

1 **MCMICRO: A scalable, modular image-processing pipeline for multiplexed tissue imaging**

2

3 Denis Schapiro<sup>1,2,3</sup>, Artem Sokolov<sup>1,2</sup>, Clarence Yapp<sup>1,2,4</sup>, Jeremy L. Muhlich<sup>1,2</sup>, Joshua Hess<sup>5</sup>, Jia-  
4 Ren Lin<sup>1,2</sup>, Yu-An Chen<sup>1,2</sup>, Maulik K. Nariya<sup>1,2</sup>, Gregory J. Baker<sup>1,2</sup>, Juha Ruukonen<sup>1,2</sup>, Zoltan  
5 Maliga<sup>1,2</sup>, Connor A. Jacobson<sup>1,2</sup>, Samouil L. Farhi<sup>3</sup>, Domenic Abbondanza<sup>3</sup>, Eliot T. McKinley<sup>6,7</sup>,  
6 Courtney Betts<sup>8</sup>, Aviv Regev<sup>3,9,10</sup>, Robert J. Coffey<sup>11</sup>, Lisa M. Coussens<sup>8,12</sup>, Sandro Santagata<sup>1,2,13</sup>  
7 and Peter K. Sorger<sup>1,2,14</sup>

8

9 The Human Tumor Atlas Network

10 <sup>1</sup>Ludwig Center for Cancer Research at Harvard, Harvard Medical School, Boston, MA

11 <sup>2</sup>Laboratory of Systems Pharmacology, Harvard Medical School, Boston, MA

12 <sup>3</sup>Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA

13 <sup>4</sup>Image and Data Analysis Core, Harvard Medical School, Boston, MA

14 <sup>5</sup>Vaccine and Immunotherapy Center, Massachusetts General Hospital, Harvard Medical School,  
15 Boston, MA

16 <sup>6</sup>Epithelial Biology Center, Vanderbilt University Medical Center, Nashville, TN

17 <sup>7</sup>Department of Cell and Developmental Biology, Vanderbilt University School of Medicine,  
18 Nashville, TN

19 <sup>8</sup>Department of Cell, Developmental & Cancer Biology, Oregon Health & Science University,  
20 Portland, OR

21 <sup>9</sup>Department of Biology, Howard Hughes Medical Institute, Massachusetts Institute of Technology,  
22 Cambridge, MA, USA

23 <sup>10</sup>Present address: Genentech, South San Francisco, CA, USA

24 <sup>11</sup>Division of Gastroenterology, Hepatology, and Nutrition, Department of Medicine, Vanderbilt  
25 University Medical Center, Nashville, TN

26 <sup>12</sup>Knight Cancer Institute, Oregon Health & Science University, Portland, OR

27 <sup>13</sup>Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA

28 <sup>14</sup>Department of Systems Biology, Harvard Medical School, Boston, MA

29

30 Corresponding author:

31 Peter K. Sorger, Harvard Medical School, 200 Longwood Avenue, Warren Alpert Building, Room  
32 440, Boston, MA 02115; Telephone: 617-432-6901; email: [peter\\_sorger@hms.harvard.edu](mailto:peter_sorger@hms.harvard.edu);

33 [sorger\\_admin@hms.harvard.edu](mailto:sorger_admin@hms.harvard.edu)

34

## 35 **ABSTRACT**

36 Highly multiplexed tissue imaging makes molecular analysis of single cells possible in a  
37 preserved spatial context. However, reproducible analysis of the underlying data poses a substantial  
38 computational challenge. Here we describe a modular and open-source computational pipeline  
39 (MCMICRO) for performing the sequential steps needed to transform large, multi-channel whole  
40 slide images into single-cell data. We demonstrate use of MCMICRO on images of different tissues  
41 and tumors acquired using multiple imaging platforms, thereby providing a solid foundation for the  
42 continued development of tissue imaging software.

43

## 44 **MAIN**

45 The recent introduction of highly multiplexed tissue imaging makes it possible to measure the  
46 levels and localization of 20-100 antigens at subcellular resolution in a preserved 3D environment  
47 (see **Table S1** for references). In a research setting, multiplexed imaging provides new insight into  
48 molecular properties of tissues and their spatial organization and, in a clinical setting, it promises to  
49 augment traditional histopathological diagnosis of disease with the molecular information needed to  
50 guide use of targeted and immuno-therapies<sup>1-4</sup>. Inadequate tools for image processing remain a  
51 substantial barrier to the routine use of multiplexed tissue imaging, particularly in the case of whole-  
52 slide imaging (WSI), in which specimens as large as 5 cm<sup>2</sup> are imaged in their entirety. Diagnostic  
53 histopathology is based on WSI, and the FDA mandates it for medical applications<sup>5,6</sup>. We have also  
54 found that multiplexed WSI is essential for accurately quantifying the mesoscale structures that  
55 organize tissues<sup>7</sup>. Whole-slide images can contain one terabyte of data, 10<sup>5</sup> to 10<sup>6</sup> cells, and involve  
56 resolvable structures with spatial scales from 100 nm to over 1 cm. This represents a substantial  
57 challenge for computational image analysis.

58 A goal common to almost all multiplexed tissue analyses is identifying cell locations,  
59 phenotypes and states based on the levels and patterns of expression of protein markers. These are  
60 usually detected using antibodies, often in conjunction with stains such as hematoxylin and eosin  
61 (H&E). Image-based single-cell analysis is a natural complement to spatial and single-cell  
62 transcriptomics<sup>8-10</sup> but faces four computational challenges: (i) image segmentation, the process of  
63 subdividing images into areas comprising single cells, is difficult when normal tissue structures are  
64 disrupted, cells are densely crowded, and nuclei have irregular morphologies – as in cancer; (ii) the  
65 fundamental units of tissue organization are highly variable, and the essential data types are not well  
66 defined; (iii) WSI generates very large files that must be available for human inspection (a 50-plex 4  
67 cm<sup>2</sup> image collected at 0.3μm lateral resolution comprises over 400 GB of data); (iv) image  
68 processing algorithms are simultaneously being developed by many research groups in parallel,

69 using different programming languages (some proprietary, such as MATLAB) without consideration  
70 of interoperability. Analogous challenges in genomics have been addressed by developing  
71 computational pipelines that streamline multi-step data analyses and can also be scaled up to cloud  
72 compute environments (e.g., Cumulus for scRNAseq)<sup>11</sup>. The use of pipelines involving software  
73 containers (e.g., Docker<sup>12</sup>) and workflow languages<sup>13</sup> makes it possible for multiple research groups  
74 to contribute to and iteratively improve complex computational tasks. In the case of tissue atlases,  
75 such as the Human Tumor Atlas Network (HTAN)<sup>14</sup>, multiple laboratories are faced with a common  
76 set of data analysis challenges, a further motivation for a standardized computational framework.

77 In this paper we describe MCMICRO (Multiple Choice MICROscopy), a scalable, modular,  
78 and open source image processing pipeline implemented in the Nextflow language<sup>15</sup> that leverages  
79 Docker/Singularity containers<sup>12,16</sup>. We show that MCMICRO can process multiplexed data acquired  
80 using at least six different imaging technologies (**Table S1**) and has attributes not found in existing  
81 workflows (**Table S2**). These include the ability to select among competing algorithms at key steps  
82 in the analysis and interactive training of machine learning models (this is particularly important for  
83 image segmentation). In common with other bioinformatics pipelines, MCMICRO is designed to  
84 complement rather than replace conventional desktop and server-deployed tools. A wide variety of  
85 algorithms can be incorporated into the MCMICRO pipeline using containers, and the results can be  
86 visualized using multiple software environments, including napari, QuPath, OMERO and histoCAT  
87 (see **Table S2** for details and references).

88 To create MCMICRO, we re-implemented as open-source software several algorithms  
89 previously available in the proprietary language MATLAB (MCQuant for quantifying marker  
90 intensities and computing morphology metrics<sup>17</sup>, and S3segmenter for watershed segmentation<sup>18</sup>,  
91 spot detection, and local thresholding). We also containerized several open-source algorithms  
92 (BaSiC<sup>19</sup> and Ilastik<sup>20</sup>), and incorporated three algorithms and associated deep learning models  
93 developed in our laboratories (UMAP/UnMicst, Coreograph and ASHLAR) (**Fig. 1A**; module names  
94 in red). All algorithms were tuned to manage very large files (~ 500GB/image) and containerized to  
95 abstract away language-specific dependencies (**Methods**). Source code, a user guide and other  
96 training materials are available via GitHub (<https://github.com/labsyspharm/mcmicro>).

97 To facilitate benchmarking, development of new algorithms and model training, we also  
98 generated a set of freely available reference images, the Exemplar Microscopy Images of Tissues and  
99 Tumors (EMIT). EMIT comprises multiplexed CyCIF images of a tissue microarray (TMA) with  
100 120 1.5 mm cores from 34 types of cancer, non-neoplastic diseases, and matched normal tissue  
101 (**Figure S1**, <https://synapse.org/EMIT>). EMIT images were processed using MCMICRO (using the  
102 Coreograph module) and all steps are documented on Synapse (<https://synapse.org/EMIT>).

103 Clustering of normal tissues and cancers by type (with some variance, because specimens came from  
104 different individuals) demonstrates that a wide range of specimens can be processed by MCMICRO  
105 to generate meaningful single cell data (**Figure S2**).

106 Processing multiplexed WSI data starts with acquisition of individual image tiles in a  
107 BioFormats-compatible format (level 1 data; **Fig. 1A**)<sup>21</sup>; each tile is typically a megapixel  
108 multichannel image, and as many as  $10^3$  tiles are required to cover a large tissue specimen at  
109 subcellular resolution. Tiles are corrected for uneven illumination, stitched together, and registered  
110 across channels to generate the first broadly useful type of data: a fully assembled, multichannel  
111 *mosaic image* in OME-TIFF format (a class of level 2 data) (**Fig. 1B**). In a large mosaic whole-slide  
112 image, length scales vary  $10^5$ -fold from the smallest resolvable feature to the largest dimension.  
113 Images are subjected to quality control, followed by segmentation. A segmentation mask (level 3  
114 data), the next object computed by MCMICRO, is available for human inspection in conjunction  
115 with underlying images to determine the quality of different segmentation approaches (**Fig. 1C**).

116 Following segmentation, the staining intensity in each channel is computed on a per-cell  
117 basis to generate a *Spatial Feature Table* (level 4 data), which is analogous to a count table in  
118 scRNAseq. In its simplest form, this table consists of the positions of cells and their integrated  
119 staining intensities in each imaging channel (morphological data, such as size, eccentricity etc. are  
120 additional table elements; **Fig. 1D**). The Spatial Feature Table can be visualized using tools designed  
121 for high dimensional data such as tSNE or UMAP, processed to identify cell types, and used for  
122 neighborhood or other types of analysis (**Fig. 1D**). It is also possible to skip segmentation altogether  
123 and perform analysis directly on images; pixel-level deep learning has already shown promise in  
124 clinical settings<sup>22,23</sup>, and many algorithms have been generalized for use with multiplexed data.  
125 Regardless of how data flows through MCMICRO, provenance is maintained by recording the  
126 identities, version numbers and parameter settings for each module, enabling full reproducibility  
127 (**Fig. S3**).

128 MCMICRO includes a newly developed tool for processing TMAs, which are widely used in  
129 research, because they enable parallel analysis of many specimens. In a TMA, a single slide carries  
130 dozens to hundreds of 0.3 to 2 mm diameter “cores”. The *Coreograph* module in MCMICRO is  
131 based on the U-Net deep learning architecture<sup>24</sup>. It finds the locations of individual cores and extracts  
132 each core as a separate, multi-channel image (**Fig. 1E**), allowing all cores to be processed in parallel  
133 by downstream modules. The robustness of the underlying neural network makes it possible for the  
134 module to accurately identify cores even in highly distorted TMAs.

135 Image processing requires user interaction and frequent visual review (see CellProfiler, for  
136 example<sup>25</sup>). To enable human-in-the-loop analysis, MCMICRO allows for training and parameter

137 adjustment to take place locally, using subsets of a large mosaic image. This iterative approach is  
138 particularly important for segmentation, since most contemporary algorithms rely on supervised  
139 machine learning. An absence of well-defined objective functions and ground truth data makes  
140 automated scoring of algorithms difficult, and different combinations of algorithms and models may  
141 be optimal for different tissues. MCMICRO therefore incorporates multiple segmentation algorithms  
142 (e.g., U-Net<sup>24</sup> or ilastik<sup>20</sup>), which can be executed in parallel and then compared (**Fig. 1C**).  
143 Additional improvement in segmentation can be achieved with the help of the EMIT data repository  
144 (<https://synapse.org/EMIT>), which includes a “classifier zoo” comprising a set of tissue-specific  
145 random forest segmentation models for ilastik. These models aid generation of robust tissue-specific  
146 segmentation masks and can also be subjected to further dataset-specific training.

147 To demonstrate the technology-agnostic capabilities of MCMICRO, we collected data from a  
148 single FFPE tonsil specimen at four different institutions using four imaging technologies: CODEX  
149 and CyCIF, which are immunofluorescence-based; mIHC, which uses multiplexed  
150 immunohistochemistry; and H&E staining (**Fig. 2A**). We also analyzed mxIF and publicly available  
151 Imaging Mass Cytometry (IMC) and MIBI data (**Table S2**). To show that MCMICRO does not have  
152 specific hardware dependencies, data processing was performed using cloud compute nodes  
153 provided either by Amazon Web Services (AWS) or the Google Cloud Platform and also using a  
154 Linux-based institutional cluster running the SLURM workload manager. MCMICRO provides  
155 detailed information on time, memory and CPU usage, making it straightforward to provision  
156 necessary computational resources (**Fig. S4**).

157 Image tiles from a variety of microscopes were subjected to stitching, registration and  
158 illumination correction using ASHLAR and BaSiC to generate mosaic level 2 image data that was  
159 visually inspected on a local workstation using napari and in the cloud using OMERO (**Fig. 2A**).  
160 Images were then segmented and staining intensities were computed on a per-cell basis using  
161 MCQuant. Cell types were visualized in the tissue context for epithelial cells of the tonsil mucosa  
162 (Keratin+/panCK+), cytotoxic T cells (CD8+) and B cells (CD20+) (**Fig. 2B**). Visual inspection of  
163 stitched and registered CyCIF, CODEX and mIHC images and derived data revealed accurate image  
164 stitching and registration, facilitating the creation of reasonable segmentation masks and the  
165 generation of correctly formatted Spatial Feature Tables. The results of cell type calling were similar  
166 (**Fig. 2C**), and when data from all three technologies were combined and visualized using tSNE,  
167 cells were separated by marker expression not imaging technology (**Fig 2D**). These findings  
168 demonstrate consistency in image acquisition and data processing.

169 A few algorithms in MCMICRO (e.g., Ashlar and BaSiC) are tissue and technology agnostic  
170 and can be used on diverse types of data with little, if any, tuning or modification. The performance

171 of other algorithms (e.g., UnMicst and Ilastik) is dependent on the properties of their learned models,  
172 which often work well for some tissues and not for others. MCMICRO facilitates identification of  
173 effective algorithms and models by executing different segmentation approaches in parallel,  
174 followed by comparison of the resulting masks. We expect continued innovation in the area of image  
175 segmentation, as well as addition of algorithms for automated quality control of images and  
176 identification of cell types based on marker intensities and cell morphologies. However, we do not  
177 anticipate that users will need to manage an endless proliferation of novel methods: multiple research  
178 consortia are actively working together on evaluation efforts (analogous to Dream Challenges<sup>26</sup>)  
179 aimed at creating best practices for highly-multiplexed image analysis. MCMICRO provides the  
180 technical foundation for such evaluations and for widespread distribution of the results. MCMICRO  
181 is also being used by the HTAN consortium to rigorously compare different image acquisition  
182 technologies.

183 In conclusion, the MCMICRO pipeline described here provides a foundation for community-  
184 wide development of FAIR (findable, accessible, interoperable and reusable)<sup>27</sup> workflows for  
185 analysis of large tissue images currently being generated by multiple international consortia and  
186 many individual laboratories. MCMICRO works with any acquisition technology that generates Bio-  
187 Formats/OME-compatible images, including the six technologies described above. The pipeline is  
188 based on widely accepted software standards and interoperates with any programming language  
189 through the use of software containers, making it easy to add new modules. The result is a user-  
190 friendly end-to-end pipeline that executes computation-intensive processes in the cloud, while  
191 enabling parameter optimization, training and quality control to be performed locally and  
192 interactively.

193



194 **METHODS**

195 Tissue samples

196 A de-identified tonsil specimen from a 4-year old Caucasian female was procured from the  
197 Cooperative Human Tissue Network (CHTN), Western Division, as part of the Human Tumor Atlas  
198 (HTAN) SARDANA trans-network project (TNP). Regulatory documents including Institutional  
199 Review Board (IRB) protocols, data use agreements and tissue use agreements were in place to  
200 ensure regulatory compliance. Standard protocols for tissue procurement and fixation were followed;  
201 a detailed protocol can be found at the link provided in Table 1. Sections were cut from a common  
202 formalin-fixed paraffin embedded (FFPE) block at a thickness of 5  $\mu\text{m}$  and mounted onto Superfrost  
203 Plus glass microscope slides (Fisher Scientific, 12-550-15) for CyCIF and mIHC or mounted on  
204 poly-L-Lysine (PLL) coated coverslips (Electron Microscopy Sciences, 72204-01; slides and FFPE  
205 sections prepared following instructions in the Akoya Biosciences CODEX User Manual Rev B.0,  
206 Chapter 3. Coverslip Preparation and Tissue Processing) for CODEX. A set of FFPE tissue sections  
207 was received by participating HTAN Centers, as indicated in **Table 1**, allowing Centers to generate a  
208 comparable spatial cell census using each Center's imaging method of choice. CHTN performed  
209 hematoxylin and eosin (H&E) staining on the first section which was subsequently imaged at  
210 Harvard Medical School (HMS).

211 For the EMIT dataset, human tissue specimens (from 42 patients) were used to construct a  
212 multi-tissue microarray (HTMA427) under an excess (discarded) tissue protocol approved by the  
213 IRB at Brigham and Women's Hospital (BWH IRB 2018P001627). Two 1.5 mm diameter cores  
214 were acquired from each of 60 tissue regions with the goal of acquiring one or two examples of as  
215 many tumors as possible (with matched normal tissue from the same resection when that was  
216 feasible), as a well several non-neoplastic medical diseases involving acute inflammation (e.g.  
217 diverticulitis and appendicitis), and secondary lymphoid tissues such as tonsil, spleen and lymph  
218 nodes. Overall, the TMA contained 120 cores plus 3 additional "marker cores," which are cores  
219 added to the TMA in a manner that makes it possible to orient the TMA in images.

220

221 **Table 1. Sample information.**

Section Number	Section Thickness ( $\mu\text{m}$ )	Center	Assay
WD-75684-01	5	Cooperative Human Tissue Network	H&E
WD-75684-02	5	Harvard Medical School	CyCIF
WD-75684-05	5	Broad Institute	CODEX

WD-75684-08	5	Harvard Medical School	CyCIF
WD-75684-12	5	Oregon Health & Science University	mIHC

222

223 CyCIF staining and imaging

224 The CyCIF method involves iterative cycles of antibody incubation, imaging and fluorophore  
 225 inactivation as described previously<sup>7</sup>. A detailed protocol can be found on protocols.io as shown in  
 226 **Table 2**. CyCIF images are 36-plex whole-slide images collected using a 20x magnification, 0.75  
 227 NA objective with 2 x 2 pixel binning, yielding images of pixel size at 0.65 μm/pixel. The image  
 228 comprises 416 and 350 image tiles for WD-75684-02 and WD-75684-08, respectively, each with  
 229 four channels, one of which is always Hoechst to stain DNA in the nucleus.

230

231 **Table 2.** List of protocols. As a part of the HTAN effort, all protocols and methods are deposited on  
 232 Protocols.io.

Category	Center	Protocols.io link
Protocol (Biospecimen)	CHTN	Tissue Procurement and Fixation in 10% NBF <a href="https://www.protocols.io/view/tissue-procurement-fixation-with-10-nbf-6y4hfyw">https://www.protocols.io/view/tissue-procurement-fixation-with-10-nbf-6y4hfyw</a>
Protocol (Characterization)	HMS	H&E <a href="https://www.protocols.io/view/tissue-procurement-fixation-with-10-nbf-6y4hfyw">wx.doi.org/10.17504/protocols.io.bsi8nchw</a>
Protocol (Characterization)	HMS	FFPE Tissue Pre-treatment Before t-CyCIF on Leica Bond RX <a href="https://www.protocols.io/view/ffpe-tissue-pre-treatment-before-t-cycif-on-leica-bji2kkge">https://www.protocols.io/view/ffpe-tissue-pre-treatment-before-t-cycif-on-leica-bji2kkge</a>
Protocol (Characterization)	HMS	Tissue Cyclic Immunofluorescence (t-CyCIF) <a href="https://www.protocols.io/view/tissue-cyclic-immunofluorescence-t-cycif-bjiukkew">https://www.protocols.io/view/tissue-cyclic-immunofluorescence-t-cycif-bjiukkew</a>
Protocol (Characterization)	Broad	CODEX <a href="https://www.protocols.io/private/FAD1B1BA64C011EB8A990A58A9FEAC2A/">https://www.protocols.io/private/FAD1B1BA64C011EB8A990A58A9FEAC2A/</a>
Protocol (Characterization)	OHSU	Multiplexed Immunohistochemistry (mIHC) <a href="https://www.protocols.io/view/mihc-staining-ohsu-coussens-lab-sop-tnp-sardana-bcdpis5n">https://www.protocols.io/view/mihc-staining-ohsu-coussens-lab-sop-tnp-sardana-bcdpis5n</a>

233

234 CODEX staining and imaging

235 Coverslips were prepared following the FFPE tissue staining protocols in the Akoya Biosciences  
 236 CODEX User Manual (Sections 5.4 – 5.6). Briefly, 5 μm FFPE tissue sections were cut onto PLL-



237 coated coverslips and baked for 20-25 minutes at 55 °C. Sections were cooled briefly before  
238 deparaffinization and were washed for 5 minutes each as follows: twice in xylene, twice in 100%  
239 ethanol, once in 90%, 70%, 50%, and 30% ethanol, and twice in deionized water. Sections were  
240 moved to 1x Citrate Buffer (Vector Laboratories, H-3300) and antigen retrieval was performed in a  
241 Tinto Retriever Pressure Cooker (BioSB, BSB 7008) at high pressure for 20 minutes. Sections were  
242 briefly washed in deionized water before being left to incubate in deionized water at room  
243 temperature for 10 minutes. Sections were briefly washed twice in Hydration Buffer (Akoya), then  
244 were left to incubate in Staining Buffer (Akoya) at room temperature for 20-30 minutes. 200  
245 µL/section of Antibody Cocktail was prepared according to manufacturer instructions. Sections were  
246 covered with the 200 µL Antibody Cocktail and left to incubate at room temperature for 3 hours in a  
247 humidity chamber. Sections were washed twice in Staining Buffer for 2 minutes, and then fixed with  
248 a mixture of 1.6% PFA in Storage Buffer (Akoya) for 10 minutes. Sections were briefly washed  
249 three times in 1x PBS, and then washed in ice-cold methanol for 5 minutes before being washed  
250 again three times in 1x PBS. Sections were stained with 190 µL of a mixture of 20 µL Fixative  
251 Reagent (Akoya) and 1 mL 1x PBS, after which they were left to incubate at room temperature for  
252 20 minutes. Sections were briefly washed three times in 1X PBS and were stored in Storage Buffer  
253 at 4 °C until the assay was ready to be run.

254

#### 255 Running the CODEX Assay

256 A 96-well plate of reporter stains with Nuclear Stain (Akoya) was prepared according to Akoya  
257 Biosciences CODEX User Manual (Sections 7.1 – 7.2). Stained Tissue sections were loaded onto the  
258 CODEX Stage Insert (Akoya) and the Reporter Plate was loaded into the CODEX Machine. The on-  
259 screen prompts were followed and the section was manually stained with a 1:2000 Nuclear Stain in  
260 1x CODEX Buffer (Akoya) for 5 minutes before proceeding with following the on-screen prompts.  
261 Imaging was performed on a Zeiss Axio Observer with Colibri 7 light source. Emission filters were  
262 BP 450/40, BP 550/100, BP 525/50, BP 630/75, BP 647/70, BP 690/50, and TBP 425/29 + 514/31 +  
263 632/100 and dichroic mirrors were QBS 405 + 492 + 575 + 653, TFT 450 + 520 + 605, TFT 395 +  
264 495 + 610, and TBS 405 + 493 + 575, all from Zeiss. Overview scans were performed at 10x  
265 magnification, after which 5 x 5 field of view regions were acquired using a Plan-Apochromat  
266 20x/0.8 M27 Air objective (Zeiss, 420650-9902-000). 20x magnification images were acquired with  
267 a 212 x 212 nm pixel size using software autofocus repeated every tile before acquiring a 17 plane z-  
268 stack with 0.49 µm spacing. Tiles were stitched using a 10% overlap.

269

270

271 mIHC staining and imaging

272 The multiplex immunohistochemistry (mIHC) platform described herein involves wet and dry-lab  
273 techniques that have been robustly developed to interrogate the tumor immune microenvironment in  
274 situ. mIHC involves a cyclic staining process optimized for FFPE tissues with panels of antibodies  
275 (12-29 per panel) designed to interrogate both lymphoid and myeloid compartments of the immune  
276 system as well as cellular functional states, as previously described<sup>28,29</sup>.

277  
278 Pipeline implementation

279 MCMICRO was implemented in Nextflow, which was chosen for its natural integration with  
280 container technologies such as Docker and Singularity, its automatic provenance tracking and  
281 parallelization of image processing tasks, and its ability to specify module dependencies that may  
282 change at runtime<sup>15</sup>.

283  
284 Illumination correction

285 BaSiC is a Fiji / ImageJ plugin for background and shading correction, producing high accuracy  
286 while requiring only a few input images<sup>19</sup>. We containerized the tool, allowing it to be executed  
287 without an explicit installation of ImageJ.

288  
289 Image stitching and registration using Ashlar

290 Cycle-based highly multiplexed microscopy produces multi-channel images of fixed cells using a  
291 standard four/five-color microscope. Registration of the images across successive cycles is made  
292 straightforward by the addition of a nuclear counterstain in every cycle. Given a set of slightly  
293 overlapping images covering a tissue, we correct for mechanical stage positioning error intrinsic to  
294 all microscopes using Ashlar (Alignment by Simultaneous Harmonization of Layer/Adjacency  
295 Registration), a Python package for efficient mosaicing and registration of highly multiplexed  
296 imagery<sup>30</sup>. The overall strategy of Ashlar is as follows: (i) align tile images from the first cycle edge-  
297 to-edge with their nearest neighbors (mosaicing) using phase correlation on the nuclear marker  
298 channel; (ii) for the second and subsequent cycles, align each tile to the greatest overlapping tile  
299 from the first cycle (registration), using phase correlation on the nuclear marker channel, and retain  
300 the corrected stage coordinates, rather than the actual merged images; (iii) use the corrected  
301 coordinates to assemble a single image covering the entire imaged area, including all channels from  
302 all cycles. This approach minimizes the compounding of alignment errors across tiles and cycles as  
303 well as temporary storage requirements for intermediate results.

304

### 305 Coreograph

306 Coreograph's function is to split, or 'dearray', a stitched TMA image into separate image stacks per  
307 core. It employs a semantic segmentation preprocessing step to assist with identifying cores that are  
308 dimmed or fragmented, which is a common issue. We trained a deep, fully connected network on  
309 two classes – core tissue and background – using the popular UNet<sup>24</sup> architecture for semantic  
310 segmentation. Training data consisted of cores that were well-separated, as well as cores that were  
311 merged and/or fragmented, which allowed for handling situations where sample integrity was highly  
312 heterogeneous. Once cores had been accentuated in the form of probability maps, they were cropped  
313 from the stitched image based on their median diameter and saved as a TIFF stack. In situations  
314 where the cores were too clumped, the median diameter was used to set the size of a Laplacian of  
315 Gaussian (LoG) kernel in order to identify local maxima from the probability maps.

### 316 317 UnMicst (U-Net model for identifying cells and segmenting tissue)

318 UnMicst is a preprocessing module in MCMICRO that aids in improving downstream segmentation  
319 accuracy by generating per-class probability maps to classify each pixel with a certain amount of  
320 confidence. Analogous to Coreograph, it employs a UNet architecture (see above). Previously, a  
321 similar UNet model was trained for nuclei segmentation to recognize two classes in Hoechst 33342 -  
322 stained tonsil tissue (nuclei contours and background). Here, we train a 3-class model to extract  
323 nuclei centers, nuclei contours, and background from manually annotated lung, tonsil, prostate and  
324 other tissues in order to ascribe a variety of nuclei shapes. Realistic augmentations, in addition to  
325 conventional on-the-fly transformations, were included by deliberately defocusing the image and  
326 increasing the exposure time of the camera to simulate focus and contrast augmentations,  
327 respectively. Training was performed using a batch size of 24 with the Adam Optimizer and a  
328 learning rate of 0.00003 until the accuracy converged. Segmentation accuracy was estimated by  
329 counting the fraction of cells in a held out test set that passed a sweeping Intersection of Union  
330 (IOU) threshold.

### 331 332 Ilastik tissue segmentation

333 Similar to UnMicst, Ilastik assigns each pixel a probability of belonging to predetermined classes  
334 (e.g., cell nucleus, membrane, background). MCMICRO relies on Ilastik's pixel classification  
335 module for training and subsequent batch-processing using a random forest classifier. Ilastik  
336 classifier training in MCMICRO is completed in several steps. First, regions of interest (ROIs) with a  
337 user-defined width and height are randomly cropped from the WSI. Second, the ROIs are manually  
338 annotated by the user on a local machine via Ilastik's graphical user interface (GUI). Third, to ensure

339 tissue portions are accurately represented in cropped images, Otsu's method is used to identify a  
340 global threshold across the WSI for a particular channel of interest (e.g., nuclear staining). Finally,  
341 the user exports the cropped sections that contain the desired proportion of pixels above the  
342 previously determined threshold. Upon completion of the random forest training, whole slide  
343 classifier predictions are deployed in headless mode (no GUI) for batch processing of large data sets  
344 within MCMICRO.

345

#### 346 Watershed segmentation via S3segmenter

347 We implemented S3segmenter, a custom marker-controlled watershed algorithm to identify nuclei  
348 from the probability maps generated by UnMicst and Ilastik. Watershed markers are obtained by  
349 convolving a LoG kernel, followed by a local maxima search across the image to identify seed  
350 points. The size of the LoG kernel and local maxima compression are tunable parameters dependent  
351 on the expected nuclei diameters in the image. As a byproduct, this method identifies false positive  
352 segments in the image background. These false positives were excluded by comparing their  
353 intensities to an Otsu-derived threshold calculated either on the raw image or on the probability map.  
354 S3segmenter currently offers three alternative methods for cytoplasm segmentation. First, traditional  
355 nonoverlapping rings (annuli) with user-defined radius are used around each nucleus. Second, a  
356 Euclidean distance transform is computed around each nucleus and masked with a user-specified  
357 channel, reflecting the overall shape of the whole tissue sample. An autofluorescence channel can be  
358 chosen if the signal-to-image background ratio is sufficiently high. Third, the cytoplasm is  
359 segmented using a marker-controlled watershed on the grayscale-weighted distance transform, where  
360 the segmented nuclei are markers and the grayscale-weighted distance transform is approximated by  
361 adding scaled versions of the distance transform and raw image together. This method is  
362 conceptually similar to that found in the CellProfiler Identify Secondary Objects module<sup>25</sup>.  
363 S3segmenter is also capable of detecting puncta by convolving a small LoG kernel across the image  
364 and identifying local maxima. Once nuclei and cytoplasm segmentation are complete, labelled masks  
365 for each region are exported as 32-bit tiff images. Two channel tiff stacks consisting of the mask  
366 outlines and raw image are also saved so that segmentation accuracy can be easily visually assessed.

367

#### 368 MCQuant

369 Semantic segmentation in MCMICRO produces 32-bit masks, which are used to quantify pixel  
370 intensity (i.e., protein expression) on multiplexed WSI for cytoplasm and nuclei. Quantification in  
371 MCMICRO is carried out using scikit-image, a popular Python-based image analysis library, and  
372 values of cellular spatial features are calculated for unique cells (cytoplasm and nuclei), in addition

373 to their mean pixel intensity (protein expression). The resulting spatial feature tables are exported as  
374 CSV files for subsequent data analysis analogous to histoCAT<sup>17</sup>, which is implemented in  
375 MATLAB.

376  
377 Data availability statement  
378 All software and code that produced the findings of the study, including all main and supplemental  
379 figures, are available at <https://github.com/labsyspharm/mcmicro>.

380 All EMIT images are available at <https://synapse.org/EMIT> and all exemplar and tonsil images are  
381 available at [https://synapse.org/MCMICRO\\_images](https://synapse.org/MCMICRO_images).

382

### 383 **ACKNOWLEDGEMENTS**

384 This work was funded by NIH grants U54-CA225088 and U2C-CA233262 to P.K.S. and S.S and by  
385 the Ludwig Cancer Center at Harvard. D.S. was funded by an Early Postdoc Mobility fellowship (no.  
386 P2ZHP3\_181475) from the Swiss National Science Foundation and is a Damon Runyon Fellow  
387 supported by the Damon Runyon Cancer Research Foundation (DRQ-03-20). Z.M. is supported by  
388 NCI grant R50-CA252138. We thank Dana-Farber/Harvard Cancer Center in Boston, MA, for the  
389 use of the Specialized Histopathology Core, which provided TMA construction and sectioning  
390 services. Dana-Farber/Harvard Cancer Center is supported in part by an NCI Cancer Center Support  
391 Grant P30 CA06516. Tissue samples were provided by the NCI Cooperative Human Tissue Network  
392 (CHTN).

### 393 **OUTSIDE INTERESTS**

394 PKS is a member of the SAB or BOD member of Applied Biomath, RareCyte Inc., and Glencoe  
395 Software, which distributes a commercial version of the OMERO database; PKS is also a member of  
396 the NanoString SAB. In the last five years the Sorger lab has received research funding from  
397 Novartis and Merck. Sorger declares that none of these relationships have influenced the content of  
398 this manuscript. SS is a consultant for RareCyte Inc. A. R. is a cofounder and equity holder of  
399 Celsius Therapeutics, an equity holder in Immunitas and, until 31 July 2020, was an SAB member of  
400 Thermo Fisher Scientific, Syros Pharmaceuticals, Neogene Therapeutics and Asimov. Since August  
401 1, 2020, A. R. has been an employee of Genentech. The other authors declare no outside interests.

402 **REFERENCES**

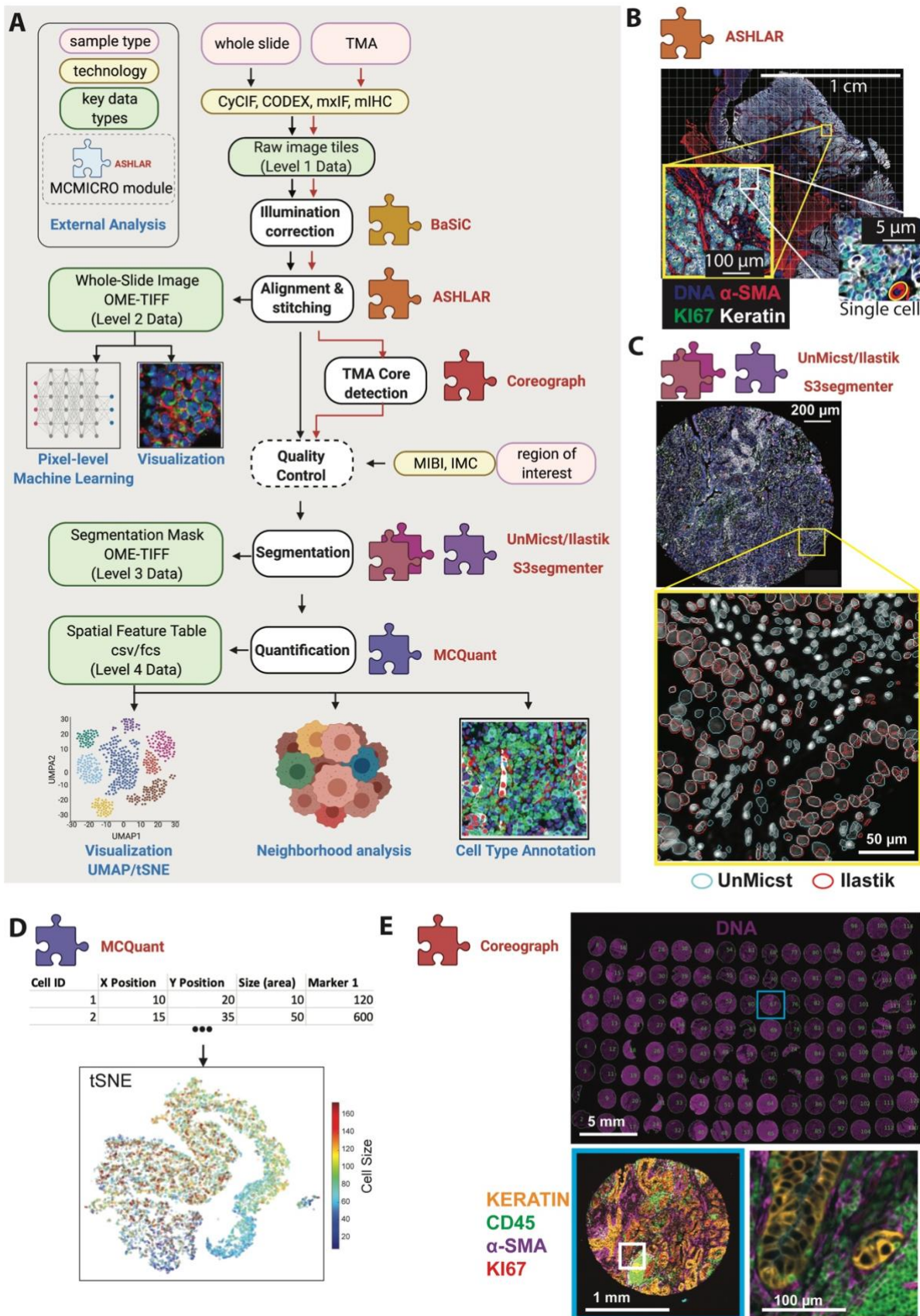
- 403 1. Keren, L. *et al.* A Structured Tumor-Immune Microenvironment in Triple Negative Breast Cancer  
404 Revealed by Multiplexed Ion Beam Imaging. *Cell* **174**, 1373-1387.e19 (2018).
- 405 2. CRUK IMAXT Grand Challenge Team *et al.* Imaging mass cytometry and multiplatform  
406 genomics define the phenogenomic landscape of breast cancer. *Nat. Cancer* **1**, 163–175 (2020).
- 407 3. Jackson, H. W. *et al.* The single-cell pathology landscape of breast cancer. *Nature* **578**, 615–620  
408 (2020).
- 409 4. Schürch, C. M. *et al.* Coordinated Cellular Neighborhoods Orchestrate Antitumoral Immunity at  
410 the Colorectal Cancer Invasive Front. *Cell* **182**, 1341-1359.e19 (2020).
- 411 5. Abels, E. & Pantanowitz, L. Current state of the regulatory trajectory for whole slide imaging  
412 devices in the USA. *J. Pathol. Inform.* **8**, 23 (2017).
- 413 6. Evans, A. J. *et al.* US Food and Drug Administration Approval of Whole Slide Imaging for  
414 Primary Diagnosis: A Key Milestone Is Reached and New Questions Are Raised. *Arch. Pathol.*  
415 *Lab. Med.* **142**, 1383–1387 (2018).
- 416 7. Lin, J.-R. *et al.* Highly multiplexed immunofluorescence imaging of human tissues and tumors  
417 using t-CyCIF and conventional optical microscopes. *eLife* **7**, e31657 (2018).
- 418 8. Vickovic, S. *et al.* High-definition spatial transcriptomics for in situ tissue profiling. *Nat. Methods*  
419 **16**, 987–990 (2019).
- 420 9. Rodrigues, S. G. *et al.* Slide-seq: A scalable technology for measuring genome-wide expression at  
421 high spatial resolution. *Science* **363**, 1463–1467 (2019).
- 422 10. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. Spatially resolved,  
423 highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090–aaa6090 (2015).
- 424 11. Li, B. *et al.* Cumulus provides cloud-based data analysis for large-scale single-cell and  
425 single-nucleus RNA-seq. *Nat. Methods* **17**, 793–798 (2020).



- 426 12. Merkel, D. Docker: Lightweight Linux Containers for Consistent Development and  
427 Deployment. *Linux J* **2014**, (2014).
- 428 13. Common Workflow Language (CWL) Workflow Description, v1.0.2.  
429 <https://www.commonwl.org/v1.0/Workflow.html>.
- 430 14. Rozenblatt-Rosen, O. *et al.* The Human Tumor Atlas Network: Charting Tumor Transitions  
431 across Space and Time at Single-Cell Resolution. *Cell* **181**, 236–249 (2020).
- 432 15. Di Tommaso, P. *et al.* Nextflow enables reproducible computational workflows. *Nat.*  
433 *Biotechnol.* **35**, 316–319 (2017).
- 434 16. Kurtzer, G. M., Sochat, V. & Bauer, M. W. Singularity: Scientific containers for mobility of  
435 compute. *PLOS ONE* **12**, e0177459 (2017).
- 436 17. Schapiro, D. *et al.* histoCAT: analysis of cell phenotypes and interactions in multiplex image  
437 cytometry data. *Nat. Methods* **14**, 873–876 (2017).
- 438 18. Saka, S. K. *et al.* Immuno-SABER enables highly multiplexed and amplified protein imaging  
439 in tissues. *Nat. Biotechnol.* **37**, 1080–1090 (2019).
- 440 19. Peng, T. *et al.* A BaSiC tool for background and shading correction of optical microscopy  
441 images. *Nat. Commun.* **8**, 14836 (2017).
- 442 20. Berg, S. *et al.* ilastik: interactive machine learning for (bio)image analysis. *Nat. Methods* **16**,  
443 1226–1232 (2019).
- 444 21. Linkert, M. *et al.* Metadata matters: access to image data in the real world. *J. Cell Biol.* **189**,  
445 777–782 (2010).
- 446 22. Iizuka, O. *et al.* Deep Learning Models for Histopathological Classification of Gastric and  
447 Colonic Epithelial Tumours. *Sci. Rep.* **10**, 1504 (2020).
- 448 23. Campanella, G. *et al.* Clinical-grade computational pathology using weakly supervised deep  
449 learning on whole slide images. *Nat. Med.* **25**, 1301–1309 (2019).

- 450 24. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional Networks for Biomedical  
451 Image Segmentation. in *Medical Image Computing and Computer-Assisted Intervention –*  
452 *MICCAI 2015* (eds. Navab, N., Hornegger, J., Wells, W. M. & Frangi, A. F.) vol. 9351 234–241  
453 (Springer International Publishing, 2015).
- 454 25. McQuin, C. *et al.* CellProfiler 3.0: Next-generation image processing for biology. *PLOS Biol.*  
455 **16**, e2005970 (2018).
- 456 26. Stolovitzky, G., Prill, R. J. & Califano, A. Lessons from the DREAM2 Challenges. *Ann. N.*  
457 *Y. Acad. Sci.* **1158**, 159–195 (2009).
- 458 27. Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and  
459 stewardship. *Sci. Data* **3**, 160018 (2016).
- 460 28. Tsujikawa, T. *et al.* Quantitative Multiplex Immunohistochemistry Reveals Myeloid-  
461 Inflamed Tumor-Immune Complexity Associated with Poor Prognosis. *Cell Rep.* **19**, 203–217  
462 (2017).
- 463 29. Banik, G. *et al.* High-dimensional multiplexed immunohistochemical characterization of  
464 immune contexture in human cancers. *Methods Enzymol.* **635**, 1–20 (2020).
- 465 30. *ASHLAR*.  
466 <https://github.com/labsyspharm/ashlar>.  
467

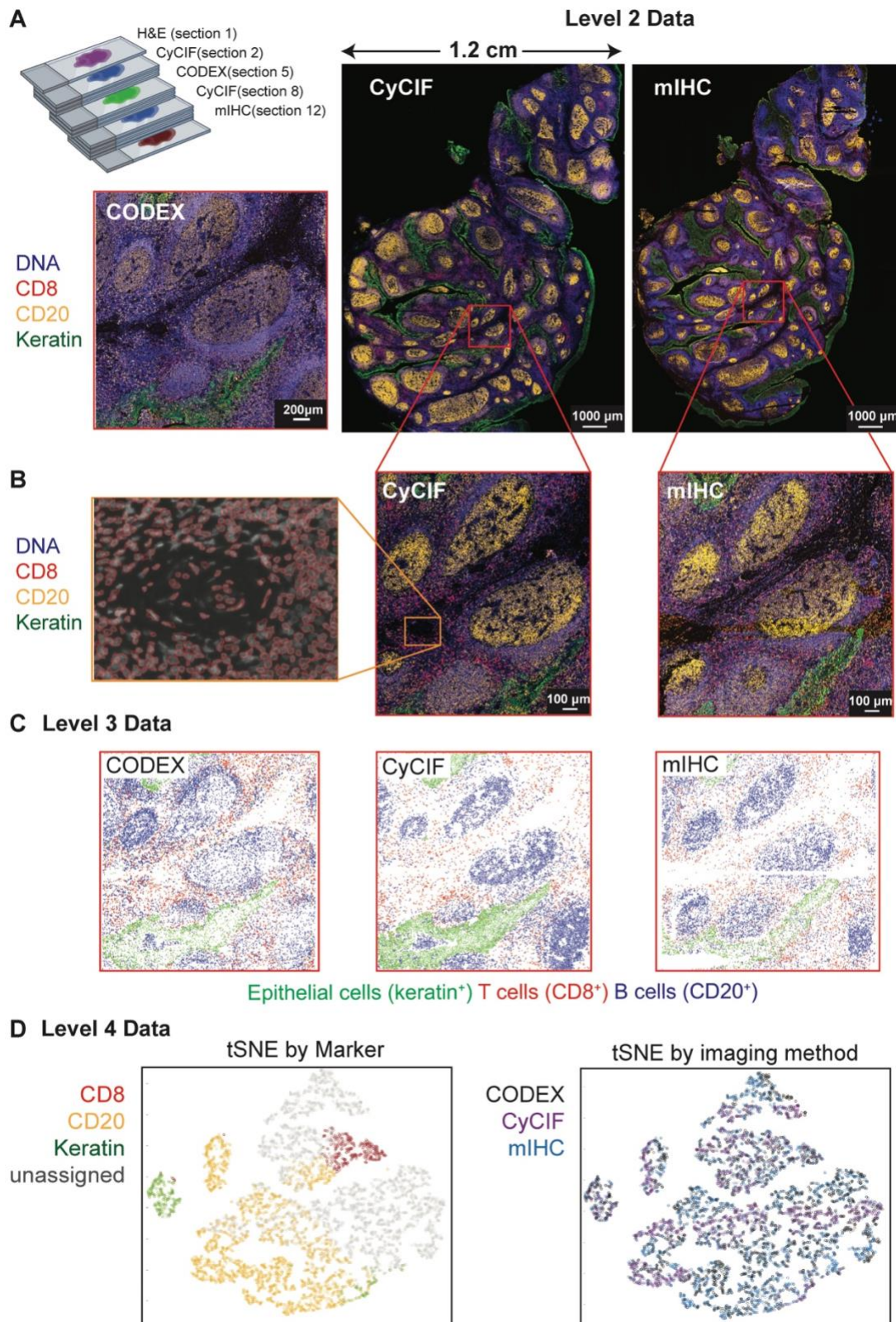
468 **FIGURE 1**



470 **Fig. 1: MCMICRO pipeline overview. Modules highlighted in bold red are developed and/or**  
471 **containerized in-house. (A)** A schematic representation of a canonical workflow for end-to-end  
472 image processing of multiplexed whole-slide and TMA using MCMICRO. Shown is a flow of inputs  
473 (pink rectangles) from imaging technologies (yellow rectangles) through image processing steps  
474 (white rectangles) that are implemented in software modules (puzzle pieces) to produce key data  
475 types (green rectangles). Data flows associated with the whole slide and TMA are represented with  
476 black and red arrows, respectively. Quality control is highlighted with a dashed border. **(B-E)**  
477 Highlights of individual software modules incorporated into MCMICRO. **B** ASHLAR is used to  
478 stitch and register individual CyCIF image tiles with subcellular accuracy (yellow zoom-in). This  
479 panel depicts a 484 tile (22 x 22) t-CyCIF, whole-slide, mosaic image of a human colorectal cancer  
480 in four channels: Hoechst 33342-stained nuclear DNA (blue),  $\alpha$ -smooth muscle actin ( $\alpha$ -SMA; red),  
481 the Ki-67 proliferation marker (green) and cytokeratin (white). An interactive on-line visualization of  
482 these data can be found at: <https://www.cycif.org/data/tnp-2020/osd-crc-case-1-ffpe-cycif-stack>. **C**  
483 Two different segmentation masks computed by UnMicst (blue) and Ilastik (red) overlaid on an  
484 image of nuclei from an EMIT TMA core. **D** A schematic of the first rows and columns of a Spatial  
485 Feature Table used for visualization using tSNE. **E** A CyCIF image of an EMIT TMA de-arrayed  
486 using Coreograph to identify individual cores, which are subsequential extracted and analyzed in  
487 high-resolution. Below, a five-color image of a single lung adenocarcinoma core is shown for  
488 channels corresponding to Hoechst 33342-stained DNA (white), cytokeratin (orange), the immune-  
489 cell marker CD45 (green),  $\alpha$ -SMA (magenta) and Ki-67 (red)).  
490



491 **FIGURE 2:**

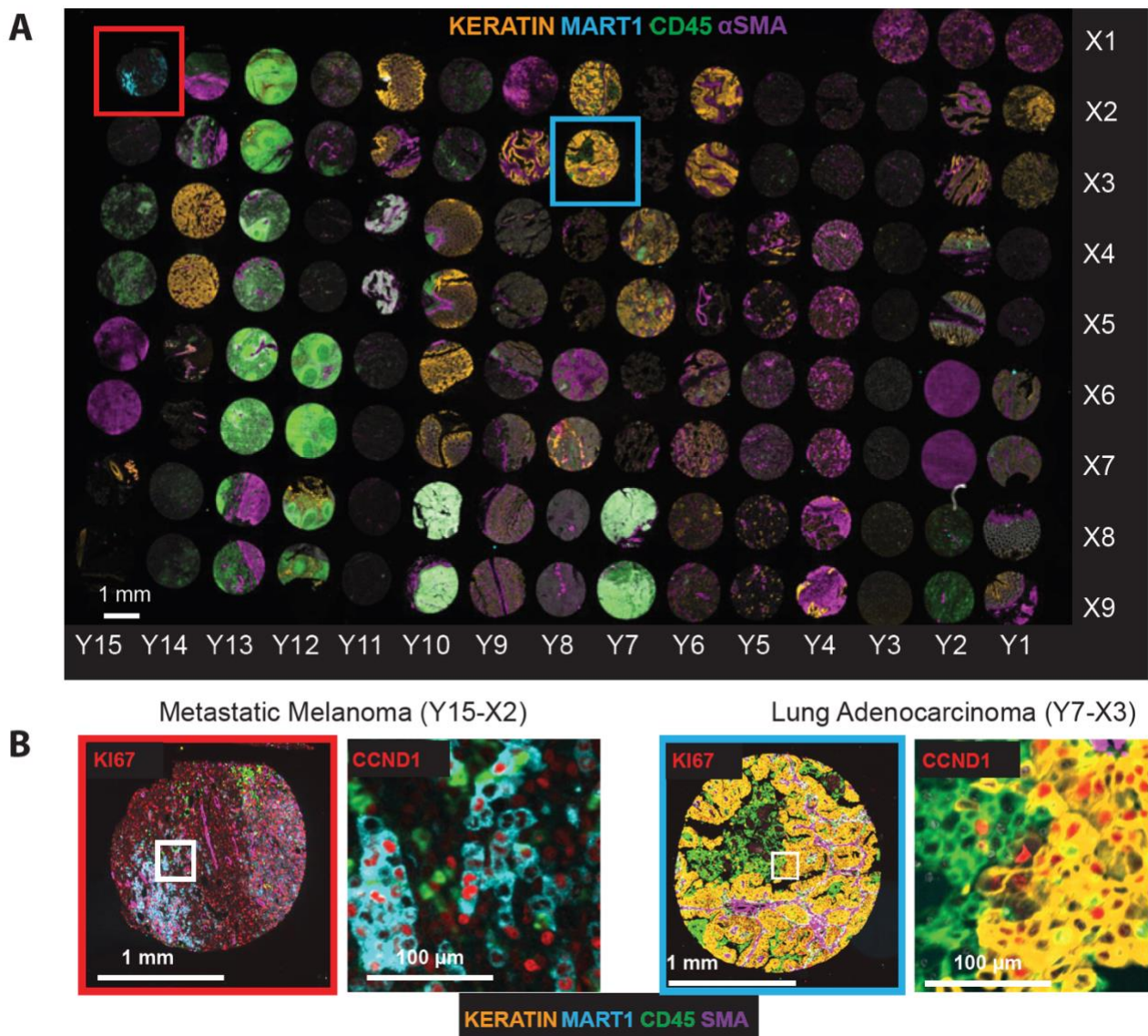


493 **Fig. 2: Comparison of images of human tonsil collected using three different technologies and**  
494 **processed using MCMICRO. A.** Serial sections of a single tissue block imaged using H&E, 11-  
495 marker CODEX, 27-marker CyCIF, and 16-marker mIHC. The sectioning plan shows the position of  
496 each 5 µm section within the block: H&E section 1, CyCIF section 2, CODEX section 5 and mIHC  
497 section 12. Images show selected channels as follows: Hoechst 33342 (blue), CD20 (orange),  
498 Keratin (green), and CD8 (red). The CODEX image shows only a specific region (red border) of the  
499 specimen visible in whole-slide images to the right. **B.** Higher magnification images of the data  
500 above highlighting individual cells and segmentation masks generated with UnMicst. **C.** Centroids of  
501 the single cell mask for CODEX, CyCIF and mIHC are colored by marker expression to identify cell  
502 types. Epithelial cells of the tonsil mucosa stain positive for pan-cytokeratin (green), cytotoxic T  
503 cells stain positive for CD8 (red); and B cells stain positive for CD20 (blue). **D.** t-SNE of combined  
504 CODEX, CyCIF and mIHC data demonstrating clustering by marker expression (left) but not  
505 imaging technology (right).  
506



507 **FIGURE S1**

## Figure S1



508

509 **Figure S1. The EMIT dataset spanning 123 tissue cores across 34 cancer, non-neoplastic**

510 **diseases, and normal tissue type. A.** CyCIF whole slide image of EMIT visualizing Hoechst

511 33342-stained nuclear DNA (white), Keratin (orange), MART1 (cyan), CD45 (green) and SMA

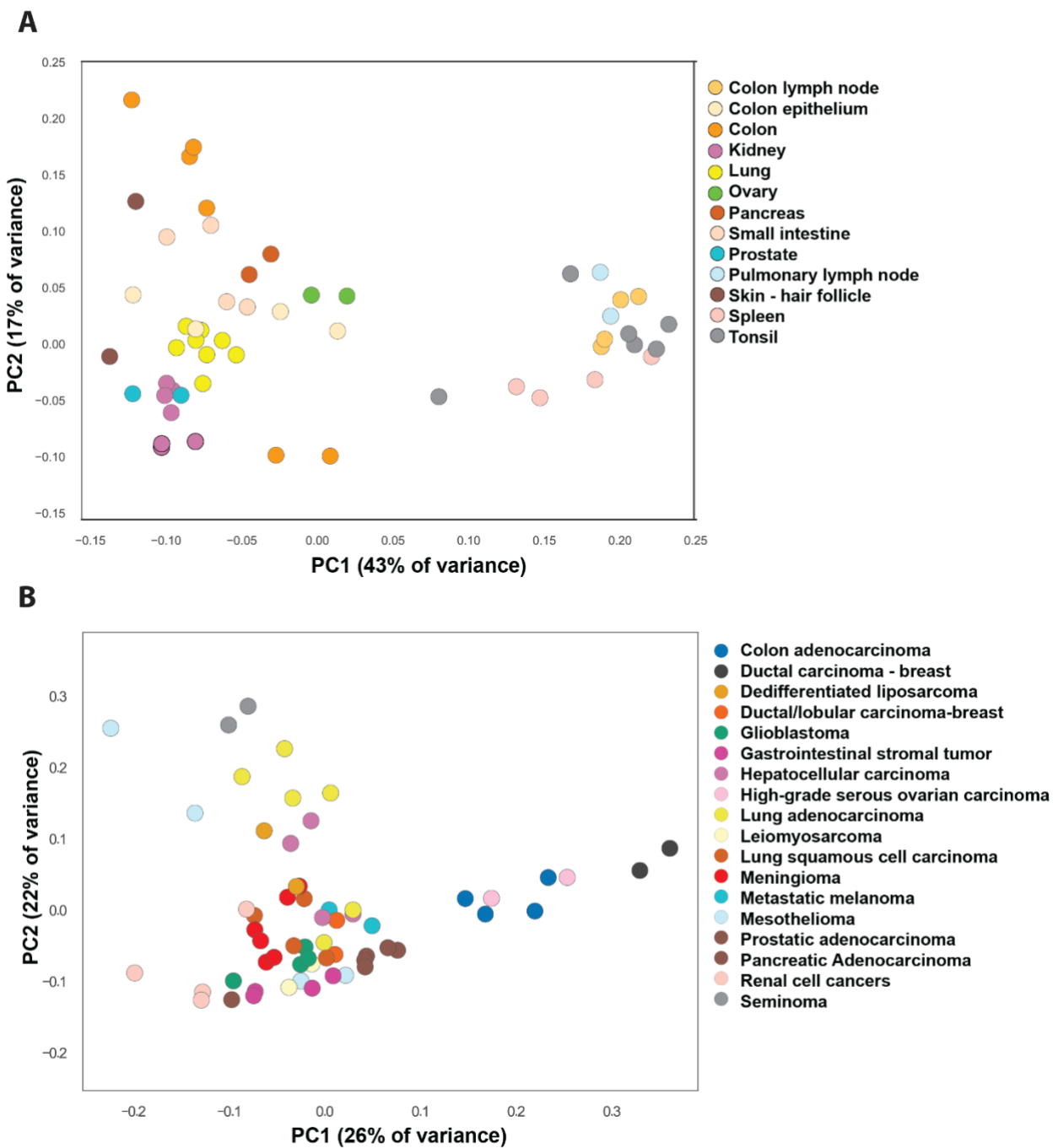
512 (purple). **B.** A zoom-in view of a metastatic melanoma (left, red box) and a lung adenocarcinoma

513 (right, blue box) core. The highest zoom level is highlighted with white boxes in the corresponding

514 low magnification images.

515

## Figure S2



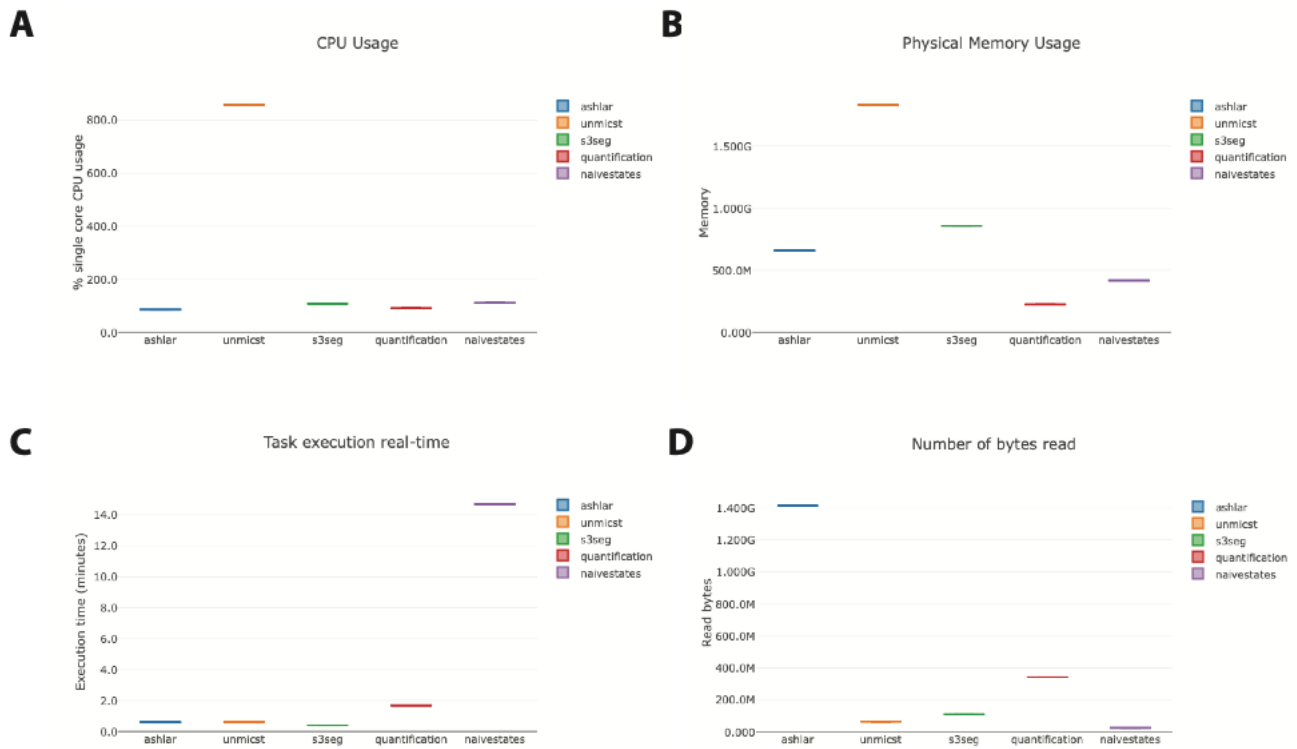
516

517 **Figure S2. Principal component analysis (PCA) of Spatial Feature Tables derived from EMIT**  
 518 **images. A.** represents normal tissues and **B.** cancer tissues. Independent cores cluster to a substantial  
 519 degree by tissue or cancer type; some variation is expected because tumors had different grades and  
 520 derive from different individuals. Data from the following antibodies was used to generate the data:  
 521 CD73, MART1, KI67, pan-cytokeratin, CD45, ECAD,  $\alpha$ -SMA, CD32, CDKN1A, CCNA2,  
 522 CDKN1C, CDKN1B, CCND1, cPARP, CCNB1, PCNA and CDK2.

523



533 **FIGURE S4**



534

535

536 **Figure S4. Detailed insight into the computational resources required by each module,**

537 **generated by Nextflow.** The data is viewed as an interactive browser-based report. **A.** CPU usage is

538 recorded as either % single core CPU usage (visualized) or % CPUs allocated. **B.** Physical memory

539 usage is recorded as either RAM only (visualized), RAM + Disk swap or % RAM allocated. **C.** Job

540 duration is recorded as either execution time (visualized) or % time allocated. **D.** Input/Output (I/O)

541 records both read (visualized) and written bytes.

542

543 **Table S1: Highly multiplexed imaging methods**

Non-cyclic metal-based	Cyclic fluorescence imaging		Cyclic immune-histochemistry
	cyclic-stained	single-stained	
Imaging Mass Cytometry (IMC) <sup>1</sup>	CODEX <sup>2</sup>	MxIF <sup>3</sup>	mIHC <sup>4</sup>
Multiplexed Ion Beam Imaging (MIBI) <sup>5</sup>	Immuno-SABER <sup>6</sup>	CyCIF <sup>7</sup>	MICSSS <sup>8</sup>
		4i <sup>9</sup>	

544

545 **Table S1:** Orange labeled methods were successfully processed by MCMICRO on publicly available  
546 datasets. Green labeled methods are additionally tested on images unique to this study with detailed  
547 description in the documentation.

548

549 **Table S2: Available open-source tools for image processing, analysis and visualization**

550

Software	Scalable	Whole slide processing	Stitching and registration	Modular	Segmentation	Analysis	GUI
MCMICRO	Yes	Yes	Yes	Yes	Yes	Yes	No
Cytokit <sup>10</sup>	Yes	No (tiles)	No	No	Yes	Yes	Yes
starfish <sup>11</sup>	Yes	No (tiles)	No	Yes	Yes	Yes	No
histoCAT <sup>12</sup>	No	Yes	No	Yes	No	Yes	Yes
QuPath <sup>13</sup>	No	Yes	No	No	Yes	Yes	Yes
CytoMAP <sup>14</sup>	No	Yes	No	No	No	Yes	Yes
Facetto <sup>15</sup>	No	Yes	No	No	No	Yes	Yes
napari <sup>16</sup>	Visualization tool						
OMERO <sup>17</sup>	Visualization tool						
Minerva <sup>18</sup>	Visualization tool						

551

552 **Table S2: List of open-source tools available for highly multiplexed image processing.**

553

554 **Supplementary table references:**

- 555 1. Giesen, C., Wang, H. A. O., Schapiro, D., Zivanovic, N., Jacobs, A., Hattendorf, B., Schüffler, P. J., Grolimund, D., Buhmann, J. M., Brandt, S.,  
556 Varga, Z., Wild, P. J., Günther, D. & Bodenmiller, B. Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry.  
557 *Nat Methods* **11**, 417–422 (2014).
- 558 2. Goltsev, Y., Samusik, N., Kennedy-Darling, J., Bhate, S., Hale, M., Vazquez, G., Black, S. & Nolan, G. P. Deep Profiling of Mouse Splenic  
559 Architecture with CODEX Multiplexed Imaging. *Cell* **174**, 968–981.e15 (2018).
- 560 3. Gerdes, M. J., Sevinsky, C. J., Sood, A., Adak, S., Bello, M. O., Bordwell, A., Can, A., Corwin, A., Dinn, S., Filkins, R. J., Hollman, D., Kamath,  
561 V., Kaanumalle, S., Kenny, K., Larsen, M., Lazare, M., Li, Q., Lowes, C., McCulloch, C. C., McDonough, E., Montalto, M. C., Pang, Z., Rittscher,  
562 J., Santamaria-Pang, A., Sarachan, B. D., Seel, M. L., Seppo, A., Shaikh, K., Sui, Y., Zhang, J. & Ginty, F. Highly multiplexed single-cell analysis  
563 of formalin-fixed, paraffin-embedded cancer tissue. *PNAS* **110**, 11982–11987 (2013).
- 564 4. Tsujikawa, T., Kumar, S., Borkar, R. N., Azimi, V., Thibault, G., Chang, Y. H., Balter, A., Kawashima, R., Choe, G., Sauer, D., El Rassi, E.,  
565 Clayburgh, D. R., Kulesz-Martin, M. F., Lutz, E. R., Zheng, L., Jaffee, E. M., Leyshock, P., Margolin, A. A., Mori, M., Gray, J. W., Flint, P. W. &  
566 Coussens, L. M. Quantitative Multiplex Immunohistochemistry Reveals Myeloid-Inflamed Tumor-Immune Complexity Associated with Poor  
567 Prognosis. *Cell Reports* **19**, 203–217 (2017).
- 568 5. Angelo, M., Bendall, S. C., Finck, R., Hale, M. B., Hitzman, C., Borowsky, A. D., Levenson, R. M., Lowe, J. B., Liu, S. D., Zhao, S., Natkunam,  
569 Y. & Nolan, G. P. Multiplexed ion beam imaging (MIBI) of human breast tumors. *Nat Med* **20**, 436–442 (2014).
- 570 6. Saka, S. K., Wang, Y., Kishi, J. Y., Zhu, A., Zeng, Y., Xie, W., Kirli, K., Yapp, C., Cicconet, M., Beliveau, B. J., Lapan, S. W., Yin, S., Lin, M.,  
571 Boyden, E. S., Kaeser, P. S., Pihan, G., Church, G. M. & Yin, P. Immuno-SABER enables highly multiplexed and amplified protein imaging in  
572 tissues. *Nat Biotechnol* **37**, 1080–1090 (2019).
- 573 7. Lin, J.-R., Izar, B., Wang, S., Yapp, C., Mei, S., Shah, P. M., Santagata, S. & Sorger, P. K. Highly multiplexed immunofluorescence imaging of  
574 human tissues and tumors using t-CyCIF and conventional optical microscopes. *eLife* **7**, e31657 (2018).
- 575 8. Remark, R., Merghoub, T., Grabe, N., Litjens, G., Damotte, D., Wolchok, J. D., Merad, M. & Gnjatic, S. In-depth tissue profiling using  
576 multiplexed immunohistochemical consecutive staining on single slide. *Science Immunology* **1**, aaf6925–aaf6925 (2016).
- 577 9. Gut, G., Herrmann, M. D. & Pelkmans, L. Multiplexed protein maps link subcellular organization to cellular states. *Science* **361**, eaar7042 (2018).
- 578 10. Czech, E., Aksoy, B. A., Aksoy, P. & Hammerbacher, J. Cytokit: a single-cell analysis toolkit for high dimensional fluorescent microscopy  
579 imaging. *BMC Bioinformatics* **20**, 448 (2019).
- 580 11. <https://spacex-starfish.readthedocs.io/en/latest/>.
- 581 12. Schapiro, D., Jackson, H. W., Raghuraman, S., Fischer, J. R., Zanotelli, V. R. T., Schulz, D., Giesen, C., Catena, R., Varga, Z. & Bodenmiller, B.  
582 histoCAT: analysis of cell phenotypes and interactions in multiplex image cytometry data. *Nat Methods* **14**, 873–876 (2017).
- 583 13. Bankhead, P., Loughrey, M. B., Fernández, J. A., Dombrowski, Y., McArt, D. G., Dunne, P. D., McQuaid, S., Gray, R. T., Murray, L. J., Coleman,  
584 H. G., James, J. A., Salto-Tellez, M. & Hamilton, P. W. QuPath: Open source software for digital pathology image analysis. *Sci Rep* **7**, 16878  
585 (2017).
- 586 14. Stoltzfus, C. R., Filipek, J., Gern, B. H., Olin, B. E., Leal, J. M., Wu, Y., Lyons-Cohen, M. R., Huang, J. Y., Paz-Stoltzfus, C. L., Plumlee, C. R.,  
587 Pöschinger, T., Urdahl, K. B., Perro, M. & Germer, M. Y. CytoMAP: A Spatial Analysis Toolbox Reveals Features of Myeloid Cell Organization in  
588 Lymphoid Tissues. *Cell Reports* **31**, 107523 (2020).
- 589 15. Krueger, R., Beyer, J., Jang, W.-D., Kim, N. W., Sokolov, A., Sorger, P. K. & Pfister, H. Facetto: Combining Unsupervised and Supervised  
590 Learning for Hierarchical Phenotypic Analysis in Multi-Channel Image Data. *IEEE Trans. Visual. Comput. Graphics* **26**, 227–237 (2020).
- 591 16. Sofroniew, N., Talley Lambert, Evans, K., Nunez-Iglesias, J., Yamauchi, K., Solak, A. C., Buckley, G., Bokota, G., Tung, T., Ziyangcz, Freeman,  
592 J., Boone, P., Winston, P., Loic Royer, Har-Gil, H., Axelrod, S., Rokem, A., Bryant, Hector, Mars Huang, Pranathi Vemuri, Dunham, R.,  
593 Jakirkham, Siqueira, A. D., Bhavya Chopra, Wood, C., Gohlke, C., Bennett, D., DragaDoncila & Perlman, E. *napari/napari: 0.3.5*. (Zenodo,  
594 2020). doi:10.5281/ZENODO.3555620
- 595 17. Allan, C., Burel, J.-M., Moore, J., Blackburn, C., Linkert, M., Loynton, S., Macdonald, D., Moore, W. J., Neves, C., Patterson, A., Porter, M.,  
596 Tarkowska, A., Loranger, B., Avondo, J., Lagerstedt, I., Lianas, L., Leo, S., Hands, K., Hay, R. T., Patwardhan, A., Best, C., Kleywegt, G. J.,  
597 Zanetti, G. & Swedlow, J. R. OMERO: flexible, model-driven data management for experimental biology. *Nat Methods* **9**, 245–253 (2012).
- 598 18. Rashid, R., Chen, Y.-A., Hoffer, J., Muhlich, J. L., Lin, J.-R., Krueger, R., Pfister, H., Mitchell, R., Santagata, S. & Sorger, P. K. *Online narrative*  
599 *guides for illuminating tissue atlas data and digital pathology images*. (Scientific Communication and Education, 2020).  
600 doi:10.1101/2020.03.27.001834



