from primary image processing are (1) a multi-channel pyramidal OME-TIFF file that includes all image channels and (2) a cell feature table with mean protein marker intensities and morphological features (area, eccentricity, orientation, and solidity) for each cell identified in the segmentation mask.

Tool Suite	Purpose	Number of Tools
MCMICRO[17]	Primary image processing	8
Viv[30]	Primary image viewing	1
Cellpose[26]	Cell segmentation	1
Mesmer[27]	Cell segmentation	1
Scimap[17]	Single cell phenotyping and spatial analysis	3
Squidpy[29]	Single cell spatial analysis	1
Vitessce[31]	Dashboard of interactive visualizations	2

Supplemental Table 1. **Tools and visualizations in Galaxy-ME**.

Single-Cell Analysis and Visualization

Single-cell analysis tools in Galaxy-ME use the cell feature table produced by primary image processing for cell phenotyping, compositional analysis, and spatial analysis (Figure 1). Galaxy-ME supports single-cell phenotyping using two complementary approaches. First, cells can be gated using a biologically-driven, semi-automated hierarchical gating approach using the SciMap[27] package from MCMICRO. The output of this approach is a distinct phenotype for each cell, such as *CD8+ T-cell* or *luminal neoplastic tumor cell*. The second approach for phenotyping cells is a data-driven approach where cells are clustered using community metrics[38,39] (e.g. Louvain, Leiden). This approach produces clusters that can then be annotated based on the markers enriched in each cluster. Galaxy-ME uses spatial analysis methods in SciMap and SquidPy[26] to quantify spatial interactions, neighbor enrichment between cell types, spatial neighborhoods, and other metrics of tissue spatial organization. Galaxy-ME uses the common ANNData format (https://anndata.readthedocs.io) for storing single-cell data, allowing for easy integration with existing Galaxy tool suites, like ScanPy[40] and Seurat[41].

There are three primary visualizations available in Galaxy-ME. Avivator is a light-weight and web-based image viewer built with the Viv[42] library that enables viewing multi-channel OME-TIFF images hosted on a web server. With Avivator, different image channels can be selected and visualized, and it is possible to pan and zoom around images. For additional data viewing, Galaxy-ME provides a tool to create Vitessce[25] interactive dashboards that include multiple connected visualizations. These visualizations include multiplex images augmented with labeled segmentation masks as well as phenotype information, UMAP plots, phenotype marker enrichment, and single-cell heatmaps.

Automated Image Processing and Cell Classification of SARDANA Samples

Registered images from the healthy Tonsil and CRC samples were cropped to match annotated regions of interest (ROIs) across all MTI assays. Nuclei were segmented using Mesmer on the Hematoxylin or DAPI channels with the appropriate image resolution. MCQuant was used to extract single-cell mean pixel intensities, morphological characteristics, and spatial coordinate positions. The resulting cell feature tables were converted to AnnData format for downstream single cell and spatial analysis.

Individual cells were labeled as positive or negative for each protein using a threshold for marker positivity that was determined automatically using Scimap (*scimap.pp.rescale*) and validated with manual gating. The marker positivity calls were used to assign cell phenotypes with the hierarchical phenotyping tool available in Scimap (*scimap.tl.phenotype_cells*). For the tonsil sample, across CycIF, mIHC and CODEX, the following cell types were called based on marker positivity: pan-cytokeratin for epithelial, CD8 for cytotoxic T-cells, and CD20 for B-cells. For the CRC sample, cyclF and mIHC had more markers in common, so a deeper immune cell classification scheme was used. For each sample set, basic compositional metrics, including total cell counts and relative frequency of each phenotype were calculated and compared across assays.

Cross-Assay Spatial Analysis of SARDANA Tonsil Sample

The Tonsil CycIF, mIHC, and CODEX single cell data, labeled with phenotype, was used for subsequent spatial analysis. Spatial neighborhood graphs were constructed for each ROI with Squidpy (*squidpy.gr.spatial_neighbors*) using default parameters (n_neighbors = 6). Spatial patterns were quantified for each assay by calculating a normalized (scaled between 0-1) spatial interaction score (*squidpy.gr.interaction_matrix*), a measure of how clustered a group of nodes in a graph are, between each phenotype using the neighborhood graphs.

Cross-Assay Spatial Analysis of SARDANA CRC Sample

For the CycIF and mIHC CRC images, spatial neighborhood matrices were generated for each ROI separately which quantified the frequency of neighbor cell phenotypes within a 30 micron radius for every epithelial cell in the tissue. The epithelial cell neighborhood matrices from all ROIs across both assays were combined and Scimap's implementation of Latent Dirichlet Allocation (n_motifs = 10) was run on the combined matrix to generate latent space weights. The resulting weights were clustered using K-means clustering (k = 20) to generate meta-clusters, or Recurrent Cellular Neighborhoods (RCNs)[27]. Hierarchical clustering was used to find biopsies that had similar RCN compositions. RCNs were annotated based on their composition.

Data Availability

The tonsil image datasets analyzed are available at https://www.synapse.org/MCMICRO_images, and the CRC images are available at https://www.synapse.org/#!Synapse:syn47164089. In addition, all datasets will be made available through the HTAN Data Portal.

Code Availability

The Galaxy-ME code repository, including galaxy tool wrappers and Docker files, are available at: https://github.com/goeckslab/tools-mti.

Acknowledgements

Competing Interests

J. Goecks has a significant financial interest in GalaxyWorks, a company that may have a commercial interest in the results of this research and technology. This competing interest is managed by OHSU, and Goecks declares that this relationship has not influenced the content of this manuscript. P.K. Sorger is a member of the SAB or BOD member of Applied Biomath, RareCyte Inc. and Glencoe Software, which distributes a commercial version of the OMERO database; P.K. Sorger is also a member of the NanoString SAB. In the last 5 years, the Sorger laboratory has received research funding from Novartis and Merck. Sorger declares that none of these relationships have influenced the content of this manuscript.

References

1. Lin, J.-R. *et al.* Highly multiplexed immunofluorescence imaging of human tissues and tumors using t-CyCIF and conventional optical microscopes. *Elife* **7**, (2018).

2. Tsujikawa, T. *et al.* Quantitative Multiplex Immunohistochemistry Reveals Myeloid-Inflamed Tumor-Immune Complexity Associated with Poor Prognosis. *Cell Rep.* **19**, 203–217 (2017).

3. Goltsev, Y. *et al.* Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed Imaging. *Cell* **174**, 968-981.e15 (2018).

4. Giesen, C. *et al.* Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nat. Methods* **11**, 417–422 (2014).

5. Angelo, M. *et al.* Multiplexed ion beam imaging of human breast tumors. *Nat. Med.* **20**, 436–442 (2014).

6. Risom, T. *et al.* Transition to invasive breast cancer is associated with progressive changes in the structure and composition of tumor stroma. *Cell* **185**, 299-310.e18 (2022).

7. Liudahl, S. M. *et al.* Leukocyte Heterogeneity in Pancreatic Ductal Adenocarcinoma: Phenotypic and Spatial Features Associated with Clinical Outcome. *Cancer Discov.* **11**, 2014–2031 (2021).

8. Keren, L. *et al.* A Structured Tumor-Immune Microenvironment in Triple Negative Breast Cancer Revealed by Multiplexed Ion Beam Imaging. *Cell* **174**, 1373-1387.e19 (2018).

9. Jackson, H. W. *et al.* The single-cell pathology landscape of breast cancer. *Nature* **578**, 615–620 (2020).

10. Berry, S. *et al.* Analysis of multispectral imaging with the AstroPath platform informs efficacy of PD-1 blockade. *Science* **372**, (2021).

11. Regev, A. *et al.* The Human Cell Atlas. *Elife* **6**, (2017).

12. HuBMAP Consortium. The human body at cellular resolution: the NIH Human Biomolecular Atlas Program. *Nature* **574**, 187–192 (2019).

13. Rozenblatt-Rosen, O. *et al.* The Human Tumor Atlas Network: Charting Tumor Transitions across Space and Time at Single-Cell Resolution. *Cell* **181**, 236–249 (2020).

14. Eng, J. *et al.* cmIF: A Python Library for Scalable Multiplex Imaging Pipelines. in *Mathematical and Computational Oncology* 37–43 (Springer International Publishing, 2019).

15. Schürch, C. M. *et al.* Coordinated Cellular Neighborhoods Orchestrate Antitumoral Immunity at the Colorectal Cancer Invasive Front. *Cell* **182**, 1341-1359.e19 (2020).

16. Blom, S. *et al.* Systems pathology by multiplexed immunohistochemistry and whole-slide digital image analysis. *Sci. Rep.* **7**, 15580 (2017).

17. Schapiro, D. *et al.* histoCAT: analysis of cell phenotypes and interactions in multiplex image cytometry data. *Nat. Methods* **14**, 873–876 (2017).

18. Czech, E., Aksoy, B. A., Aksoy, P. & Hammerbacher, J. Cytokit: a single-cell analysis toolkit for high dimensional fluorescent microscopy imaging. *BMC Bioinformatics* **20**, 448 (2019).

19. Schapiro, D. *et al.* MCMICRO: a scalable, modular image-processing pipeline for multiplexed tissue imaging. *Nat. Methods* (2021) doi:10.1038/s41592-021-01308-y.

20. McQuin, C. *et al.* CellProfiler 3.0: Next-generation image processing for biology. *PLoS Biol.* **16**, e2005970 (2018).

21. Bankhead, P. *et al.* QuPath: Open source software for digital pathology image analysis. *Sci. Rep.* **7**, 16878 (2017).

22. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).

23. Jalili, V. *et al.* The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2020 update. *Nucleic Acids Res.* (2020) doi:10.1093/nar/gkaa554.

24. Goecks, J., Nekrutenko, A., Taylor, J. & Galaxy Team. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* **11**, R86 (2010).

25. Keller, M. S. *et al.* Vitessce: a framework for integrative visualization of multi-modal and spatially-resolved single-cell data. (2021) doi:10.31219/osf.io/y8thv.

26. Palla, G. *et al.* Squidpy: a scalable framework for spatial omics analysis. *Nat. Methods* **19**, 171–178 (2022).

27. Nirmal, A. J. *et al.* The Spatial Landscape of Progression and Immunoediting in Primary Melanoma at Single-Cell Resolution. *Cancer Discov.* **12**, 1518–1541 (2022).

28. Blankenberg, D. *et al.* Dissemination of scientific software with Galaxy ToolShed. *Genome Biol.* **15**, 403 (2014).

29. Gu, Q. *et al.* Galaxy-ML: An accessible, reproducible, and scalable machine learning toolkit for biomedicine. *PLoS Comput. Biol.* **17**, e1009014 (2021).

30. de Bono, B., Grenon, P., Baldock, R. & Hunter, P. Functional tissue units and their primary tissue motifs in multi-scale physiology. *J. Biomed. Semantics* **4**, 22 (2013).

31. Peng, T. *et al.* A BaSiC tool for background and shading correction of optical microscopy images. *Nat. Commun.* **8**, 14836 (2017).

32. Muhlich, J. L. *et al.* Stitching and registering highly multiplexed whole-slide images of tissues and tumors using ASHLAR. *Bioinformatics* **38**, 4613–4621 (2022).

33. Yapp, C. *et al.* UnMICST: Deep learning with real augmentation for robust segmentation of highly multiplexed images of human tissues. *Commun Biol* **5**, 1263 (2022).

34. Saka, S. K. *et al.* Immuno-SABER enables highly multiplexed and amplified protein imaging in tissues. *Nat. Biotechnol.* **37**, 1080–1090 (2019).

35. Berg, S. *et al.* ilastik: interactive machine learning for (bio)image analysis. *Nat. Methods* **16**, 1226–1232 (2019).

36. Stringer, C., Wang, T., Michaelos, M. & Pachitariu, M. Cellpose: a generalist algorithm for cellular segmentation. *Nat. Methods* **18**, 100–106 (2021).

37. Greenwald, N. F. *et al.* Whole-cell segmentation of tissue images with human-level performance using large-scale data annotation and deep learning. *Nat. Biotechnol.* **40**, 555–565 (2022).

38. Traag, V. A., Waltman, L. & van Eck, N. J. From Louvain to Leiden: guaranteeing well-connected communities. *Sci. Rep.* **9**, 5233 (2019).

39. Levine, J. H. *et al.* Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. *Cell* **162**, 184–197 (2015).

40. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol.* **19**, 15 (2018).

41. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).

42. Manz, T. *et al.* Viv: Multiscale Visualization of High-Resolution Multiplexed Tissue Data on the Web. (2020) doi:10.31219/osf.io/wd2gu.