## Supplementary Materials

**Supplementary Table 1. Description of the components for reporting a highly multiplexed tissue imaging experiment**

| Data Level | Assay data | Assay data description | Metadata | Metadata Description |
|---|---|---|---|---|
| N/A | N/A | N/A | Clinical metadata | Participant identifiers, demographics, diagnosis, exposures, treatments, and follow-up (**Supplementary Table 5**) |
| N/A | N/A | N/A | Biospecimen metadata | Identifying and descriptive information about the method of tissue acquisition, sample processing and handling, and features pertaining to histologic assessment (**Supplementary Table 7**) |
| 1 | Raw numerical output of acquisition instruments | Vendor-specific image formats, ideally Bioformats compatible. Whole slide images recorded as individual tiles from x, y stage positions | File level metadata | Microscope specifications, image acquisition parameters, sample information (**Supplementary Table 8**) |
| 2 | Full resolution OME-TIFF | Primary images after stitching, registration, illumination correction, background subtraction, intensity normalization using automated software. File conversion from vendor specific format to universal OME-TIFF | OME-TIFF header metadata | Image spatial and bit depth properties, channel information<br><br>(**Supplementary Tables 8-10**) |
| | | | Channel level metadata | Microscope specifications, channel information, fluorophore/metal label data, experiment protocol<br><br>(**Supplementary Tables 8-10**) |
| 3 | Full resolution OME-TIFF | Level 2 image data after quality control to remove image artefacts and uninformative channels caused by tissue degradation or poor antibody staining | OME-TIFF header metadata | Image spatial and bit depth properties, channel information, rendering settings for visualization e.g., Minerva, OMERO, etc. |

| | | | | |
|---|---|---|---|---|
| | | | | (**Supplementary Tables 8-10**) |
| | | | Channel level metadata | Microscope specifications, channel information, fluorophore/metal label data, experiment protocol<br><br>(**Supplementary Tables 8-10**) |
| | Segmentation masks as OME-TIFF | Labelled masks where each cell has a unique ID. Masks generated by automated segmentation models and with some human oversight | Segmentation level metadata | Image spatial and bit depth properties, algorithm/workflow/model details<br><br>(**Supplementary Table 11**) |
| 4 | Spatial feature tables | Single cell features such as marker intensities, centroid coordinates, and area. Derived from level 2 and 3 | Spatial feature table metadata | Image spatial and bit depth properties, algorithm/model details<br><br>(**Supplementary Table 12**) |
| 5 | Data exploration | Cell type annotations | Cell state level metadata | Cell type annotations, image derived data, channel information, algorithm/workflow/method (ie. UMAP, t-SNE), rendering settings for visualization (Minerva, OMERO, etc)<br><br>(**Supplementary Tables 12, 13**) |
| | | Dimensionality-reduction from spatial feature table | | |
| | | Pixel-level machine learning | | |
| | | Annotated visualization and data exploration | | |

An example of a structured data base for these components is available at https://htan-portal-nextjs.now.sh/explore. In addition, documentation for automated data ingress, as implemented in the HTAN consortium, is available at https://ncihtan.github.io/HTAN-Data-Ingress-Docs/.

## Supplementary Table 2. Overview of relevant other initiatives / standards which guided MITI reporting guidelines

| **Initiative / Standard** |
|---|
| "Standardized" validation of antibodies for research applications. Human Protein Atlas[1] |
| The Resource Identification Initiative[2] |
| A Global View of Standards for Open Image Data Formats and Repositories[3] |
| QUAREP-LiMi: A community-driven initiative to establish guidelines for quality assessment and reproducibility for instruments and images in light microscopy[4] |
| Minimum Information guidelines for fluorescence microscopy: increasing the value, quality, and fidelity of image data[5] |
| Minimum Information About a Microarray Experiment (MIAME)[6] |
| Minimum Information about a Genome Sequence[7] |
| Guidelines for reporting single-cell RNA-seq experiments - minSCe[8] |
| Minimum Information for Biological and Biomedical Investigations[904/04/2022 07:36:00 |
| Minimum Information Specification For In Situ Hybridization and Immunohistochemistry Experiments (MISFISHIE)[10] |
| MIRO: guidelines for minimum information for the reporting of an ontology[11] |
| BINA: 4D Bioimaging North America (BINA)[12] |
| MethodsJ2[13] |
| Sharing biological data: why, when, and how[14] |
| PDX-MI: Minimal Information for Patient-Derived Tumor Xenograft Models[15] |
| REMBI: Recommended Metadata for Biological Images—enabling reuse of microscopy data in biology[16] |

**Supplementary Table 3. Open source visualization and analysis tools tailored for highly multiplexed imaging methods**

| Software | Description | Implementation | Source/ Link |
|---|---|---|---|
| napari[17] | Image viewer | local | https://github.com/napari/napari |
| Facetto[18] | Image viewer and image analysis | cloud-based | https://github.com/kruegert/facetto |
| histoCAT[19] | Image viewer and image analysis | local and cloud-based | https://github.com/BodenmillerGroup/histoCAT |
| QuPath[20] | Image viewer and image analysis | local | https://qupath.github.io/ |
| Minerva[21] | Data sharing, image viewer and image analysis | local and cloud-based | https://github.com/labsysplharm/minerva-story |
| MCMICRO[22] | Image preprocessing and image analysis | local and cloud-based | https://github.com/labsysplharm/mcmicro |
| Viv[23] | Image visualization library | cloud-based | https://github.com/hms-dbmi/viv |
| ImaCytE[24] | Image viewer and image analysis | local | https://github.com/biovault/ImaCytE |
| OMERO[25] | Image data management and viewer | cloud-based | https://github.com/ome/openmicroscopy |
| ImageJ[26] | Image viewer and image analysis | local | https://imagej.net |
| Giotto[27] | Image viewer and image analysis | local | https://github.com/RubD/Giotto/ |
| cellxgene[28] | Interactive explorer for single-cell transcriptomics data | local and cloud-based | https://github.com/chanzuckerberg/cellxgene |
| Squidpy[29] | Image viewer and image analysis | local | https://github.com/theislab/squidpy |

**Supplementary Table 4. Highly multiplexed imaging methods**

| | | | |
|---|---|---|---|
| **Optical** | **Fluorophore-based** | **Single staining and cyclic detection** | CO-Detection by indEXing (CODEX)[30]<br><br>Immunostaining with Signal Amplification By Exchange Reaction (Immuno-SABER)[31] |
| | | **Cyclic staining and detection** | Cyclic Immunofluorescence (CyCIF)[32]<br><br>Multiplex immunofluorescence (MxIF)[33]<br><br>Iterative indirect immunofluorescence imaging (4i)[34]<br><br>Multi-epitope-ligand cartography (MELC)[35] |
| | **Enzyme-based** | | Multiplex Immunohistochemistry (mIHC)[36]<br><br>Multiplexed immunohistochemical consecutive staining on single slide (MICSSS)[37] |
| **Non-optical** | **Detection based on atomic mass spectrometry of metal-labelled antibodies** | | Imaging Mass Cytometry (IMC)[38]<br><br>Multiplexed Ion Beam Imaging (MIBI)[39] |

**Supplementary Table 4:** An overview of the most commonly used antibody-based highly multiplexed tissue imaging methods. Low-plex imaging methods, such as Hematoxylin and Eosin (H&E), Immunohistochemistry (IHC), Vectra and others are supported but not included in the table.

# Supplementary Note

## 1. The components needed to describe a highly multiplexed tissue imaging experiment

The proposed metadata schema is currently designed to record information from tumor atlas samples and includes clinical data elements to provide information about biospecimens from human subjects with either cancer or precancer lesions. However, the data elements presented here can be adapted for experiments utilizing samples from patients with a range of disease types and the framework can be modified to accommodate non-human samples as well as non-tumor samples. A minimum set of clinical data elements is provided here (**Supplementary Table 1**) that covers participant demographics, diagnosis, exposures, molecular testing, treatments and follow-up. In the HTAN initiative (https://humantumoratlas.org), attributes are available for reporting additional characteristics such as prior patient history, family history, and molecular testing. The complete HTAN data standards are browsable at https://htan-portal-nextjs.now.sh/standards.

Following the example of guidelines for reporting genomic data, we have indicated the level of significance for these attributes as either "required" or "recommended." A value (specified below) must be added for attributes marked as required. The values will be represented by strings, dates, numeric or boolean variables, as appropriate. Whenever possible, valid values will be limited to either a predefined set of keywords (e.g., "Alive" or "Dead") or a numeric interval (e.g., positive integer). Values of 'Unknown', 'Not Reported' and 'Not Applicable' can be used if information is missing for certain samples. 'Unknown' indicates that it is unclear whether data was or was not reported, 'Not Reported' indicates that the information was never collected and cannot be updated, and 'Not Applicable' indicates that the attribute does not apply to the participant or the study. When "required" attributes are conditional (e.g., only available for a specific method/technology or only relevant to a specific disease type) 'Not Applicable' should be used. 'Unknown' or 'Not Reported' should be used whenever possible. The number '0' represents the numerical value.

Submitters of image data must remove Protected Health Information (PHI) and must de-identify the data prior to submission. Participant PHI that must be removed according to the safe harbor method for de-identification includes names, medical record numbers, geographical identifiers smaller than a state, contact information, all dates (other than year) related to the participant (including birthdate, encounter dates such as procedure and follow up dates, date of death). One example of an automated de-identification procedure is provided by DICOM, and some repositories may require stringent removal of all absolute dates even if they are not directly related to PHI such as the date of image acquisition.

To obfuscate PHI, dates are reported as calculated fields (in days) using reference or anchor dates ("index date") as described by the GDC. Recommended index dates are included in the attribute descriptions (**Supplementary Tables 5, 6**). As an example, the HTAN initiative utilizes the date of birth of the participant as the index date. TCGA used the date of pathologic diagnosis as the index date.

To enable effective processing by computational tools, all metadata files must follow tidy data practices[40]. Specifically, metadata tables should have rectangular shape with rows corresponding to observation instances (e.g., patients, channels, markers, etc.) and columns corresponding to attributes **(Supplementary Tables 5-12**) measured in each instance. Adherence to tidy data standards is straightforward when metadata tables are stored using standard formats like comma-separated value (CSV) or H5AD[41]. When composing metadata tables manually, users are cautioned that Microsoft Excel is known to corrupt data upon entry[42]; the issue can be avoided by using alternative spreadsheet software, e.g., Google Sheets.

### Implementation and maintenance of the MITI standard in YAML

The metadata definitions in this manuscript (Supplementary Table 1) represent a conceptual design for a standard in human-readable form. General definitions are accompanied by detailed specifications,

also in the form of human readable tables, for metadata fields and their allowed values (Supplementary Tables 5-12). To enable implementation of MITI in a real-world setting, the specification is available in a machine-readable YAML format through a publicly accessible GitHub repository (https://github.com/labsyspharm/MITI). It is the intention of MITI developers to build software tools to implement MITI based on these YAML files.

The choice of YAML as a markup language was motivated by its simple declarative structure making it both machine and human readable. Individual YAML files capture attributes, their description and significance (Supplementary Tables 5-12), but also additional information that is essential for validating specific files; this includes ensuring that data in each field has the correct attribute (e.g., boolean, integer, string, etc.) and that it meets constraints on valid values (e.g., a predefined set of keywords). MITIv1.0 YAML files are an exact match to the content presented here, but MITI is expected to undergo regular updates as it is deployed on a large scale by the imaging community; collaborative revision will be facilitated by the version control and source code management functionality provided by Git (and GitHub).

## 2. Data Levels for antibody-based multiplexed tissue Imaging

As of August 2021, the Minimal Information and Tissue Imaging standard (MITI) is still in an "request for comment" period, but we expect the final data levels for multiplexed tissue images generated using antibody reagents to be close to what is described below.

The concept of "data levels" (or tiers) was first developed by the Cancer Genome Atlas (TCGA), Genomic Data Commons (GDC) and the database of Genotypes and Phenotypes (dbGAP) to standardize the transformation of raw data (as generated by a measurement apparatus) into processed and interpreted data as used in research publications. Use of data levels promotes uniform and reproducible data analysis and interpretation [33]. Data levels for genomic data also distinguish between restricted access data that carries a de-identification risk (Level 1 and 2 data such as primary sequence) and open access data (Level 3 data such as RNA-seq gene counts). In the case of tissue images de-identification is not considered a risk and only the extent of data processing is considered in establishing a data level.

**Level 1 MITI data** comprise the raw numerical output of acquisition instruments (microscopes, slide scanner, mass cytometers etc.). These data may be in a variety of vendor-specific formats, although all microscope vendors and investigators are strongly encouraged to conform to universally recognized *Bioformats* standards. For a whole slide image, Level 1 will contain many individual image tiles (images recorded from different x, y positions in the specimen).

FASTq files are a common type of Level 1 data in genomics.

**Level 2 MITI data** comprise full-resolution primary images in the universal OME-TIFF format that have undergone stitching, registration, illumination correction, background subtraction, intensity normalization etc. to generate high quality mosaic images. The processing of Level 1 data to generate Level 2 data must be performed using automated software routines (not human intervention), ideally using open-source algorithms whose operation is transparent. The generation of an image mosaic from multiple image tiles using e.g., ASHLAR[43] is a prototypical Level 1 to Level 2 transformation.

BAM files are a common type of Level 2 data in genomics.

**Level 3 MITI data** are the results of image processing and include segmentation masks, images labelled by humans (e.g., to identify nuclei or annotated histology) or by software algorithms. The generation of Level 3 data may involve human interpretation, which should be recorded as part of the image metadata. To provide conformity with GDC data access concepts, full-resolution primary OME-TIFF images that have been subjected to human, or human-assisted software-based quality control are considered Level 3 data. A typical transformation from Level 2 to 3 images involves removing channels

in which staining failed (a bad reagent batch) or cyclic data acquisition was interrupted (e.g., a dropped slide).

We anticipate that Level 3 images will be the primary types of images made available via public-facing data repositories. A key feature of microscopy in general, and tissue imaging in particular, is that virtually all types of human and computational analysis require access to full resolution data files, which can be very large. In contrast, in genomics, many types of analysis are possible using highly processed and compressed files; sustaining access to Level 3 images therefore imposes a substantial burden.

For tissue imaging, level 3 data types include:

- Level 3 Image mosaics that have been subjected to quality control, typically to remove staining and channel specific issues, and for cyclic methods, channels in which tissue damage has reached unacceptable levels.
- Segmentation masks, which are typically generated using software, such as UnMICST[44] but subjected to some level of human oversight or, in the case of machine learning, to supervised training. The models used to generate masks should be recorded.
- "Data Overviews" involving browser-based tools such as MINERVA[21] (or OME Viewers if an OMERO[25] database is available) that make it possible to browse images without having to download them. Viewing in these cases should involve as little additional interpretation as possible.

mRNA expression levels (RNA-seq gene count tables) are a common type of Level 3 genomic data.

**Level 4** data are numerical data generated from processing Level 3 data, most commonly to create "spatial feature tables" describing marker intensities, cell coordinates and other single-cell features (the analogy is with count tables in RNA sequencing).

**Level 5** data are results (e.g., cell type annotation) derived from Levels 4 spatial feature tables and level 3 images. Typical level 5 data include:

- MINERVA "Data Explorations" that use digital docents and human-generated annotation to guide users through the features of a complex set of images. The analogy is with a traditional figure.
- Dimensionality-reduced version of Level 4 data including all model parameters
- Machine learned models (other than segmentation models) from images or other numerical data
- Models that integrate image data with other data modalities
- Cell type and state annotations
- Tissue architecture information such as ducts in normal tissue and tumor nests in malignant tissue

## 3. Clinical and Patient-Derived Metadata

Clinical Data Elements are used to report patient related metadata including demographics, diagnosis, exposures, treatment and follow-up and are based on the Genomic Data Commons (GDC) Data Model. 'NOS' ('Not otherwise specified') indicates that a general diagnosis was possible but sufficient information was not available to provide a specific diagnosis. Complete data standards, including valid values, for HTAN are available on the HTAN Data Portal at https://htan-portal-nextjs.now.sh/standards. Note that metadata describing patient-derived material used to create a model may also be reported using these data elements where applicable.

### Supplementary Table 5. Clinical and Patient-Derived Data Elements for Highly Multiplexed Tissue Imaging Experiments

| Attribute | Description | Valid Values | Significance |
|---|---|---|---|
| **Participant ID** | Participant Identifier | | REQUIRED |

| | | | |
|---|---|---|---|
| **Species** | Text that identifies species of tissue sample. | Human, Primate, Mouse, Other | REQUIRED |
| Demographics | | | |
| **Ethnicity** | Text designations that identify ethnicity. | Hispanic or Latino, Not Hispanic or Latino | REQUIRED |
| **Gender** | Text designations that identify gender. | Female, Male, non-binary, self-identify/other: (please specify) | REQUIRED |
| **Race** | Text designations that identify race. | White, American Indian or Alaska native, Black or African American, Asian, Native Hawaiian or other Pacific islander, other | REQUIRED |
| **Vital status at last follow up** | The survival state of the person at the last follow up contact. | Alive, Dead | REQUIRED |
| **Year of death** | Numeric value to represent the year of the death of an individual. | Integer (YYYY) | RECOMMENDED |
| **Cause of death** | Text term to identify the cause of death for a patient. | Cancer Related, Cardiovascular Disorder, NOS, End-stage Renal Disease, Infection, Not Cancer Related, Renal Disorder, NOS, Spinal Muscular Atrophy, Surgical Complications, Toxicity | RECOMMENDED |
| **Days to death** | Age at the time of death expressed in the number of days since birth. | Number | RECOMMENDED |
| Diagnosis | | | |
| **Age at diagnosis** | Age at the time of diagnosis expressed in the number of days since birth. | Number | REQUIRED |
| **Known genetic predisposition mutation** | A yes/no/unknown indicator to identify whether there is a known genetic predisposition mutation present in the patient. | Yes, No | RECOMMENDED, IF APPLICABLE |
| **Hereditary cancer predisposition syndrome** | Inherited genetic predisposition syndrome that confers heightened susceptibility to cancer in the patient. | Examples for Oncology are listed in the NCI CDE Browser | RECOMMENDED, IF APPLICABLE |
| **Primary diagnosis** | Text term used to describe the patient's histologic diagnosis. | Examples for Oncology are listed in the NCI CDE Browser | REQUIRED |
| **Morphology** | Histologic Morphology Code based on International Classification of Diseases $10^{th}$ Edition. | Examples for Oncology are listed in the NCI CDE Browser | REQUIRED |
| **Site of resection or biopsy** | The text term used to describe the anatomic site of origin, of the patient's disease. | Examples for Oncology are listed in the NCI CDE Browser | REQUIRED |

| | | | |
|---|---|---|---|
| **Tissue or organ of origin** | The text term used to describe the anatomic site of origin, of the patient's disease. | Examples for Oncology are listed in the NCI CDE Browser | REQUIRED |
| **Classification of tumor** | Text that describes the kind of disease present in the tumor specimen as related to a specific time point. | Primary, Metastasis, Recurrence, Progression, Premalignant | RECOMMENDED, IF APPLICABLE |
| **Prior treatment** | Yes/no/unknown indicator related to the administration of therapeutic agents received before the body specimen was collected. | Yes, No | RECOMMENDED, IF APPLICABLE |
| **Days to last follow up** | Time interval from the date of last follow up to the date of initial pathologic diagnosis, represented as a calculated number of days. | Number | REQUIRED |
| **Last known disease status** | Text term that describes the last known state or condition of an individual's neoplasm. | Distant met recurrence/progression, Loco-regional recurrence/progression, Biochemical evidence of disease without structural correlate, Tumor free, Unknown tumor status, With tumor | REQUIRED, IF APPLICABLE |
| **Days to last known disease status** | Time interval from the date of last follow up to the date of initial pathologic diagnosis, represented as a calculated number of days. | Number | REQUIRED, IF APPLICABLE |
| **Progression or recurrence** | Yes/No/Unknown indicator to identify whether a patient has had a new tumor event after initial treatment. | Yes, No | REQUIRED, IF APPLICABLE |
| **Days to progression** | Number of days between the date of the patient’s first diagnosis and the date the patient's disease progressed. | Number | REQUIRED, IF APPLICABLE |
| **Days to progression free** | Number of days between the date used for index (DOB) and the date the patient’s disease was formally confirmed as progression-free. | Number | REQUIRED, IF APPLICABLE |
| **Tumor grade** | Numeric value to express the degree of abnormality of cancer cells, a measure of differentiation and aggressiveness. | G1, G2, G3, G4, GX, GB, High Grade, Low Grade | REQUIRED, IF APPLICABLE |
| Exposures | | | |
| **Pack years smoked** | Numeric computed value to represent lifetime tobacco exposure defined as number of cigarettes smoked per day x number of years. | Number | RECOMMENDED |
| **Years smoked** | Numeric value to represent the number of years a person has been smoking. | Number | RECOMMENDED |
| Canonical mutational information | | | |

| | | | |
|---|---|---|---|
| **Gene symbol** | The text term used to describe a gene targeted or included in molecular analysis. | Refer to HTAN Data Portal for examples and complete list | REQUIRED, IF AVAILABLE |
| **Molecular analysis method** | The text term used to describe the method used for molecular analysis. | Refer to HTAN Data Portal for examples and complete list | REQUIRED, IF AVAILABLE |
| **Test result** | The text term used to describe the result of the molecular test. | Refer to HTAN Data Portal for examples and complete list | REQUIRED, IF AVAILABLE |
| Treatment | | | |
| **Treatment type** | Text term that describes the kind of treatment administered. | Examples for Oncology are listed in the NCI CDE Browser | REQUIRED |
| **Initial disease status** | Text term used to describe the status of the patient's malignancy when the treatment began. | Initial Diagnosis, Progressive Disease, Recurrent Disease, Residual Disease | RECOMMENDED, IF APPLICABLE |
| **Therapeutic agents** | Text identification of the individual agents used as part of a treatment regimen. | String | REQUIRED |
| Follow-Up | | | |
| **Days to follow up** | Number of days between the date of sample acquisition and the date of the patient's last follow-up appointment or contact. | Number | RECOMMENDED |
| **Lost to follow up** | Yes/No/Unknown indicator to identify whether a patient was lost to follow up. | Yes, No | RECOMMENDED |

## 4. Non-Human Metadata Elements for Highly Multiplexed Tissue Imaging Experiments

Separate data elements report on the creation, quality assurance, and study of tissues from non-human model organisms for imaging experiments. With mouse tissues being a prominent example, Supplementary Table 6 aggregates the existing standards defined for mouse models by the Mouse Tumor Biology (MTB) Database[45,46] and the Patient-Derived Xenograft Minimum Information (PDX-MI) Standards[15], in accordance with guidelines set forth by the Alliance for Genome Resources[47].

**Supplementary Table 6. Mouse Metadata Elements for Multiplexed Tissue Imaging Experiments**

| Attribute | Significance |
|---|---|
| **Submitter Patient ID** | REQUIRED |
| **Submitter Sample ID** | REQUIRED |
| **Disease tissue of origin** | REQUIRED |
| **Disease type** | REQUIRED |
| **Is sample from untreated patient?** | REQUIRED |
| **Original sample type** | RECOMMENDED |
| **Sample from an existing PDX model** | RECOMMENDED |

| | |
|---|---|
| **Submitter PDX ID** | REQUIRED |
| **Mouse strain (and source)** | REQUIRED |
| **Strain immune system humanized?** | REQUIRED |
| **Type of humanization** | REQUIRED |
| **Sample preparation** | REQUIRED |
| **Injection type and site** | REQUIRED |
| **Mouse treatment for engraftment** | RECOMMENDED |
| **Engraftment rate** | RECOMMENDED |
| **Engraftment time** | RECOMMENDED |
| **Sample characterization technology** | REQUIRED |
| **Sample confirmed not to be of mouse/EBV origin** | REQUIRED |
| **Response to standard of care (pharmacological positive control)** | REQUIRED |
| **Animal health status** | REQUIRED |
| **Passage QA performance** | REQUIRED |
| **Treatment, passage** | RECOMMENDED |
| **Treatment protocol** | RECOMMENDED |
| **Treatment response** | RECOMMENDED |
| **Sample OMICS** | RECOMMENDED |
| **Lag time/doubling time** | RECOMMENDED |
| **PDX model availability?** | RECOMMENDED |

## 5. Biospecimen Metadata

For each participant in a study, one or several biospecimen samples may be analyzed. Each biospecimen should be assigned a unique label (Biospecimen ID). Relationships between participants and biospecimens can be indicated using the Parent ID; Parent IDs may be used to indicate the source of the biospecimen and when multiple biospecimens are derived from the same source. For tissue imaging, samples are either frozen or fixed and then sectioned and mounted onto slides. In this case each slide is a unique biospecimen from the same parent tissue block. An implementation example is included in Section 11 below.

These biospecimen metadata attributes include information about the method of tissue acquisition, sample processing and handling, and histologic assessment.

### Supplementary Table 7. Biospecimen Attributes for Highly Multiplexed Tissue Imaging Experiments

| Attribute | Description | Valid Values | Significance |
|---|---|---|---|

| | | | |
|---|---|---|---|
| **Biospecimen ID** | Identifier for the Biospecimen | | REQUIRED |
| **Parent ID** | Parent Identifier from which the biospecimen was obtained. The parent could be another biospecimen or a research participant. | | REQUIRED |
| **Adjacent Biospecimen ID** | List of Identifiers (separated by commas) of adjacent biospecimens cut from the same sample | | REQUIRED |
| **Biospecimen Type** | Biospecimen Type | Tissue, Bone Marrow, Cell Block, Blood, Fluids | REQUIRED |
| **Analyte Type** | Text term that represents the kind of molecular specimen analyte. | Tissue section | REQUIRED |
| Collection and Processing | | | |
| **Protocol Link** | Identifier that describes the protocol by which the sample was obtained or generated. E.g. Protocols.io | A valid Digital Object Identifier (DOI) | REQUIRED |
| **Timepoint Label** | Label to identify the time point at which the biospecimen was obtained. | Baseline, End of Treatment, Overall survival, Final | REQUIRED |
| **Collection Days From Index** | Number of days from the research participant's index date that the biospecimen was obtained. | Number | REQUIRED |
| **Acquisition Method Type** | Records the method of acquisition or source for the specimen under consideration. | Autopsy, Biopsy, Fine needle aspirate, Surgical Resection, Punch biopsy, Shave biopsy, Excision, Re-excision, Sentinel node biopsy, Lymphadenectomy | REQUIRED |
| **Post-mortem Interval** | Number of days from the research participant’s date of death that the biospecimen was obtained. | Number | RECOMMENDED |
| **Tumor tissue type** | Text that describes the kind of disease present in the tumor specimen as related to a specific time point. | Primary Tumor, Local Tumor Recurrence, Distant Tumor Recurrence, Metastatic, Premalignant | REQUIRED, IF APPLICABLE |
| **Preservation method** | Text term that represents the method used to preserve the sample. | Formaldehyde fixed paraffin embedded (FFPE), Frozen | REQUIRED |
| **Fixative Type** | The field to identify the type of fixative used to preserve a tissue specimen | Formaldehyde | REQUIRED |

| | | | |
|---|---|---|---|
| **Fixation Duration** | The length of time, from beginning to end, required to process or preserve biospecimens in fixative (measured in minutes) | Number | REQUIRED |
| **Slide Charge Type** | A description of the charge on the glass slide. | Uncharged, Charged, Coverslip, Coated, Not applicable | REQUIRED |
| **Section Thickness Value** | Numeric value to describe the thickness of a slice to tissue taken from a biospecimen, measured in microns (um). | Number | REQUIRED |
| **Days to sectioning** | Number of days between the date of sample acquisition and the date that the specimen was sectioned | Number | REQUIRED |
| **Storage Method** | The method by which a biomaterial was stored after preservation or before another protocol was used. | Ambient temperature, 4°C, -20°C, -80°C | REQUIRED |
| **Days to processing** | Number of days between the date of sample acquisition and the date that the biospecimen was processed | Number | REQUIRED |
| **Shipping Conditions** | Text descriptor of the shipping environment of a biospecimen. | Ambient Pack, Cold Pack, Dry Ice, Ice Pack, Liquid Nitrogen, Other Shipping Environment, Specimen at Room Temperature, Not shipped | REQUIRED |
| Histologic Assessment | | | |
| **Histology Assessments By** | Text term describing who made the histological assessments of the sample | Pathologist, Research Scientist, Other, Unknown | REQUIRED |
| **Histology Assessment Medium** | The method of assessment used to characterize histology | Digital, Microscope | RECOMMENDED |
| **Tumor infiltrating lymphocytes** | Fraction of tumor-infiltrating lymphocytes | Number | RECOMMENDED, IF APPLICBALE |
| **Degree of dysplasia** | Information related to the presence of cells that look abnormal under a microscope but are not cancer. Records the degree of dysplasia for the cyst or lesion under consideration. | Normal, basal cell hyperplasia, or metaplasia<br>Mild dysplasia, Moderate dysplasia, Severe dysplasia, Carcinoma in Situ | RECOMMENDED, IF APPLICABLE |
| **Dysplasia fraction** | Information related to the presence of cells that look abnormal under a microscope but are not cancer. Records the degree of dysplasia for the cyst or lesion under | Number | RECOMMENDED, IF APPLICABLE |

| | | | |
|---|---|---|---|
| | consideration. | | |
| **Number proliferating cells** | Numeric value that represents the fraction of proliferating cells determined during pathologic review of the sample slide(s). | Number | RECOMMENDED, IF APPLICABLE |
| **Percent Necrosis** | Numeric value to represent the percentage of cell death in a malignant tumor sample or specimen. | Number | RECOMMENDED, IF APPLICABLE |
| **Percent Normal Cells** | Numeric value to represent the percentage of normal cell content in a malignant tumor sample or specimen. | Number | RECOMMENDED, IF APPLICABLE |
| **Percent Stromal Cells** | Numeric value to represent the percentage of reactive cells that are present in a malignant tumor sample or specimen but are not malignant such as fibroblasts, vascular structures, etc. | Number | RECOMMENDED, IF APPLICABLE |
| **Percent Tumor Cells** | Numeric value to represent the percentage of infiltration by tumor cells in a sample. | Number | RECOMMENDED, IF APPLICABLE |
| **Percent Tumor Nuclei** | Numeric value to represent the percentage of tumor nuclei in a malignant neoplasm sample or specimen. | Number | RECOMMENDED, IF APPLICABLE |
| Imaging | | | |
| **Fiducial marker** | Fiducial markers for the alignment of images taken across multiple rounds of imaging. | Nuclear Stain (DAPI), Fluorescent beads, Grid slides (hemocytometer), Adhesive markers | REQUIRED (unless N/A) |
| **Mounting medium** | The solution in which the specimen is embedded, generally under a cover glass. It may be liquid, gum or resinous, soluble in water, alcohol or other solvents and be sealed from the external atmosphere by non-soluble ringing media. | Aqueous (water based) medium, Non-Aqueous (Solvent) based, Xylene, Toluene, Antifade with DAPI, Antifade without DAPI, PBS | RECOMMENDED |
| **Slicing method** | The method by which the tissue was sliced. | Vibratome, Cryosectioning, Tissue molds, Sliding microtome, Sectioning | RECOMMENDED |
| **Control** | Control tissue / area | Cell line, tissue, mutation, matrix, metal coating, fiducials | RECOMMENDED |

## 6. File-level Metadata (data level 1)

Raw files without any preprocessing steps are considered Level 1 and the following attributes are needed to sufficiently describe the individual files. These files are sometimes in proprietary formats. For some imaging assays such as whole slide scanning of H&E and IHC stained sections where the output of the machine is immediately usable and doesn't require any pre-processing (e.g., in Leica/Aperio-generated .svs-format files), a format translation may be the only step required to create the "Level 2" OME-TIFF file.

### Supplementary Table 8: File-level Metadata for Highly Multiplexed Tissue Imaging Experiments

| Attribute | Description | Valid Values | Significance |
|---|---|---|---|
| **Data File Id** | unique<br>e.g., "self" ID for data file described by this metadata | see HTAN identifier SOP or create a new one for your project | REQUIRED |
| **Filename** | name of data file as submitted to repository | no spaces,<br>no special- characters | REQUIRED |
| **File Format** | format of data file | SVS, OME-TIFF, CZI | REQUIRED |
| **Channel-Level Metadata CSV Filename** | "companion" CSV file containing channel-level metadata for multi-channel/multi-cyclic methods | no spaces,<br>no special- characters | REQUIRED<br>(unless N/A) |
| **Imaging modality** | imaging technology | epifluorescence, lightsheet, confocal, ion beam, laser ablation | REQUIRED |
| **Assay Type** | method / technology used | H&E, t-CyCIF, IHC, mIHC, MxIF, SABER, IMC, CODEX, MIBI, MELC | REQUIRED |
| **Protocol ID(s)** | unique identifier (e.g. protocols.io) -- *may be a delimited **list** of identifiers* | DOI | REQUIRED |
| **Acquisition software** | method(s) and parameters used for any acquisition/ transformation/ processing to create this image | DOI or github repo/commit or name/version | RECOMMENDED |
| **Image Acquisition Date** | | YYYY-MM-DD | RECOMMENDED |
| **Microscope** | Microscope type (manufacturer, model, etc) used for this experiment | | RECOMMENDED |
| **Objective** | objective lens used (descriptor or other identifier) | | RECOMMENDED |
| **Nominal Magnification** | nominal optical magnification | 10X, 40X | RECOMMENDED |
| **Lens NA** | lens numerical aperture | (floating point) | RECOMMENDED |
| **Working Distance** | distance from objective to coverslip | (floating point) | RECOMMENDED |
| **Working Distance Unit** | default = microns | | RECOMMENDED |
| **Immersion Medium** | the imaging medium affects the working NA of the objective | Air, Water, Glycerin, Oil | RECOMMENDED |

| | | | |
|---|---|---|---|
| **Comment** | Free text field to use if necessary | | RECOMMENDED |
| **Frame_ Averaging** | Number of frames averaged together (if no averaging, set to 1) | 1, 2, 3, ... | RECOMMENDED |
| **Fov_Number** | Index of FOV (as it pertains to its place in the Experiment) | 1, 2, 3, ... | REQUIRED (unless N/A) |
| **Fov_Sizex** | FOV size in X-dimension (micron) | | REQUIRED (unless N/A) |
| **Fov_Sizey** | FOV size in Y-dimension (micron) | | REQUIRED (unless N/A) |
| **Pyramid** | whether or not image file contains a pyramid stack | True or False (or T/F) | REQUIRED (unless N/A) |
| **Z_Stack** | whether or not image file contains a Z-stack | True or False (or T/F, N/A) | REQUIRED (unless N/A) |
| **T_Series** | whether or not image file contains a time series | True or False (or T/F, N/A) | REQUIRED (unless N/A) |

## 7. Channel-level Metadata Attributes (data levels 1 and 2)

Channel-level metadata can be provided in a companion CSV file with each OME-TIFF file associated with one such table. This companion CSV, referenced by filename in the corresponding level 1 and 2 file level metadata table **(Supplementary Table 7),** should have as many rows as necessary to describe each channel in the image data file -- with at least one row of data for each channel within the image data file. Each column in the CSV will represent a single attribute defined below, for all channels. Antibodies must be identified with their RRID identifiers[2] and validation should follow previously described practices[1,48] although validation results are currently external to the MITI standard.

The first two attributes listed below are required in order to align other information in this CSV to the metadata in the OME-TIFF header and in the file-level metadata. Beyond these two, the submitting center has flexibility to choose what information to provide -- *but ideally the list of attributes should be finite and agreed-upon*. If there are channels which represent more than one target, or more than one antibody or fluorophore, then that information should be provided on multiple rows (and the corresponding Channel ID and Channel Name repeated) to follow tidy data standards[40].

### Supplementary Table 9. Channel-level Metadata for Highly Multiplexed Tissue Imaging Experiments

| Attribute | Description | Valid Values | Significance |
|---|---|---|---|
| **Channel ID** | this must match the corresponding field in the OME-XML / TIFF header | “Channel:0:1” | REQUIRED |
| **Channel Name** | this must match the corresponding field in the OME-XML / TIFF header | “Blue” or “CD45” or “E-cadherin” | REQUIRED |
| **Cycle Number** | the cycle # in which the co-listed reagent(s) was(were) used | 1, 2, 3, … (up to number of cycles) | RECOMMENDED |
| **Sub-Cycle #** | sub-cycle # | 1, 2, 3, ... | RECOMMENDED |
| **Target name** | short descriptive name (abbreviation) for this target (antigen) | “Keratin”, “CD163”, “DNA” | REQUIRED |
| **Antibody Name** | short descriptive name for this antibody | “Keratin-570”, “CD8a-488” | REQUIRED |

| | | | |
|---|---|---|---|
| **Antibody Role** | where appropriate, indicate whether antibody is primary or secondary | “Primary” or “Secondary” | RECOMMENDED |
| **RRID identifier** | Research Resource Identifier | “RRID: AB_394606” | REQUIRED |
| **Fluorophore** | Fluorescent dye label | “Alexa Fluor 488” | REQUIRED |
| **Clone** | Unique clone identifier | “OX-8” | REQUIRED |
| **Lot** | lot number from vendor | | REQUIRED |
| **Vendor** | Vendor name | Abcam, CST, eBioscience | REQUIRED |
| **Catalog #** | catalog number from vendor | | REQUIRED |
| **Excitation wavelength** | center/peak of the excitation spectrum (nm) | 499 | RECOMMENDED |
| **Emission wavelength** | center/peak of the emission spectrum (nm) | 520 | RECOMMENDED |
| **Excitation bandwidth** | nominal width of excitation spectrum (nm) | 30 | RECOMMENDED |
| **Emission bandwidth** | nominal width of emission spectrum (nm) | 30 | RECOMMENDED |
| **Metal isotope: Element** | Element abbreviation | “La” or “Nd” | REQUIRED |
| **Metal isotope: Mass** | Element mass number | “139” or “142” | REQUIRED |
| **Oligo barcode: upper strand** | DNA barcode used for labeling | “AATGGTAC” | REQUIRED |
| **Oligo barcode: lower strand** | DNA barcode used for labeling | “AATGGTAC” | REQUIRED |
| **Dilution** | Final dilution ratio used in experiment | 1:1000 | RECOMMENDED |
| **Concentration** | Final concentration used in experiment | 10 ug/mL | RECOMMENDED |
| **Passed QC** | Identify stains that did not pass QC but are included in the dataset. | T/F | RECOMMENDED |
| **QC details** | Comment on why QC failed | text | RECOMMENDED |
| **Comment** | Free text field to use if necessary | | RECOMMENDED |

## 8. OME-TIFF Header Metadata Attributes (data level 2)

OME-TIFF is an open file format that combines the TIFF format for storing binary pixel data with an OME-based XML metadata header. Using the Bio-Formats software plug-in proprietary file formats are converted into OME-based files that can be opened using any Bio-Formats compatible software.

### Supplementary Table 10. OME-TIFF Header Metadata for Highly Multiplexed Tissue Imaging Experiments

| Attribute | Description | Valid Values | Significance |
|---|---|---|---|
| **Image ID** | unique internal Image identifier | “Image:0” | REQUIRED |
| **Pixels BigEndian** | boolean | true, false | REQUIRED |
| **DimensionOrder** | internal ordering of dimensions | “XYZCT” | REQUIRED |

| | | | |
|---|---|---|---|
| **PhysicalSizeX** | physical size of one pixel in x-dimension | “0.650” | REQUIRED |
| **PhysicalSizeXUnit** | unit for PhysicalSizeX | “µm” | REQUIRED |
| **PhysicalSizeY** | physical size of one pixel in y-dimension | “0.650” | REQUIRED |
| **PhysicalSizeYUnit** | unit for PhysicalSizeY | “µm” | REQUIRED |
| **PhysicalSizeZ** | physical size of one pixel in z-dimension | “0.650” | REQUIRED if size of Z > 1 |
| **PhysicalSizeZUnit** | unit for PhysicalSizeZ | “µm” | REQUIRED if size of Z > 1 |
| **SizeC** | number of channels | integer > 0 | REQUIRED |
| **SizeT** | number of time-points | integer > 0 | REQUIRED |
| **SizeX** | number of pixels in x-dimension | integer > 0 | REQUIRED |
| **SizeY** | number of pixels in y-dimension | integer > 0 | REQUIRED |
| **SizeZ** | number of z-planes | integer > 0 | REQUIRED |
| **Type** | bit depth for each pixel value | “uint16” or “float” | REQUIRED |
| **PlaneCount** | total number of planes in this Image | integer > 0 | REQUIRED |
| **Channel Name** | channel label for each channel in this image | “Blue” or “CD45” or “E-cadherin” | REQUIRED |
| **Comment** | Free text field to use if necessary | | RECOMMENDED |

## 9. Processing- and Segmentation level Metadata Attributes (data level 3)

Processing-level attributes describe steps taken to address issues at the pixel level that compromise downstream analyses. This may include novel steps to identify and correct for image artefacts and reject channels that do not meet data standards. Thus, the output file will be an OME-TIFF image of similar bit depth, lateral size and resolution to the level 2 OME-TIFF but may have fewer channels.

Segmentation-level attributes primarily include information on the generation of the individual masks (e.g., segmentation method, individual thresholds, cellular expansion etc.) and the represented object classes (e.g., Nuclei, Cytoplasm, Cell, Tumor, Stroma etc.). Thus, segmentation masks can represent either cellular compartments or larger cellular communities / tissue structures as well as the same classes segmented with different parameters / methods. This allows dozens of masks to be associated with individual images and image-derived features. Individual pixels in segmentation masks should be labelled with unique object IDs within the scope of each image mask. Background pixels are set to zero. All masks should be saved as TIFF files containing integer values (8/16/32bit). The size and resolution should correspond to the level 2 OME-TIFF image.

Intermediate steps during segmentation (probability maps, e.g., UNET) / overlapping mask from instance segmentation (e.g., MASK-R-CNN) should be converted to a single channel labeled mask as described above. Those files can be stored alongside level 3 metadata.

## Supplementary Table 11. Processed-data level Metadata for Highly Multiplexed Tissue Imaging Experiments

| Attribute | Description | Valid Values | Significance |
|---|---|---|---|
| **Data Type** | Specify if OME-TIFF provided is a segmentation mask or QC-checked image | mask, image | REQUIRED |
| **Data File ID** | unique<br>e.g., “self” ID for data file described by this metadata | see identifier SOP or create a new one for your project | REQUIRED |
| **Filename** | name of data file as submitted to repository | no spaces, no special-characters | REQUIRED |
| **File Format** | format of data file | OME-TIFF | REQUIRED |
| **Parent Data File ID** | ID of Data File from which this Data File was derived | see identifier SOP or create a new one for your project | REQUIRED |
| **Protocol ID(s)** | unique identifier (e.g., protocols.io) *-- may be a delimited **list** of identifiers* | DOI | REQUIRED |
| **Software and Version** | method(s) and parameters used for any quality control, transformation, and visualization to create this image | DOI or github repo/commit or name/version | REQUIRED |
| **Commit SHA** | Short SHA for software version | 8 hexadecimal characters (for github) | REQUIRED |
| **Passed QC** | did all masks pass QC | True or False (or T/F) | If ‘Data type’ == ‘image’, REQUIRED, otherwise RECOMMENDED |
| **Comment** | Free text field to use if necessary | | RECOMMENDED |
| **Object Class** | Defines if the mask delineates nucleus, the cytoplasm, plasma membrane, the whole cell or other | nucleus, cytoplasm, cell, plasma membrane, other | If `Data type` = `mask`, REQUIRED |
| **Object Class Description** | Free text description of object class | eg. “organelle” | If Object Class = ‘’other’’, REQUIRED, otherwise RECOMMENDED |
| **Minimum Intensity Display Range** | Lower-bound intensity value | eg “100” | RECOMMENDED |
| **Maximum Intensity Display Range** | Upper-bound intensity value | eg “65000” | RECOMMENDED |
| **FOV_number** | Index of FOV (as it pertains to its place in the Experiment) | 1, 2, 3, … (where valid values start at ) | RECOMMENDED |

| | | | |
|---|---|---|---|
| **FOV_SizeX** | FOV size in X-dimension (micron) | | RECOMMENDED |
| **FOV_SizeY** | FOV size in Y-dimension (micron) | | RECOMMENDED |
| **Pyramid** | whether or not image file contains a pyramid stack | True or False (or T/F) | REQUIRED |
| **Z_stack** | whether or not image file contains a Z-stack | True or False (or T/F) | REQUIRED |
| **T_series** | whether or not image file contains a time series | True or False (or T/F) | REQUIRED |
| **PhysicalSizeX** | physical size of one pixel in x-dimension | eg “0.650” | REQUIRED |
| **PhysicalSizeXUnit** | unit for PhysicalSizeX | eg “µm” | REQUIRED |
| **PhysicalSizeY** | physical size of one pixel in y-dimension | eg “0.650” | REQUIRED |
| **PhysicalSizeYUnit** | unit for PhysicalSizeY | eg “µm” | REQUIRED |
| **PhysicalSizeZ** | physical size of one pixel in z-dimension | eg “0.200” | REQUIRED if size of Z > 1 |
| **PhysicalSizeZUnit** | unit for PhysicalSizeZ | eg “µm” | REQUIRED if size of Z > 1 |
| **Type** | bit depth for each pixel value | eg “uint16” | REQUIRED |

## 10. Object-level Metadata Attributes (data level 4)

Segmentation-level attributes are associated with a CSV, FCS, or H5AD, etc file containing single cell level measurements across all channels (see channel-level; **Supplementary Table 8**) that have successfully passed QC. Therefore, masks for single cells (see segmentation-level; **Supplementary Table 10**) that have also successfully passed QC are combined with the individual channels for quantification. Non-cellular level information can be added as a separate column into the CSV, FCS, or H5AD file and recorded in the cell-level attributes of the files. A Python package ANNData, used for handling similar information for single-cell omics[41], is particularly well-suited here.

### Supplementary Table 12. Cell-level Attributes for Highly Multiplexed Tissue Imaging Experiments

| Attribute | Description | Valid Values | Significance |
|---|---|---|---|
| **Data File ID** | unique<br>e.g., “self” ID for data file described by this metadata | see identifier SOP or create a new one for your project | REQUIRED |
| **Filename** | name of data file as submitted to repository | no spaces,<br>no special- characters | REQUIRED |
| **File Format** | format of data file | CSV, FCS, h5ad | REQUIRED |
| **Parent ID** | Parent Identifier from which the biospecimen was obtained. The parent could be another biospecimen or a | see identified SOP | REQUIRED |

| | | | |
|---|---|---|---|
| | research participant. | | |
| **Software and Version** | method(s) and parameters used for any acquisition/ transformation/ processing to create this image | DOI for protocols.io or github repo/commit or name/version | REQUIRED |
| **Commit SHA** | Short SHA for software version | 8 hexadecimal characters (for github) | REQUIRED |
| **Passed QC** | did all images/channels pass QC | True or False (or T/F) | REQUIRED |
| **Comment** | Free text field to use if necessary | | RECOMMENDED |
| **Header size** | How many columns in the CSV / FCS | e.g. 120 | REQUIRED |
| **Object classes included** | Defines which cell compartment this mask pertains or which tissue structure | Nucleus, cytoplasm, whole cell, plasma membrane, other | REQUIRED |
| **Object Class Description** | Free text description of object class | eg. “organelle” | If Object Class = ‘’other’’, REQUIRED, otherwise RECOMMENDED |
| **FOV_number** | Index of FOV (as it pertains to its place in the Experiment) | 1,2,3 (where valid values start at ) | RECOMMENDED |
| **FOV_SizeX** | FOV size in X-dimension (micron) | eg “100” | RECOMMENDED |
| **FOV_SizeY** | FOV size in Y-dimension (micron) | eg “100” | RECOMMENDED |
| **Pyramid** | whether or not image file contains a pyramid stack | True or False (or T/F) | REQUIRED |
| **Z_stack** | whether or not image file contains a Z-stack | True or False (or T/F) | REQUIRED |
| **T_series** | whether or not image file contains a time series | True or False (or T/F) | REQUIRED |
| **Type** | Bit depth of image | uint8, uint16, uint32 | REQUIRED |
| **PhysicalSizeX** | physical size of one pixel in x-dimension | eg “0.650” | REQUIRED |
| **PhysicalSizeXUnit** | unit for PhysicalSizeX | eg “µm” | REQUIRED |
| **PhysicalSizeY** | physical size of one pixel in y-dimension | eg “0.650” | REQUIRED |
| **PhysicalSizeYUnit** | unit for PhysicalSizeY | eg “µm” | REQUIRED |
| **PhysicalSizeZ** | physical size of one pixel in z-dimension | eg “0.200” | REQUIRED if size of Z > 1 |
| **PhysicalSizeZUnit** | unit for PhysicalSizeZ | eg “µm” | REQUIRED if size of Z > |

| | | | |
|---|---|---|---|
| | | | 1 |
| **Cell-state level** | Are cell types / states included? | Yes / No | REQUIRED |

## 11. Cell-state Level Metadata Attributes (data level 5)

Attributes associated with cell types and cell states can be specified as a separate Level 5 table or in additional columns that augment Level 4 information (**Supplementary Table 11**) to reduce storage size. Each cell can be associated with multiple types/states, but each association must be specified on a separate row in the corresponding level 5 data file. The cell type/state annotations are encoded by a set of keywords from a predefined dictionary cataloguing all possible states and allowing for "Other/Unknown". The dictionary itself is typically derived from known marker-cell type associations (e.g., databases, literature, etc.) or in a data-driven fashion via clustering and cluster annotations. In the metadata table, the dictionary is stored a semicolon-delimited list of keywords (**Supplementary Table 12**). We envision that future versions of the MITI standard will define cell type ontologies to capture hierarchical relationships between the various cell types and states (e.g., "T-cell" is a child node of "Lymphocyte", which is itself a child node of "Immune"), much like GO ontologies currently catalogue hierarchical associations between protein function terms[49]. When implemented, the dictionary field in the metadata table will be replaced by a reference to the cell type ontology resource and its specific version.

## Supplementary Table 13. Cell-state Attributes for Highly Multiplexed Tissue Imaging Experiments

| Attribute | Description | Valid Values | Significance |
|---|---|---|---|
| **Data File ID** | unique<br>e.g., "self" HTAN ID for data file described by this metadata | see HTAN identifier SOP or create a new one for your project | REQUIRED |
| **Filename** | name of data file as submitted to repository | no spaces,<br>no special- characters | REQUIRED |
| **File Format** | format of data file | CSV or FCS | REQUIRED |
| **Parent ID** | Parent Identifier from which the biospecimen was obtained. The parent could be another biospecimen or a research participant. | | REQUIRED |
| **Protocol ID(s)** | unique identifier (e.g. protocols.io) -- *may be a delimited **list** of identifiers* | DOI | REQUIRED |
| **Software and Version** | method(s) and parameters used for any acquisition/ transformation/ processing to create this image; can be a reference to a workflow defined in a standard workflow language (e.g., CWL, Nextflow, Snakemake) | DOI or github repo/commit or name/version | REQUIRED |
| **Comment** | Free text field to use if necessary | | RECOMMENDED |
| **Header size** | How many columns do have in the CSV / FCS / hdf5 | 120 | REQUIRED |
| **Possible cell type / states format** | A semicolon-delimited list of possible keywords that each cell will | "tumor;immune;stromal;proliferating;quiesc | RECOMMENDED |

| | | | |
|---|---|---|---|
| | be annotated with | ent;...;other;unknown” | |
| **Cell type clustering / calling methods** | Methods used (connected to protocols.io) | PhenoGraph, k-means, etc. | RECOMMENDED |
| **Algorithm parameters** | Semicolon-delimited list and description of algorithm parameters | k=15;metric=Euclidean;... | RECOMMENDED |

## 12. Implementations

Example metadata is provided below for two different types of samples. First, a colorectal cancer specimen acquired from the Cooperative Human Tissue Network (CHTN) and used as part of an HTAN trans-network project (HTAN TNP CRC1). Second, a COVID-19 patient specimen. Clinical, Biospecimen and Imaging metadata (e.g., for t-CyCIF) are provided.

Full example implementation can be found here:
https://docs.google.com/spreadsheets/d/1ZSSAxLJ1ci8XQqZch93VmTgZNY6zlt3lgvNiN4e_n-Y/edit?usp=sharing

### Example 1: Overview colorectal cancer example (CRC1)

| Attributes | Example values from HTAN HTAN TNP CRC1 |
|---|---|
| Participant ID | HTA13_1 |
| Gender | Male |
| Race | White |
| Vital status | Alive |
| Age at diagnosis | 25185 days |
| Primary diagnosis | Malignant adenocarcinoma |
| Morphology | Mixed mucinous and signet ring cell adenocarcinoma |
| Site of resection or biopsy | Cecum |
| Tissue or organ of origin | Cecum |
| Classification of tumor | Primary |
| Prior treatment | No |
| Days to last follow up | 720 |
| Last known disease status | Tumor free |
| Days to last known disease status | 720 |
| Progression or recurrence | Yes |

| | |
|---|---|
| Days to recurrence | 365 |
| Tumor grade | G3 |
| Gene symbol | Not applicable |
| Molecular analysis method | Microsatellite analysis |
| Test result | High |
| Gene symbol | BRAF |
| Molecular analysis method | Targeted sequencing |
| Test result | Positive |
| Pack years smoked | 10 |
| Years smoked | 10 |
| Treatment type | Chemotherapy |
| Initial disease status | Initial Diagnosis |
| Therapeutic agents | Oxaliplatin |
| Days to follow up | 25905 |
| Biospecimen Type | Tissue |
| Analyte type | Tissue section |
| Protocol link | dx.doi.org/10.17504/protocols.io.bji2kkge |
| Timepoint label | Initial diagnosis |
| Acquisition method type | Surgical Resection |
| Tumor tissue type | Primary Tumor |
| Preservation method | Formalin fixed paraffin embedded (FFPE) |
| Fixative type | Formalin |
| Fixation duration | 24 |
| Slide charge type | Charged |
| Section thickness value | 5 |
| Storage method | Refrigerated at 4 degrees |
| Shipping conditions | Ambient pack |
| Histology assessments by | Pathologist |

| | |
|---|---|
| Degree of dysplasia | Carcinoma in situ |
| Percent stromal cells | 40 |
| Percent tumor cells | 40 |
| Mounting medium | Xylene |
| Slicing method | Sectioning |

**Example 2: COVID-19 example**

| Attributes | Example values from study of SARS-CoV-2 |
|---|---|
| Participant ID | S01 |
| Gender | Female |
| Race | Black/African American |
| Vital status | Dead |
| Age at diagnosis | 24090 days |
| Known genetic predisposition mutation | Not applicable |
| Hereditary cancer predisposition syndrome | Not applicable |
| Primary diagnosis | COVID-19 pneumonia |
| Morphology | U07.1 |
| Site of resection or biopsy | Lung, right upper lobe |
| Tissue or organ of origin | Lung, right upper lobe |
| Classification of tumor | Not applicable |
| Prior treatment | Not applicable |
| Days to last follow up | 21 |
| Last known disease status | Not applicable |
| Days to last known disease status | Not applicable |
| Progression or recurrence | Not applicable |
| Days to recurrence | Not applicable |
| Tumor grade | Not applicable |
| Pack years smoked | 0 |
| Years smoked | 0 |

| | |
|---|---|
| Initial disease status | Not applicable |
| Therapeutic agents | Azithromycin, Hydroxychloroquine, Tocilizumab |
| Days to follow up | 24111 |
| Biospecimen Type | Tissue |
| Analyte type | Tissue section |
| Protocol link | dx.doi.org/10.17504/protocols.io.bji2kkge |
| Timepoint label | Autopsy |
| Acquisition method type | Autopsy |
| Post-mortem interval | 2 |
| Tumor tissue type | Not applicable |
| Preservation method | Formalin fixed paraffin embedded (FFPE) |
| Fixative type | Formalin |
| Fixation duration | Not reported |
| Slide charge type | Charged |
| Section thickness value | 5 |
| Storage method | Refrigerated at 4 degrees |
| Shipping conditions | Not shipped |
| Histology assessments by | Pathologist |
| Degree of dysplasia | Not applicable |
| Percent stromal cells | Not applicable |
| Percent tumor cells | Not applicable |
| Mounting medium | Xylene |
| Slicing method | Sectioning |

## 13. Reporting Guidelines Discussion

These guidelines for reporting highly multiplexed tissue imaging experiments have been developed as part of discussions in the Human Tumor Atlas Network (HTAN) working groups (Clinical/Biospecimen Working Group, Molecular Characterization Working Group, Data Analysis Working Group, Policy Working Group), and have been modified through a series of HTAN-wide request for comments led by the HTAN Data Coordinating Center (DCC) and through discussions with members of the Image Analysis Working Group of the Cancer Systems Biology Consortium and Physical Sciences-Oncology Network (CSBC/PS-ON) communities, the National Cancer Institute Imaging Data Commons (IDC),

Human Cell Atlas (HCA), the Human BioMolecular Atlas Program (HuBMAP), cBioPORTAL for Cancer Genomics, and the Open Microscopy Environment (OMERO).

**Supplementary References**


1. Edfors, F., Hober, A., Linderbäck, K., Maddalo, G., Azimi, A., Sivertsson, Å., Tegel, H., Hober, S., Szigyarto, C. A.-K., Fagerberg, L., von Feilitzen, K., Oksvold, P., Lindskog, C., Forsström, B. & Uhlen, M. Enhanced validation of antibodies for research applications. *Nat Commun* **9,** 4130 (2018).
2. Bandrowski, A., Brush, M., Grethe, J. S., Haendel, M. A., Kennedy, D. N., Hill, S., Hof, P. R., Martone, M. E., Pols, M., Tan, S., Washington, N., Zudilova-Seinstra, E., Vasilevsky, N., & Resource Identification Initiative Members are listed here: https://www.force11.org/node/4463/members. The Resource Identification Initiative: A cultural shift in publishing. *F1000Res.* **4,** 134 (2015).
3. Swedlow, J. R., Kankaanpää, P., Sarkans, U., Goscinski, W., Galloway, G., Sullivan, R. P., Brown, C. M., Wood, C., Keppler, A., Loos, B., Zullino, S., Longo, D. L., Aime, S. & Onami, S. A Global View of Standards for Open Image Data Formats and Repositories. *arXiv [q-bio.OT]* (2020). at <http://arxiv.org/abs/2010.10107>
4. Nelson, G., Boehm, U., Bagley, S., Bajcsy, P., Bischof, J., Brown, C. M., Dauphin, A., Dobbie, I. M., Eriksson, J. E., Faklaris, O., Fernandez-Rodriguez, J., Ferrand, A., Gelman, L., Gheisari, A., Hartmann, H., Kukat, C., Laude, A., Mitkovski, M., Munck, S., North, A. J., Rasse, T. M., Resch-Genger, U., Schuetz, L. C., Seitz, A., Strambio-De-Castillia, C., Swedlow, J. R., Alexopoulos, I., Aumayr, K., Avilov, S., Bakker, G.-J., Bammann, R. R., Bassi, A., Beckert, H., Beer, S., Belyaev, Y., Bierwagen, J., Birngruber, K. A., Bosch, M., Breitlow, J., Cameron, L. A., Chalfoun, J., Chambers, J. J., Chen, C.-L., Conde-Sousa, E., Corbett, A. D., Cordelieres, F. P., Del Nery, E., Dietzel, R., Eismann, F., Fazeli, E., Felscher, A., Fried, H., Gaudreault, N., Goh, W. I., Guilbert, T., Hadleigh, R., Hemmerich, P., Holst, G. A., Itano, M. S., Jaffe, C. B., Jambor, H. K., Jarvis, S. C., Keppler, A., Kirchenbuechler, D., Kirchner, M., Kobayashi, N., Krens, G., Kunis, S., Lacoste, J., Marcello, M., Martins, G. G., Metcalf, D. J., Mitchell, C. A., Moore, J., Mueller, T., Nelson, M. S., Ogg, S., Onami, S., Palmer, A. L., Paul-Gilloteaux, P., Pimentel, J. A., Plantard, L., Podder, S., Rexhepaj, E., Royon, A., Saari, M. A., Schapman, D., Schoonderwoert, V., Schroth-Diez, B., Schwartz, S., Shaw, M., Spitaler, M., Stoeckl, M. T., Sudar, D., Teillon, J., Terjung, S., Thuenauer, R., Wilms, C. D., Wright, G. D. & Nitschke, R. QUAREP-LiMi: A community-driven initiative to establish guidelines for quality assessment and reproducibility for instruments and images in light microscopy. *arXiv [q-bio.OT]* (2021). at <http://arxiv.org/abs/2101.09153>
5. Huisman, M., Hammer, M., Rigano, A., Farzam, F., Gopinathan, R., Smith, C., Grunwald, D. & Strambio-De-Castillia, C. Minimum Information guidelines for fluorescence microscopy: increasing the value, quality, and fidelity of image data. *arXiv [q-bio.QM]* (2019). at <http://arxiv.org/abs/1910.11370>
6. Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C. A., Causton, H. C., Gaasterland, T., Glenisson, P., Holstege, F. C. P., Kim, I. F., Markowitz, V., Matese, J. C., Parkinson, H., Robinson, A., Sarkans, U., Schulze-Kremer, S., Stewart, J., Taylor, R., Vilo, J. & Vingron, M. Minimum information about a microarray experiment (MIAME)—toward standards for microarray data. *Nat Genet* **29,** 365–371 (2001).
7. Field, D., Garrity, G., Gray, T., Morrison, N., Selengut, J., Sterk, P., Tatusova, T., Thomson, N., Allen, M. J., Angiuoli, S. V., Ashburner, M., Axelrod, N., Baldauf, S., Ballard, S., Boore, J., Cochrane, G., Cole, J., Dawyndt, P., De Vos, P., DePamphilis, C., Edwards, R., Faruque, N., Feldman, R., Gilbert, J., Gilna, P., Glöckner, F. O., Goldstein, P., Guralnick, R., Haft, D., Hancock, D., Hermjakob, H., Hertz-Fowler, C., Hugenholtz, P., Joint, I., Kagan, L., Kane, M., Kennedy, J., Kowalchuk, G., Kottmann, R., Kolker, E., Kravitz, S., Kyrpides, N., Leebens-Mack, J., Lewis, S. E., Li, K., Lister, A. L., Lord, P., Maltsev, N., Markowitz, V., Martiny, J., Methe, B., Mizrachi, I., Moxon, R., Nelson, K., Parkhill, J., Proctor, L., White, O., Sansone, S.-A., Spiers, A., Stevens, R., Swift, P., Taylor, C., Tateno, Y., Tett, A., Turner, S., Ussery, D., Vaughan, B., Ward, N., Whetzel, T., San Gil, I., Wilson, G. & Wipat, A. The minimum information about a genome sequence (MIGS) specification. *Nat. Biotechnol.* **26,** 541–547 (2008).
8. Füllgrabe, A., George, N., Green, M., Nejad, P., Aronow, B., Fexova, S. K., Fischer, C., Freeberg, M. A., Huerta, L., Morrison, N., Scheuermann, R. H., Taylor, D., Vasilevsky, N., Clarke, L., Gehlenborg, N., Kent, J., Marioni, J., Teichmann, S., Brazma, A. & Papatheodorou, I. Guidelines for reporting single-cell RNA-seq experiments. *Nat. Biotechnol.* **38,** 1384–1386 (2020).
9. Taylor, C. F., Field, D., Sansone, S.-A., Aerts, J., Apweiler, R., Ashburner, M., Ball, C. A., Binz, P.-A., Bogue, M., Booth, T., Brazma, A., Brinkman, R. R., Michael Clark, A., Deutsch, E. W., Fiehn, O.,

Fostel, J., Ghazal, P., Gibson, F., Gray, T., Grimes, G., Hancock, J. M., Hardy, N. W., Hermjakob, H., Julian, R. K., Jr, Kane, M., Kettner, C., Kinsinger, C., Kolker, E., Kuiper, M., Le Novère, N., Leebens-Mack, J., Lewis, S. E., Lord, P., Mallon, A.-M., Marthandan, N., Masuya, H., McNally, R., Mehrle, A., Morrison, N., Orchard, S., Quackenbush, J., Reecy, J. M., Robertson, D. G., Rocca-Serra, P., Rodriguez, H., Rosenfelder, H., Santoyo-Lopez, J., Scheuermann, R. H., Schober, D., Smith, B., Snape, J., Stoeckert, C. J., Jr, Tipton, K., Sterk, P., Untergasser, A., Vandesompele, J. & Wiemann, S. Promoting coherent minimum reporting guidelines for biological and biomedical investigations: the MIBBI project. *Nat. Biotechnol.* **26,** 889–896 (2008).

10. Deutsch, E. W., Ball, C. A., Bova, G. S., Brazma, A., Bumgarner, R. E., Campbell, D., Causton, H. C., Christiansen, J., Davidson, D., Eichner, L. J., Goo, Y. A., Grimmond, S., Henrich, T., Johnson, M. H., Korb, M., Mills, J. C., Oudes, A., Parkinson, H. E., Pascal, L. E., Quackenbush, J., Ramialison, M., Ringwald, M., Sansone, S.-A., Sherlock, G., Stoeckert, C. J., Jr, Swedlow, J., Taylor, R. C., Walashek, L., Zhou, Y., Liu, A. Y. & True, L. D. Development of the Minimum Information Specification for In Situ Hybridization and Immunohistochemistry Experiments (MISFISHIE). *OMICS* **10,** 205–208 (2006).

11. Matentzoglu, N., Malone, J., Mungall, C. & Stevens, R. MIRO: guidelines for minimum information for the reporting of an ontology. *J. Biomed. Semantics* **9,** 6 (2018).

12. Hammer, M., Huisman, M., Rigano, A., Boehm, U., Chambers, J. J., Gaudreault, N., North, A. J., Pimentel, J. A., Sudar, D., Bajcsy, P., Brown, C. M., Corbett, A. D., Faklaris, O., Lacoste, J., Laude, A., Nelson, G., Nitschke, R., Farzam, F., Smith, C. S., Grunwald, D. & Strambio-De-Castillia, C. *Towards community-driven metadata standards for light microscopy: tiered specifications extending the OME model*. (Bioinformatics, 2021). doi:10.1101/2021.04.25.441198

13. Ryan, J., Pengo, T., Rigano, A., Llopis, P. M., Itano, M. S., Cameron, L. A., Marqués, G., Strambio-De-Castillia, C., Sanders, M. A. & Brown, C. M. MethodsJ2: a software tool to capture metadata and generate comprehensive microscopy methods text. *Nat Methods* (2021). doi:10.1038/s41592-021-01290-5

14. Wilson, S. L., Way, G. P., Bittremieux, W., Armache, J., Haendel, M. A. & Hoffman, M. M. Sharing biological data: why, when, and how. *FEBS Lett* **595,** 847–863 (2021).

15. Meehan, T. F., Conte, N., Goldstein, T., Inghirami, G., Murakami, M. A., Brabetz, S., Gu, Z., Wiser, J. A., Dunn, P., Begley, D. A., Krupke, D. M., Bertotti, A., Bruna, A., Brush, M. H., Byrne, A. T., Caldas, C., Christie, A. L., Clark, D. A., Dowst, H., Dry, J. R., Doroshow, J. H., Duchamp, O., Evrard, Y. A., Ferretti, S., Frese, K. K., Goodwin, N. C., Greenawalt, D., Haendel, M. A., Hermans, E., Houghton, P. J., Jonkers, J., Kemper, K., Khor, T. O., Lewis, M. T., Lloyd, K. C. K., Mason, J., Medico, E., Neuhauser, S. B., Olson, J. M., Peeper, D. S., Rueda, O. M., Seong, J. K., Trusolino, L., Vinolo, E., Wechsler-Reya, R. J., Weinstock, D. M., Welm, A., Weroha, S. J., Amant, F., Pfister, S. M., Kool, M., Parkinson, H., Butte, A. J. & Bult, C. J. PDX-MI: Minimal Information for Patient-Derived Tumor Xenograft Models. *Cancer Res.* **77,** e62–e66 (2017).

16. Sarkans, U., Chiu, W., Collinson, L., Darrow, M. C., Ellenberg, J., Grunwald, D., Hériché, J.-K., Iudin, A., Martins, G. G., Meehan, T., Narayan, K., Patwardhan, A., Russell, M. R. G., Saibil, H. R., Strambio-De-Castillia, C., Swedlow, J. R., Tischer, C., Uhlmann, V., Verkade, P., Barlow, M., Bayraktar, O., Birney, E., Catavitello, C., Cawthorne, C., Wagner-Conrad, S., Duke, E., Paul-Gilloteaux, P., Gustin, E., Harkiolaki, M., Kankaanpää, P., Lemberger, T., McEntyre, J., Moore, J., Nicholls, A. W., Onami, S., Parkinson, H., Parsons, M., Romanchikova, M., Sofroniew, N., Swoger, J., Utz, N., Voortman, L. M., Wong, F., Zhang, P., Kleywegt, G. J. & Brazma, A. REMBI: Recommended Metadata for Biological Images—enabling reuse of microscopy data in biology. *Nat Methods* (2021). doi:10.1038/s41592-021-01166-8

17. Sofroniew, N., Talley Lambert, Evans, K., Nunez-Iglesias, J., Winston, P., Bokota, G., Peña-Castellanos, G., Yamauchi, K., Bussonnier, M., Solak, A. C., Ziyangczi, Pop, D. D., Buckley, G., Pam, Alisterburt, Hilsenstein, V., Sweet, A., Hector, Freeman, J., Boone, P., Jordão Bragantini, Lowe, A. R., Gohlke, C., Loic Royer, Kir0ul, Har-Gil, H., Lukasz Migas, Kittisopikul, M., Haase, R. & Axelrod, S. *napari/napari: 0.4.10rc0*. (Zenodo, 2021). doi:10.5281/ZENODO.3555620

18. Krueger, R., Beyer, J., Jang, W.-D., Kim, N. W., Sokolov, A., Sorger, P. K. & Pfister, H. Facetto: Combining Unsupervised and Supervised Learning for Hierarchical Phenotype Analysis in Multi-Channel Image Data. *IEEE Trans. Visual. Comput. Graphics* **26,** 227–237 (2020).

19. Schapiro, D., Jackson, H. W., Raghuraman, S., Fischer, J. R., Zanotelli, V. R. T., Schulz, D., Giesen, C., Catena, R., Varga, Z. & Bodenmiller, B. histoCAT: analysis of cell phenotypes and interactions in multiplex image cytometry data. *Nature Methods* **14,** nmeth.4391 (2017).
20. 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, H. G., James, J. A., Salto-Tellez, M. & Hamilton, P. W. QuPath: Open source software for digital pathology image analysis. *Sci Rep* **7,** 16878 (2017).
21. 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 guides for illuminating tissue atlas data and digital pathology images*. (Scientific Communication and Education, 2020). doi:10.1101/2020.03.27.001834
22. Schapiro, D., Sokolov, A., Yapp, C., Muhlich, J. L., Hess, J., Lin, J.-R., Chen, Y.-A., Nariya, M. K., Baker, G. J., Ruokonen, J., Maliga, Z., Jacobson, C. A., Farhi, S. L., Abbondanza, D., McKinley, E. T., Betts, C., Regev, A., Coffey, R. J., Coussens, L. M., Santagata, S. & Sorger, P. K. *MCMICRO: A scalable, modular image-processing pipeline for multiplexed tissue imaging*. (Bioinformatics, 2021). doi:10.1101/2021.03.15.435473
23. Manz, T., Gold, I., Patterson, N. H., McCallum, C., Keller, M. S., Herr, B. W., Börner, K., Spraggins, J. M. & Gehlenborg, N. *Viv: Multiscale Visualization of High-Resolution Multiplexed Bioimaging Data on the Web*. (Open Science Framework, 2020). doi:10.31219/osf.io/wd2gu
24. Somarakis, A., Van Unen, V., Koning, F., Lelieveldt, B. P. F. & Hollt, T. ImaCytE: Visual Exploration of Cellular Microenvironments for Imaging Mass Cytometry Data. *IEEE Trans. Visual. Comput. Graphics* 1–1 (2019). doi:10.1109/TVCG.2019.2931299
25. Allan, C., Burel, J.-M., Moore, J., Blackburn, C., Linkert, M., Loynton, S., Macdonald, D., Moore, W. J., Neves, C., Patterson, A., Porter, M., Tarkowska, A., Loranger, B., Avondo, J., Lagerstedt, I., Lianas, L., Leo, S., Hands, K., Hay, R. T., Patwardhan, A., Best, C., Kleywegt, G. J., Zanetti, G. & Swedlow, J. R. OMERO: flexible, model-driven data management for experimental biology. *Nat Methods* **9,** 245–253 (2012).
26. Schindelin, J., Rueden, C. T., Hiner, M. C. & Eliceiri, K. W. The ImageJ ecosystem: An open platform for biomedical image analysis. *Mol. Reprod. Dev.* **82,** 518–529 (2015).
27. Dries, R., Zhu, Q., Dong, R., Eng, C.-H. L., Li, H., Liu, K., Fu, Y., Zhao, T., Sarkar, A., Bao, F., George, R. E., Pierson, N., Cai, L. & Yuan, G.-C. Giotto: a toolbox for integrative analysis and visualization of spatial expression data. *Genome Biol.* **22,** 78 (2021).
28. Megill, C., Martin, B., Weaver, C., Bell, S., Prins, L., Badajoz, S., McCandless, B., Pisco, A. O., Kinsella, M., Griffin, F., Kiggins, J., Haliburton, G., Mani, A., Weiden, M., Dunitz, M., Lombardo, M., Huang, T., Smith, T., Chambers, S., Freeman, J., Cool, J. & Carr, A. *cellxgene: a performant, scalable exploration platform for high dimensional sparse matrices*. (Systems Biology, 2021). doi:10.1101/2021.04.05.438318
29. Palla, G., Spitzer, H., Klein, M., Fischer, D., Schaar, A. C., Kuemmerle, L. B., Rybakov, S., Ibarra, I. L., Holmberg, O., Virshup, I., Lotfollahi, M., Richter, S. & Theis, F. J. *Squidpy: a scalable framework for spatial single cell analysis*. (Bioinformatics, 2021). doi:10.1101/2021.02.19.431994
30. Goltsev, Y., Samusik, N., Kennedy-Darling, J., Bhate, S., Hale, M., Vazquez, G., Black, S. & Nolan, G. P. Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed Imaging. *Cell* **174,** 968-981.e15 (2018).
31. 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., Boyden, E. S., Kaeser, P. S., Pihan, G., Church, G. M. & Yin, P. Immuno-SABER enables highly multiplexed and amplified protein imaging in tissues. *Nat Biotechnol* **37,** 1080–1090 (2019).
32. Lin, J.-R., Izar, B., Wang, S., Yapp, C., Mei, S., Shah, P. M., Santagata, S. & Sorger, P. K. Highly multiplexed immunofluorescence imaging of human tissues and tumors using t-CyCIF and conventional optical microscopes. *eLife* **7,** e31657 (2018).
33. 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, V., Kaanumalle, S., Kenny, K., Larsen, M., Lazare, M., Li, Q., Lowes, C., McCulloch, C. C., McDonough, E., Montalto, M. C., Pang, Z., Rittscher, 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 of formalin-fixed, paraffin-embedded cancer tissue. *PNAS* **110,** 11982–11987 (2013).

34. Gut, G., Herrmann, M. D. & Pelkmans, L. Multiplexed protein maps link subcellular organization to cellular states. *Science* **361,** eaar7042 (2018).
35. Schubert, W., Bonnekoh, B., Pommer, A. J., Philipsen, L., Böckelmann, R., Malykh, Y., Gollnick, H., Friedenberger, M., Bode, M. & Dress, A. W. M. Analyzing proteome topology and function by automated multidimensional fluorescence microscopy. *Nat Biotech* **24,** 1270–1278 (2006).
36. 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., 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. & Coussens, L. M. Quantitative Multiplex Immunohistochemistry Reveals Myeloid-Inflamed Tumor-Immune Complexity Associated with Poor Prognosis. *Cell Reports* **19,** 203–217 (2017).
37. Remark, R., Merghoub, T., Grabe, N., Litjens, G., Damotte, D., Wolchok, J. D., Merad, M. & Gnjatic, S. In-depth tissue profiling using multiplexed immunohistochemical consecutive staining on single slide. *Science Immunology* **1,** aaf6925–aaf6925 (2016).
38. Giesen, C., Wang, H. A. O., Schapiro, D., Zivanovic, N., Jacobs, A., Hattendorf, B., Schüffler, P. J., Grolimund, D., Buhmann, J. M., Brandt, S., Varga, Z., Wild, P. J., Günther, D. & Bodenmiller, B. Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nat Methods* **11,** 417–422 (2014).
39. 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, Y. & Nolan, G. P. Multiplexed ion beam imaging (MIBI) of human breast tumors. *Nat Med* **20,** 436–442 (2014).
40. Wickham, H. Tidy Data. *Journal of Statistical Software, Articles* **59,** 1–23 (2014).
41. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. *Genome Biol* **19,** 15 (2018).
42. Ziemann, M., Eren, Y. & El-Osta, A. Gene name errors are widespread in the scientific literature. *Genome Biol.* **17,** 177 (2016).
43. Muhlich, J., Chen, Y.-A., Russell, D. & Sorger, P. K. *Stitching and registering highly multiplexed whole slide images of tissues and tumors using ASHLAR software*. (Bioinformatics, 2021). doi:10.1101/2021.04.20.440625
44. Yapp, C., Novikov, E., Jang, W.-D., Chen, Y.-A., Cicconet, M., Maliga, Z., Jacobson, C. A., Wei, D., Santagata, S., Pfister, H. & Sorger, P. K. *UnMICST: Deep learning with real augmentation for robust segmentation of highly multiplexed images of human tissues*. (Bioinformatics, 2021). doi:10.1101/2021.04.02.438285
45. Krupke, D. M., Begley, D. A., Sundberg, J. P., Richardson, J. E., Neuhauser, S. B. & Bult, C. J. The Mouse Tumor Biology Database: A Comprehensive Resource for Mouse Models of Human Cancer. *Cancer Res* **77,** e67–e70 (2017).
46. Conte, N., Mason, J. C., Halmagyi, C., Neuhauser, S., Mosaku, A., Yordanova, G., Chatzipli, A., Begley, D. A., Krupke, D. M., Parkinson, H., Meehan, T. F. & Bult, C. C. PDX Finder: A portal for patient-derived tumor xenograft model discovery. *Nucleic Acids Research* **47,** D1073–D1079 (2019).
47. The Alliance of Genome Resources Consortium, Agapite, J., Albou, L.-P., Aleksander, S., Argasinska, J., Arnaboldi, V., Attrill, H., Bello, S. M., Blake, J. A., Blodgett, O., Bradford, Y. M., Bult, C. J., Cain, S., Calvi, B. R., Carbon, S., Chan, J., Chen, W. J., Cherry, J. M., Cho, J., Christie, K. R., Crosby, M. A., Pons, J. D., Dolan, M. E., Santos, G. dos, Dunn, B., Dunn, N., Eagle, A., Ebert, D., Engel, S. R., Fashena, D., Frazer, K., Gao, S., Gondwe, F., Goodman, J., Gramates, L. S., Grove, C. A., Harris, T., Harrison, M.-C., Howe, D. G., Howe, K. L., Jha, S., Kadin, J. A., Kaufman, T. C., Kalita, P., Karra, K., Kishore, R., Lauledерkind, S., Lee, R., MacPherson, K. A., Marygold, S. J., Matthews, B., Millburn, G., Miyasato, S., Moxon, S., Mueller, H.-M., Mungall, C., Muruganujan, A., Mushayahama, T., Nash, R. S., Ng, P., Paulini, M., Perrimon, N., Pich, C., Raciti, D., Richardson, J. E., Russell, M., Gelbart, S. R., Ruzicka, L., Schaper, K., Shimoyama, M., Simison, M., Smith, C., Shaw, D. R., Shrivatsav, A., Skrzypek, M., Smith, J. R., Sternberg, P. W., Tabone, C. J., Thomas, P. D., Thota, J., Toro, S., Tomczuk, M., Tutaj, M., Tutaj, M., Urbano, J.-M., Auken, K. V., Slyke, C. E. V., Wang, S.-J., Weng, S., Westerfield, M., Williams, G., Wong, E. D., Wright, A. & Yook, K. Alliance of Genome Resources Portal: unified model organism research platform. *Nucleic Acids Research* **48,** D650–D658 (2020).

48. Du, Z., Lin, J.-R., Rashid, R., Maliga, Z., Wang, S., Aster, J. C., Izar, B., Sorger, P. K. & Santagata, S. Qualifying antibodies for image-based immune profiling and multiplexed tissue imaging. *Nat. Protoc.* **14,** 2900–2930 (2019).
49. Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M. & Sherlock, G. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25,** 25–29 (2000).