MITI minimum information guidelines for highly multiplexed tissue images

The imminent release of tissue atlases combining multichannel microscopy with single-cell sequencing and other omics data from normal and diseased specimens creates an urgent need for data and metadata standards to guide data deposition, curation and release. We describe a Minimum Information about Highly Multiplexed Tissue Imaging (MITI) standard that applies best practices developed for genomics and for other microscopy data to highly multiplexed tissue images and traditional histology.

Denis Schapiro, Clarence Yapp, Artem Sokolov, Sheila M. Reynolds, Yu-An Chen, Damir Sudar, Yubin Xie, Jeremy Muhlich, Raquel Arias-Camison, Sarah Arena, Adam J. Taylor, Milen Nikolov, Madison Tyler, Jia-Ren Lin, Erik A. Burlingame, Human Tumor Atlas Network, Young H. Chang, Samouil L. Farhi, Vésteinn Thorsson, Nithya Venkatamohan, Julia L. Drewes, Dana Pe'er, David A. Gutman, Markus D. Herrmann, Nils Gehlenborg, Peter Bankhead, Joseph T. Roland, John M. Herndon, Michael P. Snyder, Michael Angelo, Garry Nolan, Jason R. Swedlow, Nikolaus Schultz, Daniel T. Merrick, Sarah A. Mazzili, Ethan Cerami, Scott J. Rodig, Sandro Santagata and Peter K. Sorger

Highly multiplexed tissue imaging using any of a variety of optical and mass-spectrometry-based methods (Supplementary Table 1) combines deep molecular insight into the biology of single cells with spatial information traditionally acquired using histological methods, such as hematoxylin and eosin (H&E) staining and immunohistochemistry (IHC)[1]. As currently practiced, multiplexed tissue imaging of proteins involves 20–60 channels of two-dimensional (2D) data, with each channel corresponding to a different antibody or colorimetric stain (Fig. 1). Multiple inter-institutional and international projects, such as the Human Tumor Atlas Network (HTAN)[2], the Human BioMolecular Atlas Program (HuBMAP)[3] and the LifeTime Initiative[4], aim to combine such highly multiplexed tissue images with single-cell sequencing and other types of omics data to create publicly accessible atlases of normal and diseased tissues. Easy public access to primary and derived data is an explicit goal of these atlases and is expected to encompass native-resolution images, segmented single-cell data, anonymized clinical metadata and treatment history (for human specimens), genetic information (particularly for animal models) and specification of the protocols used to acquire and process the data. Given the imminent release of the first atlases, an urgent need exists for data and metadata standards consistent with emerging Findable, Accessible, Interoperable, and Reusable (FAIR) standards[5]. In this commentary, we establish the Minimum

Information about Highly Multiplexed Tissue Imaging (MITI) standard and associated data-level definitions; we also discuss the relationship of MITI to existing standards, practical implementations and future developments.

Scope and target audiences

MITI covers biospecimen, reagent, data acquisition and data analysis metadata, as well as data levels for imaging with antibodies, aptamers, peptides, dyes and similar detection reagents (Supplementary Table 1). The standard is also compatible with images based on H&E staining, low-plex immunofluorescence (IF) and IHC. A working group is currently extending MITI to cover subcellular-resolution imaging of nucleic acids using methods such as MERFISH[6]. Although conceived with today's 2D images in mind (typically involving 5–10-μm-thick sections of fixed or frozen specimens), MITI accommodates three-dimensional (3D) datasets acquired using confocal, deconvolution and light-sheet microscopes[7]. MITI has been established as its own organization with its own GitHub repository, governing structure and procedures for proposing and incorporating revisions. The definition of MITI is available in the machine-readable YAML format (https://github.com/miti-consortium/MITI) along with other relevant information. MITI has also been implemented in practice (https://github.com/ncihtan/data-models) and used to structure metadata available via the HTAN data portal (https://htan-portal-nextjs.vercel.app).

However, MITI is independent of HTAN or any single research consortium.

Highly multiplexed imaging is derived from methods such as IHC and IF that are in widespread use in preclinical research using cultured cells and model organisms, and in clinical practice with human tissue specimens. Many standards and best practices have been established for these types of data (Supplementary Table 2), but high-plex imaging presents unique challenges: images are expensive to collect and can be very large (up to 1 TB in size), specimens are often difficult to acquire and may have data use restrictions, and accurate clinical and genomic annotation is a necessity. Recent interest in highly multiplexed tissue imaging has been driven by applications in oncology, largely due to the importance of the tumor microenvironment in immuno-editing and responsiveness to immunotherapy, but the approach is broadly applicable to studying normal development, infectious disease, immunology and other topics. HuBMAP[3], for example, is using high-plex imaging to study a range of normal human tissues. MITI is also relevant to studies with model organisms, and data tables have already been created to store data from genetically engineered mouse models (GEMMs) in a standardized manner.

Multiplexed imaging also promises to influence the pathological diagnosis of diseases, which is rapidly switching to digital approaches[8]. For over a century, histological analysis of anatomic specimens (from biopsies and surgical resection) has
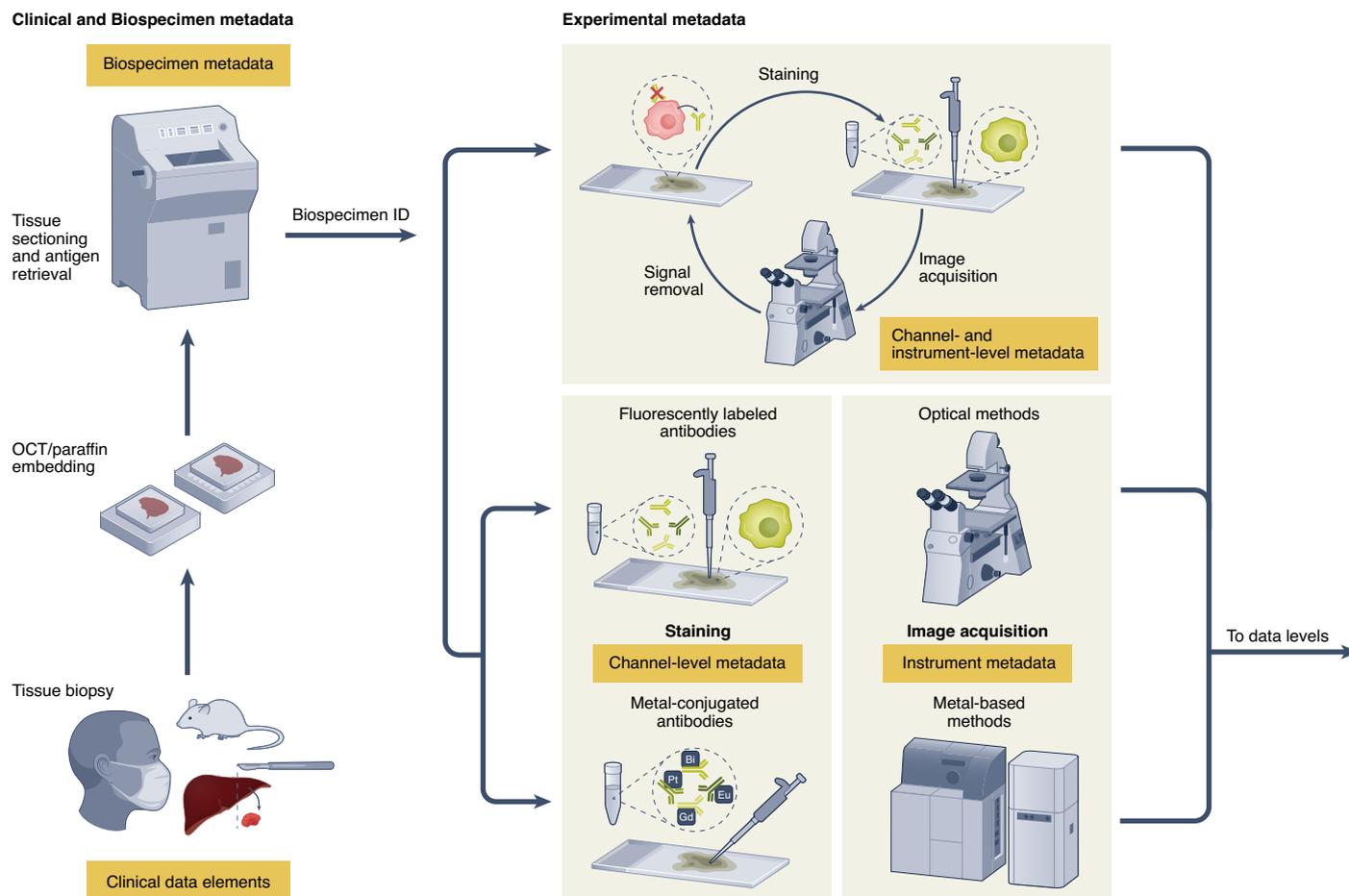
Fig. 1 | The steps in a canonical multiplexed tissue imaging experiment and the associated metadata. In a typical workflow, samples collected from patient biopsies and resections or from animal models are formaldehyde fixed and paraffin embedded or frozen, and are then sectioned and mounted onto either a standard glass microscope slide (for CyCIF, mIHC, IMC, MELC or mxIF), fluidic chamber (for CODEX) or specialized carriers (for MIBI). Clinical and biospecimen metadata (extracted from clinical records, for example) are linked to all other levels of metadata via a unique ID (Biospecimen ID). Data are acquired using cyclical or noncyclical staining and imaging methods, and both reagent and experimental metadata (consisting of antibody, reagent and instrument metadata) are collected. In both cyclic and noncyclic methods, sections undergo preprocessing, antigen retrieval and antibody incubation, and images are acquired. In cyclic imaging methods, fluorophores or chromogens are inactivated or removed and additional antibodies and/or visualization reagents are applied and data acquisition repeated. Channel and instrument metadata capture these essential details.

been the primary method to diagnose diseases such as cancer[9], and this remains true today, despite the impact of gene sequencing. Multiplexed tissue imaging promises to augment such conventional pathological diagnosis with the detailed molecular information needed to specify the use of contemporary precision therapies. This is therefore an opportune time to seek alignment of research and diagnostic approaches by establishing public standards able to take full advantage of the detailed molecular information revealed by emerging imaging methods.
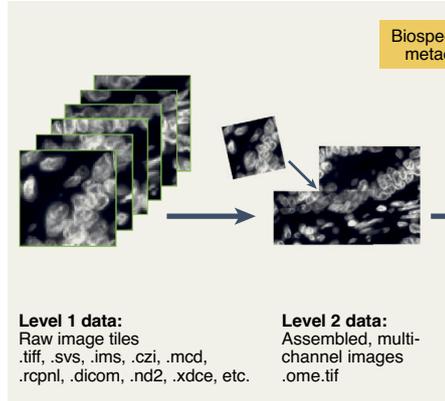
Existing standards and approaches
The Human Genome Project, the Cancer Genome Atlas (TCGA)[10] and similar large-scale genomic programs have

developed several approaches to data management that are of immediate relevance to tissue atlases. The first is the concept of "minimum information" metadata, which has been employed in microarrays (the MIAME standard)[11], genome sequences (MIGS)[12] and biological investigation in general (MIBBI)[13]. The second is the idea of "data levels" (https://gdc.cancer.gov/resources/tcga-users/tcga-code-tables/data-levels), which specify the extent of data processing (raw, normalized, aggregated or region of interest, corresponding to data levels 1–4) and access control. Access control is required because even anonymized DNA sequencing data pose a risk of re-identification[14]. As a result, the database of Genotypes and Phenotypes (dbGaP), the US National Cancer Institute (NCI) Genome

Data Commons (GDC)[15] and the US Federal Register (79 FR 51345) control access to primary sequencing data (so-called level 1 and 2 sequencing data) based on policies set by a data access committee. Higher-level genomic data, which are generally more consolidated, involve information aggregated from many patients and pose little or no re-identification risk, can be freely shared[16] (Fig. 2). When datasets are combined, they acquire the most stringent restriction applied to any constituent element. Although we are not aware of any policies addressing the anonymity of histological images, consultation with our Institutional Review Boards (IRBs, or ethics committees) has led us to conclude that public release of tissue images does not constitute a risk to patient privacy.

a MITI data levels 1 and 2 (access may be limited):

b MITI data levels 3–5 (unrestricted access):

Biospecimen metadata

Level 1 data:
Raw image tiles
.tiff, .svs, .ims, .czi, .mcd,
.rcpnl, .dicom, .nd2, .xdce, etc.

Level 2 data:
Assembled, multi-
channel images
.ome.tif

Level 3 data:
Quality controlled, assembled
images, segmentation masks
.ome.tif

Level 4 data:
Spatial feature table
.fcs, .csv, .h5ad

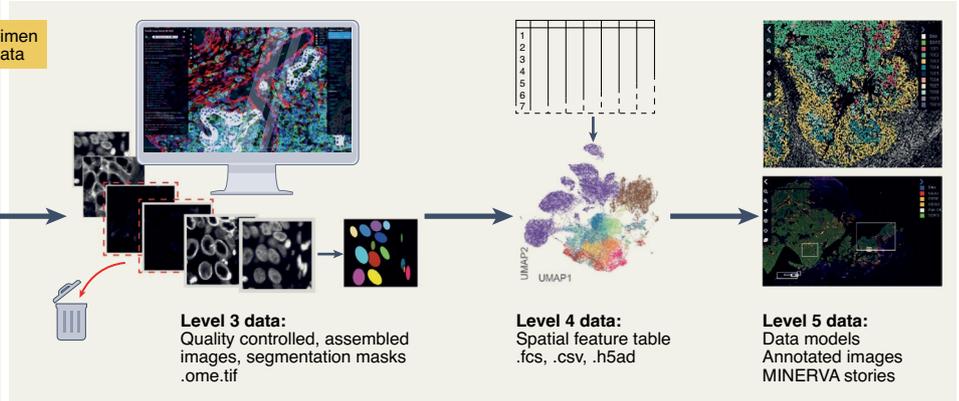**Level 5 data:**
Data models
Annotated images
MINERVA stories

Fig. 2 | MITI data levels and formats. Data levels specify the extent of data processing and, in the case of sequencing data, whether access requires the approval of a data access committee. In common practice, data at levels 3 and up are freely shared. Primary data arising from microscopes and data acquisition instruments correspond to level 1 data. Because the raw image data acquired from one slide usually consist of separate image fields, possibly from proprietary formats, they are processed to correct for uneven illumination and other instrumentation artifacts and assembled into a single multichannel image in the OME-TIFF format (level 2 data). OME-TIFF image mosaics undergo quality control (including artifact removal, channel rejection and evaluation of staining quality) to generate full-resolution, assembled and curated level 3 image data; segmentation algorithms generate one or more label masks that also comprise level 3 data. The great majority of users will want to access these level 3 images. Each label mask (for example, nuclei, cytoplasmic regions, whole cells, organelles, etc.) is used to compute quantitative features, such as the mean signal intensity, spatial coordinates of individual cells and morphological features, which are stored as level 4 spatial feature tables (where rows represent single cells and columns the extracted cellular features); these data are suitable for analysis using the dimensionality reduction and visualization tools used for other types of single-cell data (for example, UMAP plots). Spatial models computed from images and spatial feature tables or by direct application of machine learning to images, as well as images annotated by humans, comprise level 5 data.

MITI data levels are nonetheless consistent with the existing GDC and dbGaP practice whereby data intended for unrestricted distribution are classified as level 3 and up. In the case of images adhering to the MITI standard, level 3 data have been subjected to quality control and some degree of human annotation, making them more useful in a shared environment than raw images. We anticipate that IRBs and government agencies will in the future provide further guidance on the sharing of datasets that combine clinical history, sequence information and tissue images; MITI will be adapted to accommodate such guidance.

The MITI standard also draws extensively on image formats developed for cultured cells and model organisms and on a wide variety of open-source software tools (Supplementary Table 3). Noteworthy among these are the Open Microscopy Environment (OME) TIFF standard[17] and the BioFormats[18] approach to standardization of microscopy data. MITI field definitions are harmonized with the QUality Assessment and REProducibility for Instruments and Images in Light Microscopy (QUAREP-LiMi)[19] effort, the Resource Identification Initiative[20] and antibody standardization efforts by the Human Protein Atlas[21] and are also compliant with the recently developed Recommended Metadata for Biological Images initiative[22]. Metadata on model organisms (particularly

GEMMs and patient-derived xenografts (PDXs)) are aligned with existing standards, many developed for genomic information (see Supplementary Table 2 for a full list of antecedent resources). Well-curated clinical information is essential for the interpretation of data from human specimens, but standardizing such information has proven to be a major challenge in the past, for example in TCGA[23,34]. Thus, HTAN and other current NCI projects focused on human specimens are emphasizing standardization of clinical metadata, and the MITI standard is designed to closely align with the GDC Data Model[24] in this regard (Supplementary Tables 5 and 6).

All imaging methods generate data that comprise a sequence of intensity values on a raster; multispectral imaging simply adds new dimensions to the raster. The cameras that collect H&E and IHC images from bright-field microscopes or high-plex images from fluorescence microscopes generate rasters; ablation-based mass-spectrometry imaging (for example, multiplexed ion beam imaging (MIBI) and imaging mass cytometry (IMC)) is also raster based. As currently defined, MITI specifies that raster images should be stored in the OME-TIFF 6 standard, but OME formats are currently being migrated to a set of next-generation file formats (collectively called OME-NGFF)[25] to improve their scalability and performance

in the cloud. MITI will be updated to align with these new formats as they come into general use. Another area of translational and clinical research in which imaging is commonly encountered is radiology, which is almost entirely digital, and which uses data interchange standards governed by the Digital Imaging and Communications in Medicine (DICOM) standard (https://www.dicomstandard.org/). DICOM has recently been extended to accommodate both radiology data and OME-TIFF standards[26]. The NCI's ongoing program to create an Imaging Data Commons[27] is expected to be based on this dual standard, or on a successor using OME-NGFF. MITI is, or will be, compatible with these foundational data standards.

In highly multiplexed tissue imaging, antibodies either are conjugated to fluorophores directly or via oligonucleotides or are bound to secondary antibodies (Fig. 1, Supplementary Table 4). Images are then acquired serially, one to six channels at a time, to assemble data from 20–60 antibodies. In ablation-based methods, antibodies are labeled with metals and vaporized with lasers or ion beams, after which they are detected by atomic mass spectrometry (Supplementary Table 4). In all cases, the raw output of data acquisition instruments comprises level 1 MITI data (Fig. 2), analogous to the level 1 FASTq files in genomics.

Whole-slide imaging is required for clinical applications[28] and also necessary to ensure adequate power in preclinical studies[29]. However, resolution and field of view have a reciprocal relationship—with respect both to optical physics and to the practical process of mapping image fields onto the fixed raster of a camera (or ablating beam). Whole-slide images of histological specimens[8] must therefore be acquired by dividing a large specimen into contiguous tiles. This usually involves acquisition of ~100–1,000 tiles by moving the microscope stage in both the x and y dimensions, with each tile being a multidimensional, subcellular-resolution TIFF image. Tiles are combined at subpixel accuracy into a mosaic image, in a process known as stitching. When high-plex images are assembled from multiple rounds of lower-plex imaging, it is also necessary to register channels to each other across imaging cycles and to correct for any unevenness in illumination (so-called flat-fielding)[30]. Stitched and registered mosaics can be as large as 50,000 × 50,000 pixels × 100 channels and require ~500 GB of disk space. They correspond to level 2 MITI data and represent full-resolution primary images that have undergone automated stitching, registration, illumination correction, background subtraction and intensity normalization and have been stored in a standardized OME format. The level of processing is analogous to that of BAM files, a common type of level 2 data in genomics.

Level 3 data represent images that have been processed with some interpretive intent, which may include (i) full-resolution images following quality control or artifact removal, (ii) segmentation masks computed from such images, (iii) machine-generated spatial models and (iv) images with human- or machine-generated annotations. Level 3 MITI data are roughly analogous to level 3 mRNA expression data in genomics. However, whereas many users of genomic data only require access to processed level 3 and 4 data, which are usually quite compact, quantitative analysis of tissue images adds a requirement for full-resolution primary images so that images and computed features can be examined in parallel[31]. Level 3 MITI data are intended to be the primary type of image data distributed by tissue atlases and similar projects.

Assembled level 3 images are typically segmented to identify single cells[31], which are quantified to produce a 'spatial feature table' that describes marker intensities, cell coordinates and other single-cell features. The level 4 data in spatial feature tables are a natural complement to count tables in single-cell sequencing data (for

example, from scRNA-seq, scATAC-seq or scDNA-seq) and can be analyzed using many of the same dimensionality reduction methods (for example, PCA, t-SNE and U-MAP)[32] and online browsers such as cellxgene (Supplementary Table 3)[33]. These types of tabular data are all examples of 'feature observation matrixes', which are themselves being standardized across domains of biology to improve their utility and intercompatibility. Level 5 MITI data comprise results computed from spatial feature tables or primary images. Because access to terabyte-size full-resolution image data is impractically burdensome when reading a manuscript or browsing a large dataset, a specialized type of level 5 image data has been developed to enable panning and zooming across images using a standard web browser[35]. In the case of level 5 images viewed with MINERVA software, the aim is to exploit functionality and concepts similar to those in Google Maps or electronic museum guides. The inclusion of digital docents with images makes it possible to combine pan and zoom with guided narratives that greatly facilitate comprehension of complex datasets and promote new hypothesis generation[35].

For any metadata standard to be used, a balance must be struck between ease of data entry, which minimizes noncompliance by data generators, and level of detail, which must be sufficient for data retrieval, analysis and publication in a reproducible manner. Moreover, specification of a metadata standard is separate from the essential task of developing a practical and reliable means for capturing information needed to ensure adherence to the standard. Two approaches have proven most effective in addressing this requirement. One, exemplified by OMeta[36], involves a relational database and web interface that data generators use to input necessary information in a controlled manner. Another approach, exemplified by MAGE-TAB[37], involves a standardized format for collecting metadata via a series of structured documents, which are then used to populate web pages and databases[38]. As a practical test of MITI, we have implemented the latter approach in a JSON schema (https://github.com/ncihtan/data-models) that also conforms to the design principles of SCHEMA.org. These principles focus on the creation, maintenance and promotion of schemas for structured data that are supported by major web search engines, thereby enhancing discoverability. In this TAB-like approach, the MITI standard is exposed to data collectors as Google Sheets with dropdowns representing controlled vocabularies and highlighting required or optional elements;

many fields are automatically validated upon entry. These documents are ingested using SCHEMATIC (Schema Engine for Manifest Ingress and Curation; https://github.com/Sage-Bionetworks/schematic), automatically linked to primary imaging data and stored as cloud assets. These implementations continue to evolve, and entirely different approaches are possible: nothing in a MITI-type standard constrains how data are collected.

Whereas many research agencies and countries have made a major investment in curating, storing and distributing genomic data, fewer repositories exist for primary image data. The Image Data Resource[39] maintained by the European Bioinformatics Institute (EBI) is an exception, but as the volume of image data grows, other means of data distribution will almost certainly be required. In the US, in the absence of a major public investment in data storage, the development of "requester pays"[40] access to datasets is a promising development. The primary cost associated with the creation and maintenance of a dataset on a commercial cloud service involves data download, not data ingress and storage. In a 'requester pays' model, a user seeking access to a dataset pays the cost of data egress directly to the cloud provider, making access both secure and anonymous (moreover, the cost of egress into another account on the same commercial cloud is low). Although this approach might appear to create an impediment to research, the actual cost of egress is quite low (currently, about US $100/TB) compared to any form of data acquisition, and a key goal is to avoid a tragedy of the commons in which frequent, duplicate downloads overwhelm the system. A combination of a MITI implementation on a cloud service (as described above) with 'requester pays' cloud access will also make it possible for individuals to distribute very large FAIR image datasets at relatively low cost. Such an approach does not obviate the need for public investments, such as those being made by the EBI, but it does represent a practical way forward for democratizing the release of standardized data—some of which can then be incorporated into publicly supported resources. Regardless, the MITI standard described here is available for immediate use, without being affected by how access to the primary data is provisioned.

Public data and metadata standards have been essential for the success of genomics and other fields of biomedicine, but the creation of a new standard is no guarantee of successful adoption. An outpouring of effort 10–20 years ago led to the development of widely adopted

and well-maintained standards such as MIAME[11], MIGS[12] and MIBBI[13], and these have been consolidated and further documented by the Digital Curation Center (https://www.dcc.ac.uk/), FairSharing.org and similar projects. However, many other minimum information projects have been left unattended[41], and it remains unclear whether existing metadata adequately conform to user needs[42]. The development of MITI and of the initial HTAN implementation enjoys NCI support and is expected to become part of the NCI Cancer Research Data Commons[27], helping ensure its viability. However, individuals and organizations are invited to join in the further development of MITI and should make contact via the image.sc forum or submit pull requests (that is, requests for inclusion in the MITI 'code base' at https://github.com/miti-consortium/MITI). Because high-plex tissue imaging is in its infancy and MITI has attracted the great majority of developers of existing high-plex tissue image acquisition methods, it represents a solid beginning for what will need to be an evolving standard. By having its own repository and governance structure, independent of any particular research program or constituency, MITI also conforms with other requirements of successful open standards[43].

Data availability
The detailed specification of the guidelines outlined in this manuscript are available at https://github.com/miti-consortium/MITI and https://www.miti-consortium.org/.

Code availability
The detailed specification of the guidelines outlined in this manuscript are available at https://github.com/miti-consortium/MITI and https://www.miti-consortium.org/. □

Denis Schapiro[1,2,3,4,41], Clarence Yapp[1,5,41], Artem Sokolov[1,6,41], Sheila M. Reynolds[7], Yu-An Chen[1], Damir Sudar[8], Yubin Xie[9], Jeremy Muhlich[1], Raquel Arias-Camison[1], Sarah Arena[1], Adam J. Taylor[10], Milen Nikolov[10], Madison Tyler[1], Jia-Ren Lin[1], Erik A. Burlingame[11,40], Human Tumor Atlas Network*, Young H. Chang[11], Samouil L. Farhi[2], Vésteinn Thorsson[7], Nithya Venkatamohan[12], Julia L. Drewes[13], Dana Pe'er[9], David A. Gutman[14], Markus D. Herrmann[15], Nils Gehlenborg[6], Peter Bankhead[16], Joseph T. Roland[17], John M. Herndon[18], Michael P. Snyder[19], Michael Angelo[19], Garry Nolan[19], Jason R. Swedlow[20], Nikolaus Schultz[21], Daniel T. Merrick[22], Sarah A. Mazzili[23], Ethan Cerami[12], Scott J. Rodig[24], Sandro Santagata[1,24✉] and Peter K. Sorger[1,25✉]

[1]Laboratory of Systems Pharmacology, Ludwig Center for Cancer Research at Harvard, Harvard Medical School, Boston, MA, USA. [2]Klarman Cell Observatory, Broad Institute of MIT and Harvard, Cambridge, MA, USA. [3]Institute for Computational Biomedicine, Faculty of Medicine, Heidelberg University Hospital and Heidelberg University, Heidelberg, Germany. [4]Institute of Pathology, Heidelberg University Hospital, Heidelberg, Germany. [5]Image and Data Analysis Core, Harvard Medical School, Boston, MA, USA. [6]Department of Biomedical Informatics, Harvard Medical School, Boston, MA, USA. [7]Institute for Systems Biology, Seattle, WA, USA. [8]Quantitative Imaging Systems LLC, Portland, OR, USA. [9]Program in Computational and Systems Biology, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [10]Sage Bionetworks, Seattle, WA, USA. [11]Oregon Health and Science University, Portland, OR, USA. [12]Dana-Farber Cancer Institute, Boston, MA, USA. [13]Johns Hopkins University School of Medicine, Baltimore, MD, USA. [14]School of Medicine, Emory University, Atlanta, GA, USA. [15]Department of Pathology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. [16]Edinburgh Pathology, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh, UK. [17]Vanderbilt University School of Medicine, Nashville, TN, USA. [18]Department of Surgery, Washington University School of Medicine, St. Louis, MO, USA. [19]School of Medicine, Stanford University, Stanford, CA, USA. [20]Division of Computational Biology and Centre for Gene Regulation and Expression, University of Dundee, Dundee, UK. [21]Department of Epidemiology & Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY, USA. [22]Pathology, University of Colorado, Aurora, CO, USA. [23]Boston University, Boston, MA, USA. [24]Department of Pathology, Brigham and Women's Hospital, Boston, MA, USA. [25]Department of Systems Biology, Harvard Medical School, Boston, MA, USA. [40]Present address: Indica Labs, Albuquerque, NM, USA. [41]These authors contributed equally: Denis Schapiro, Clarence Yapp, Artem Sokolov. *A list of authors and their affiliations appears at the end of the paper.
✉e-mail: ssantagata@bics.bwh.harvard.edu; peter_sorger@hms.harvard.edu

Human Tumor Atlas Network

Daniel L. Abravanel[26], Samuel Achilefu[18], Foluso O. Ademuyiwa[18], Andrew C. Adey[11], Rebecca Aft[18], Khung Jun Ahn[27], Fatemeh Alikarami[27], Shahar Alon[28], Orr Ashenberg[1], Ethan Baker[1], Gregory J. Baker[29], Shovik Bandyopadhyay[30], Peter Bayguinov[18], Jennifer Beane[23], Winston Becker[19], Kathrin Bernt[27], Courtney B. Betts[11], Julie Bletz[10], Tim Blosser[31], Adrienne Boire[9], Genevieve M. Boland[26], Edward S. Boyden[28], Elmar Bucher[11], Raphael Bueno[26], Qiuyin Cai[17], Francesco Cambuli[9], Joshua Campbell[23], Song Cao[18], Wagma Caravan[18], Ronan Chaligné[9], Joseph M. Chan[9], Sara Chasnoff[18], Deyali Chatterjee[18], Alyce A. Chen[29], Changya Chen[27], Chia-hui Chen[27], Bob Chen[17], Feng Chen[18], Siqi Chen[18], Milan G. Chheda[18], Koei Chin[11], Hyeyoung Cho[11], Jaeyoung Chun[9], Luis Cisneros[32], Robert J. Coffey[17], Ofir Cohen[1], Graham A. Colditz[18], Kristina A. Cole[27], Natalie Collins[16], Daniel Cotter[19], Lisa M. Coussens[11], Shannon Coy[24], Allison L. Creason[11], Yi Cui[28], Daniel Cui Zhou[18], Christina Curtis[19], Sherri R. Davies[18], Inode Bruijn[9], Toni M. Delorey[1], Emek Demir[11], David Denardo[18], Dinh Diep[33], Li Ding[18], John DiPersio[18], Steven M. Dubinett[34], Timothy J. Eberlein[18], James A. Eddy[10], Edward D. Esplin[19], Rachel E. Factor[35], Kayvon Fatahalian[19], Heidi S. Feiler[11], Jose Fernandez[19], Andrew Fields[11], Ryan C. Fields[18], James A. J. Fitzpatrick[18], James M. Ford[19], Jeff Franklin[17], Bob Fulton[18], Giorgio Gaglia[24], Luciano Galdieri[18], Karuna Ganesh[9], Jianjiong Gao[9], Benjamin L. Gaudio[29], Gad Getz[1], David L. Gibbs[7], William E. Gillanders[18], Jeremy Goecks[11], Daniel Goodwin[28], Joe W. Gray[11], William Greenleaf[19], Lars J. Grimm[35], Qiang Gu[11], Jennifer L. Guerriero[16], Tuhin Guha[19], Alexander R. Guimaraes[11], Belen Gutierrez[19], Nir Hacohen[36], Casey Ryan Hanson[19], Coleman R. Harris[17], William G. Hawkins[18], Cody N. Heiser[17], John Hoffer[29], Travis J. Hollmann[9], James J. Hsieh[18], Jeffrey Huang[27], Stephen P. Hunger[27], Eun-Sil Hwang[35], Christine Iacobuzio-Donahue[9], Michael D. Iglesia[18], Mohammad Islam[27], Benjamin Izar[16], Connor A. Jacobson[29], Samuel Janes[37], Reyka G. Jayasinghe[18], Tiarah Jeudi[16], Bruce E. Johnson[16], Brett E. Johnson[11], Tao Ju[18], Humam Kadara[38], Elias-Ramzey Karnoub[9], Alla Karpova[18], Aziz Khan[19], Warren Kibbe[35], Albert H. Kim[18], Lorraine M. King[35], Elyse Kozlowski[16], Praveen Krishnamoorthy[18], Robert Krueger[31], Anshul Kundaje[19], Uri Ladabaum[19], Rozelle Laquindanum[19], Clarisse Lau[7], Ken Siu Kwong Lau[17], Nicole R. LeBoeuf[26], Hayan Lee[19], Marc Lenburg[23], Ignaty Leshchiner[1], Rochelle Levy[16], Yize Li[18], Christine G. Lian[24], Wen-Wen Liang[18], Kian-Huat Lim[18], Yiyun Lin[38], David Liu[16], Qi Liu[17], Ruiyang Liu[18], Joseph Lo[35], Pierrette Lo[10], William J. Longabaugh[7], Teri Longacre[19], Katie Luckett[9], Cynthia Ma[18], Chris Maher[18], Allison Maier[29], Danika Makowski[27], Carlo Maley[32], Zoltan Maliga[29], Parvathy Manoj[9], John M. Maris[27], Nick Markham[17], Jeffrey R. Marks[35], Daniel Martinez[27],

Jay Mashl[18], Ignas Masilionis[9],
Joan Massague[9], Marciej A. Mazurowski[35],
Eliot T. McKinley[17], Joshua McMichael[18],
Matthew Meyerson[16], Gordon B. Mills[11],
Zahi I. Mitri[11], Andrew Moorman[9],
Jacqueline Mudd[18], George F. Murphy[24],
Nataly Naser Al Deen[18], Nicholas E. Navin[38],
Tal Nawy[9], Reid M. Ness[17], Stephanie Nevins[19],
Ajit Johnson Nirmal[16], Edward Novikov[29],
Stephen T. Oh[18], Derek A. Oldridge[30],
Kouros Owzar[35], Shishir M. Pant[29],
Wungki Park[9], Gary J. Patti[18], Kristina Paul[19],
Roxanne Pelletier[29], Daniel Persson[11],
Candi Petty[18], Hanspeter Pfister[31],
Kornelia Polyak[16], Sidharth V. Puram[18], Qi Qiu[30],
Álvaro Quintanal Villalonga[9],
Marisol Adelina Ramirez[17], Rumana Rashid[24],
Ashley N. Reeb[18], Mary E. Reid[39], Jan Remsik[9],
Jessica L. Riesterer[11], Tyler Risom[19],
Cecily Claire Ritch[24], Andrea Rolong[17],
Charles M. Rudin[9], Marc D. Ryser[35],
Kazuhito Sato[18], Cynthia L. Sears[17],
Yevgeniy R. Semenov[36], Jeanne Shen[19],
Kooresh I. Shoghi[18], Martha J. Shrubsole[17],
Yu Shyr[17], Alexander B. Sibley[35],
Alan J. Simmons[17], Anubhav Sinha[28],
Shamilene Sivagnanam[11], Sheng-Kwei Song[18],
Austin Southar-Smith[18], Avrum E. Spira[23],
Jeremy St. Cyr[16], Stephanie Stefankiewicz[27],
Erik P. Storrs[18], Elizabeth H. Stover[16],
Siri H. Strand[19], Cody Straub[35],
Cherease Street[18], Timothy Su[17], Lea F. Surrey[27],
Christine Suver[10], Kai Tan[27],
Nadezhda V. Terekhanova[18], Luke Ternes[11],
Anusha Thadi[27], George Thomas[11],
Rob Tibshirani[19], Shigeaki Umeda[9],
Yasin Uzun[27], Tuulia Vallius[29],
Eliezer R. Van Allen[16], Simon Vandekar[17],
Paige N. Vega[17], Deborah J. Veis[18],
Sujay Vennam[19], Ana Verma[24],
Sebastien Vigneau[16], Nikhil Wagle[16],
Richard Wahl[18], Thomas Walle[9],
Liang-Bo Wang[18], Simon Warchol[31],
M. Kay Washington[17], Cameron Watson[11],
Annika K. Weimer[19], Michael C. Wendl[18],
Robert B. West[19], Shannon White[19],
Annika L. Windon[17], Hao Wu[30], Chi-Yun Wu[30],
Yige Wu[18], Matthew A. Wyczalkowski[18],
Jason Xu[30], Lijun Yao[18], Wenbao Yu[27],
Kun Zhang[33] and Xiangzhu Zhu[17]

[26]Brigham and Women's Hospital, Boston, MA,
USA. [27]The Children's Hospital of Philadelphia,
Philadelphia, PA, USA. [28]Massachusetts Institute
of Technology, Cambridge, MA, USA. [29]Harvard
Medical School, Boston, MA, USA. [30]Perelman
School of Medicine, University of Pennsylvania,
Philadelphia, PA, USA. [31]Harvard University,
Cambridge, MA, USA. [32]Arizona State University,
Tempe, AZ, USA. [33]University of California, San
Diego, San Diego, CA, USA. [34]David Geffen School
of Medicine at UCLA, Los Angeles, CA, USA. [35]Duke
University Medical Center, Durham, NC, USA.
[36]Massachusetts General Hospital, Boston, MA, USA.
[37]Division of Medicine, University College London,
London, UK. [38]MD Anderson Cancer Center,
Houston, TX, USA. [39]Roswell Park Comprehensive
Cancer Center, Buffalo, NY, USA.

References

1. Bodenmiller, B. *Cell Syst* **2**, 225–238 (2016).
2. Rozenblatt-Rosen, O. et al. *Cell* **181**, 236–249 (2020).
3. HuBMAP Consortium. *Nature* **574**, 187–192 (2019).
4. Rajewsky, N. et al. *Nature* **587**, 377–386 (2020).
5. Wilkinson, M. D. et al. *Sci. Data* **3**, 160018 (2016).
6. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. *Science* **348**, aaa6090 (2015).
7. Fischer, R. S., Wu, Y., Kanchanawong, P., Shroff, H. & Waterman, C. M. *Trends Cell Biol* **21**, 682–691 (2011).
8. Ghaznavi, F., Evans, A., Madabhushi, A. & Feldman, M. *Annu. Rev. Pathol.* **8**, 331–359 (2013).
9. Amin, M. B. et al. *CA Cancer J. Clin* **67**, 93–99 (2017).
10. The Cancer Genome Atlas Research Network. et al. *Nat. Genet.* **45**, 1113–1120 (2013).
11. Brazma, A. et al. *Nat. Genet.* **29**, 365–371 (2001).
12. Field, D. et al. *Nat. Biotechnol.* **26**, 541–547 (2008).
13. Taylor, C. F. et al. *Nat. Biotechnol.* **26**, 889–896 (2008).
14. Benitez, K. & Malin, B. *J. Am. Med. Inform. Assoc.* **17**, 169–177 (2010).
15. Grossman, R. L. et al. *N. Engl. J. Med* **375**, 1109–1112 (2016).
16. Byrd, J. B., Greene, A. C., Prasad, D. V., Jiang, X. & Greene, C. S. *Nat. Rev. Genet.* **21**, 615–629 (2020).
17. Swedlow, J. R., Goldberg, I., Brauner, E. & Sorger, P. K. *Science* **300**, 100–102 (2003).
18. Li, S. et al. *Methods* **96**, 27–32 (2016).
19. Nelson, G. et al. *J. Microscopy* **284**, 56–73 (2021).
20. Bandrowski, A. et al. *F1000 Res.* **4**, 134 (2015).
21. Edfors, F. et al. *Nat. Commun.* **9**, 4130 (2018).
22. Sarkans, U. et al. *Nat. Methods* **18**, 1418–1422 (2021).
23. Liu, J. et al. *Cell* **173**, 400–416.e11 (2018).
24. Zhang, Z. et al. *Nat. Commun.* **12**, 1226 (2021).
25. Moore, J. et al. *Nat. Methods* **18**, 1496–1498 (2021).
26. Clunie, D. A. *J. Pathol. Inform.* **10**, 12 (2019).
27. Fedorov, A. et al. *Cancer Res.* **81**, canres.0950.2021 (2021).
28. Center for Devices and Radiologic Health. *US Food and Drug Administration* http://www.fda.gov/regulatory-information/search-fda-guidance-documents/technical-performance-assessment-digital-pathology-whole-slide-imaging-devices (2019).
29. Lin, J.-R. et al. Preprint at bioRxiv https://doi.org/10.1101/2021.03.31.437984 (2021).
30. Peng, T. et al. *Nat. Commun.* **8**, 14836 (2017).
31. Schapiro, D. et al. *Nat. Methods*, https://doi.org/10.1038/s41592-021-01308-y (2021).
32. Heiser, C. N. & Lau, K. S. *Cell Rep* **31**, 107576 (2020).
33. Megill, C. et al. Preprint at bioRxiv https://doi.org/10.1101/2021.04.05.438318 (2021).
34. Gutman, D. A. et al. *Cancer Res* **77**, e75–e78 (2017).
35. Rashid, R. et al. *Nat. Biomed. Eng.* https://doi.org/10.1038/s41551-021-00789-8 (2021).
36. Singh, I. et al. *BMC Bioinformatics* **20**, 8 (2019).
37. Rayner, T. F. et al. *BMC Bioinformatics* **7**, 489 (2006).
38. Martínez-Romero, M. et al. *Database (Oxford)* **2019**, baz059 (2019).
39. Williams, E. et al. *Nat. Methods* **14**, 775–781 (2017).
40. Amazon Web Services. https://docs.aws.amazon.com/AmazonS3/latest/userguide/RequesterPaysBuckets.html (accessed 11 February 2022).
41. Kacprzak, E. et al. *J. Web Semant.* **55**, 37–55 (2019).
42. Löffler, F., Wesp, V., König-Ries, B. & Klan, F. *PLoS One* **16**, e0246099 (2021).
43. Swedlow, J. R. et al. *Nat. Methods* **18**, 1440–1446 (2021).

Author contributions

D. Schapiro, C.Y. and A.S. initiated and implemented the MITI guidelines with extensive guidance from other authors and direct supervision by P.K.S. and S.S. All authors contributed to and reviewed the final MITI guidelines. D. Schapiro, C.Y, A.S., P.K.S. and S.S. wrote the manuscript with input from all authors.

Competing interests

P.K.S. is a member of the Scientific Advisory Board (SAB) or Board of Directors of Applied Biomath, RareCyte Inc. and Glencoe Software, which distributes a commercial version of the OMERO data management platform; P.K.S. is also a member of the NanoString SAB and a consultant to Merck and Montai Health. In the past five years, the Sorger laboratory has received research funding from Novartis and Merck. P.K.S. declares that none of these relationships have influenced the content of this manuscript. S.S. is a consultant for RareCyte Inc. N.G. is a co-founder and equity owner of Datavisyn. D. Schapiro is a consultant for Roche Glycart AG. J.R.S. is founder and CEO of Glencoe Software, which distributes a commercial version of the OMERO data management platform. S.J.R. receives research funding from Bristol Myers Squibb, Merck, Affimed and Kite/Gilead and is on the SAB of Immunitas Therapeutics. D. Sudar is employed by Quantitative Imaging Systems LLC. E.A.B. is an employee of Indica Labs.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41592-022-01415-4.

Peer review information *Nature Methods* thanks Matthew Hartley and Anja Hauser for their contribution to the peer review of this work.