



Published in final edited form as:

Nat Biomed Eng. 2022 May ; 6(5): 515–526. doi:10.1038/s41551-021-00789-8.

## Narrative online guides for the interpretation of digital-pathology images and tissue-atlas data

Rumana Rashid<sup>1,2,3</sup>, Yu-An Chen<sup>1</sup>, John Hoffer<sup>1</sup>, Jeremy L. Muhlich<sup>1</sup>, Jia-Ren Lin<sup>1,2</sup>, Robert Krueger<sup>1,4</sup>, Hanspeter Pfister<sup>4</sup>, Richard Mitchell<sup>3</sup>, Sandro Santagata<sup>1,2,3,\*</sup>, Peter K. Sorger<sup>1,2,5,\*</sup>

<sup>1</sup>Laboratory of Systems Pharmacology, Harvard Medical School, Boston, MA

<sup>2</sup>Ludwig Center for Cancer Research at Harvard, Harvard Medical School, Boston, MA

<sup>3</sup>Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA

<sup>4</sup>School of Engineering and Applied Sciences, Harvard University, Cambridge, MA

<sup>5</sup>Department of Systems Biology, Harvard Medical School, Boston, MA

### Abstract

Multiplexed tissue imaging facilitates the diagnosis and understanding of complex disease traits. However, the analysis of such digital images heavily relies on the experience of anatomical pathologists for the review, annotation, and description of tissue features. In addition, the wider use of data from tissue atlases in basic and translational research and in classrooms would benefit from software that facilitates the easy visualization and sharing of the images and the results of their analyses. In this Perspective, we describe the ecosystem of software available for the analysis of tissue images and discuss the need for interactive online guides that help histopathologists make complex images comprehensible to non-specialists. We illustrate this idea via a software interface (Minerva), accessible via web browsers, that integrates multi-omic and tissue-atlas features. We argue that such interactive narrative guides can effectively disseminate digital histology data and aid their interpretation.

Diagnostic pathology relies on the microscopic analysis of histological specimens stained with colorimetric dyes or antibodies. Multiplexed high-resolution imaging<sup>1–4</sup> can provide information on the expression levels and subcellular localization of 20–60 proteins, facilitating the study of cell states, tissue biomarkers and cell–cell interactions in normal and diseased conditions, and enabling the molecular profiling of tissues while preserving the native tissue architecture<sup>5</sup>. A common application of tissue imaging in oncology involves the identification and quantification of immune cell types and the mapping of their locations

\* Co-corresponding authors: Peter K. Sorger, Harvard Medical School, 200 Longwood Avenue, Warren Alpert Building, Room 440, Boston, MA 02115; Telephone: 617-432-6901; peter\_sorger@hms.harvard.edu; sorger\_admin@hms.harvard.edu, Sandro Santagata, Brigham and Women's Hospital, 60 Fenwood Road, HBTM 8002P, Boston, MA 02115; Telephone: 617-525-5686; ssantagata@bics.bwh.harvard.edu.

#### Author contributions

R.R., Y.-A.C., J.H., J.L.M. and R.K. contributed to researching information for the writing of this article, to the development of the Minerva software and to curating data. R.M. contributed to the content and discussion. H.P. contributed to data visualization. S.S. and P.K.S. contributed to all aspects of the article. All authors contributed to the writing of the manuscript.

relative to the tumour boundaries and to stromal cells<sup>5</sup>. This type of spatially resolved data is particularly relevant to the study of the mechanisms of action of immunotherapies, such as the immune checkpoint inhibitors targeting the programmed cell death protein 1 (PD-1) or its ligand PD-L1<sup>6</sup>, which function by blocking juxtracrine signalling between immune cells and tumour cells. However, images of tissues can be complex owing to the range of sizes of the biologically relevant structures: from the micrometre scale of subcellular vesicles and nuclear granules to the millimetre scale of cell configurations in blood and lymphatic vessels, to the centimetre scale required to visualize the interactions among endothelia, epithelia, muscle and the other cell types forming functional tissues.

There are a number of multiplexed tissue-imaging methods in current use. Imaging mass cytometry (IMC)<sup>1</sup> and multiplexed ion beam imaging (MIBI)<sup>2</sup> detect antigens via antibodies labelled with metal isotopes, followed by tissue ablation and atomic mass spectrometry. Methods such as multiplexed immunofluorescence (MxIF)<sup>3</sup>, co-detection by indexing (CODEX)<sup>7</sup>, tissue-based cyclic immunofluorescence (t-CyCIF)<sup>4</sup>, multiplexed immunohistochemistry (mIHC)<sup>8,9</sup>, and immunostaining with signal amplification by exchange reaction (immuno-SABER)<sup>10</sup> use fluorescently labelled (or enzyme-linked) antibodies followed by microscopy. These methods differ in the number of antigens that they can detect on a single tissue section (at present, ~12 for multiplexed IHC and ~40–60 for IMC, MxIF, CODEX and t-CyCIF). Some methods are restricted to selected fields of view (in particular, ~1 mm<sup>2</sup> for MIBI and IMC), whereas others can perform whole-slide imaging (WSI) on areas ~100–400 times larger (this is the case for MxIF, t-CyCIF, CODEX and mIHC). Most multiplexed tissue-imaging methods are in active development, and hence their strengths and limitations regarding imaging speed, sensitivity, resolution and other performance parameters are in flux. However, they all generate multiplexed two-dimensional (2D) images of cells and supporting tissue structures in situ. It is also possible to generate three-dimensional (3D) images via optical-sectioning and tissue-clearing methods<sup>11</sup>.

## Tissue and tumour atlases

International projects are currently underway to create publicly accessible atlases of normal human tissues and tumours (Table 1). These include the Human Cell Atlas<sup>12</sup>, the Human BioMolecular Atlas Program (HuBMAP)<sup>13</sup>, and the Human Tumor Atlas Network (HTAN)<sup>14</sup>. For example, HTAN is envisioned as the spatially resolved counterpart of the well-established Cancer Genome Atlas<sup>15</sup> (TCGA) (Figure 1). HTAN atlases aim to integrate genetic and molecular information from dissociative single-cell methods, such as single-cell RNA sequencing, with morphological and spatial details obtained from tissue imaging and spatial transcriptomics<sup>13,14</sup>. These atlases will combine images from many specimens into common reference systems to enable systematic inter-patient and cross-disease analysis of the data<sup>13</sup>. Conceptually, integration across samples and data types is easiest to achieve at the level of derived features, such as a census of cell types and positions (from imaging data) or transcript levels (from single-cell RNA sequencing). Adding mesoscopic-scale information from images—such as the arrangement of supporting stroma, membranes, blood and lymphatic vessels—is more challenging.

Computational methods that automatically extract spatial information from images are still under development; therefore, the practice of annotating slides by anatomical pathologists remains essential. Ongoing research to better understand how pathologists make diagnoses from tissue specimens<sup>16</sup> and to quantify connections between features computed from cellular neighbourhoods and clinical outcome<sup>17</sup> will provide useful insight into the development of automated image-analysis pipelines. However, human oversight of tissue images will remain essential for relating morphology to pathophysiology and for assessing the quality of image-processing algorithms and training machine-learning classifiers. Therefore, any software-based solution aimed at users of digital atlases must enable free exploration of image data while also allowing users to benefit from the expertise of pathologists who currently work almost entirely with physical specimens on glass slides. In this Perspective, we describe software interfaces that meet these considerations. The software interfaces allow users working in research or in clinical practice to browse increasingly large image datasets at low cost and without the need for complex software. The interfaces accomplish this by using 'digital docents' and narration so that users can benefit from the expert description of a complex specimen, while also facilitating access to quantitative information derived from computational analysis. They can also augment reports and publications so that reported imaging data are not restricted to showing selected small fields of view.

## Tissue imaging in a clinical setting

An expansion in the use of multiplexed tissue-imaging methods in research settings is happening at the same time as a transition to digital technologies in clinical histopathology<sup>18</sup>; the two developments are, however, not optimally linked or coordinated (Figure 1). Although genetics is increasingly pertinent to disease diagnosis, particularly in cancer and in inherited diseases, histology and cytology remain the central pillars of routine clinical work for the confirmation or diagnosis of many diseases. In current practice, tissue samples recovered by biopsy or surgical resection are formaldehyde-fixed and paraffin-embedded, sliced into 5- $\mu$ m sections, and stained with haematoxylin and eosin (H&E). Samples for immediate study (used when a patient is undergoing surgery, for example) are frozen, sectioned and stained (these are often called OCT samples, because of the 'optimal cutting temperature' medium in which they are embedded). Liquid haematological samples are spread on a slide to create blood smears, which are also stained with colorimetric dyes (Romanowsky–Giemsa staining). H&E staining imparts a characteristic pink-blue colour on cells and other structures, and pathologists review these specimens using simple bright-field microscopes; other stains are analysed in a similar way. Some clinical samples are also subjected to IHC to obtain information on the expression of one or few protein biomarkers per slide<sup>19</sup>. Although cost-effective and widely used, many histopathology methods were developed over a century ago, and IHC is itself 75 years old<sup>20</sup> (Figure 1). Moreover, diagnoses based on these techniques generally do not capture the depth of molecular information needed to optimally select targeted therapies. The latter remains the purview of mRNA and DNA sequencing, which, in a clinical setting, often involves exome sequencing of several hundred selected genes.

Recently, pathologists have started to leverage the pattern-recognition capabilities of machine-learning algorithms to extract deeper diagnostic insight into histological data. The digital analysis of histological specimens first became possible with the introduction of bright-field WSI instruments 20 years ago<sup>21,22</sup>, but it was not until 2016 that the United States Food and Drug Administration (FDA) released guidance on the technical requirements for the use of digital imaging in diagnosis<sup>23</sup>. Digital-pathology instruments, software and start-ups have proliferated over the past few years, fuelled in large part by the development of machine-learning algorithms that assist in the interpretation of H&E-stained slides<sup>24</sup>, which histopathology laboratories need to process at high volumes (often more than 1 million slides per year in a single hospital). Machine learning of images has shown its worth in several areas of medicine<sup>25,26</sup>, and promises to assist practitioners by increasing the efficiency and reproducibility of pathological diagnoses<sup>24</sup>. The pathology departments at several comprehensive cancer centres (institutions that are part of a network coordinated by the United States National Cancer Institute) have recently introduced multiplexed image-based immunoprofiling services for the identification of patients most likely to benefit from immune checkpoint inhibitors. Nonetheless, in most hospitals, the vast majority of diagnostic pathology still involves the visual inspection of physical specimens; only a minority of slides are scanned and digitized for concurrent or subsequent review on a computer screen. This is widely anticipated to change over the next decade; in fact, several European countries have national digital-pathology programs. As clinical pathology incorporates new measurement modalities and becomes increasingly digital, compatible and interoperable software and standards for clinical and research purposes need further development, and this is likely to be particularly important in the conduct of clinical trials. From a technical perspective, standards for digital image files (such as the OME-TIFF format) developed for preclinical research can work well with multiplexed images from clinical services. However, software for clinical use also requires security, workflows and billing features that are not necessary components of software used for research.

## Sharing imaging data from tissue atlases

Algorithms, software and standards for high-dimensional image data<sup>27,28</sup> remain underdeveloped in comparison to the tools available for almost all types of genomic information. Moreover, with sequencing data, the information present in primary data files (such as FASTQ files) are fully retained (or enhanced) when reads are aligned or when count tables are generated; it is rarely necessary to re-access the primary data. In contrast, methods for the extraction of features from tissue images are immature, and their visual inspection by experts for diagnostic or scientific purposes, or for the testing of new algorithms, requires repeated access to the original images at their native resolution.

As far back as 2008, the Journal of Cell Biology worked with the founders of the Open Microscopy Environment (OME)<sup>29</sup> to deploy a JCB DataViewer<sup>30</sup>, which provided direct access to primary high-resolution microscopy data (much of it from tissue culture cells and model organisms). Economic pressures ended this ambitious effort<sup>31</sup>, which emphasizes that funds have long been available to purchase expensive microscopes but not to distribute the resulting data. Currently, most H&E, IHC and multiplexed tissue images are shared only as figure panels in manuscripts, a form that typically provides access to a few

selected fields of view at a single (and often downsampled) resolution. In the best case, these data might be available via public data repositories such as figshare<sup>32</sup>. The European Bioinformatics Institute (EBI) Image Data Resource (IDR) is a notable exception; IDR uses the OME-compatible OMERO<sup>33</sup> server to provide full-resolution access to selected microscopy data<sup>34</sup>. The importance of sharing image data is particularly pronounced in the case of research biopsies, including those being used to assemble tumour atlases. These biopsies are obtained to expand scientific knowledge rather than to inform the treatment of individual patients, and there is an ethical obligation for the resulting data (appropriately anonymized) to be made widely available in an open and useful form for research purposes<sup>35</sup>. Open-source data and code are widely recognized as valuable tools in genomics<sup>36,37</sup> and in basic cell-and-developmental-biology research, but availability has only recently been flagged as a significant barrier to scientific progress<sup>38</sup>. More generally, digital pathology and medical imaging are disciplines in which the goal of making research findable, accessible, interoperable and reusable (FAIR)<sup>39</sup> is desirable, but the computational infrastructure is insufficient to meet these aspirations.

## Software for image analysis and interpretation

A wide variety of academic and commercial microscopy software systems are available either as a desktop system, for data analysis, or as a client–server relational database management system (RDBMS), for image management (Figure 2a and Boxes 1 and 2). Desktop software is particularly good for interactive image analysis because it exploits graphics cards for rapid image rendering as well as high-bandwidth connections to local data for computation. RDBMSs are instead ideal for data management because they enable relational queries, support multiple simultaneous users, ensure data integrity and effectively manage access to large-scale local and cloud-based computer resources (Figure 2b provides a detailed comparison of available software).

However, with the introduction of high-plex and whole-slide tissue images, new software is required to guide users through the extraordinary complexity of images that encompass multiple square centimetres of tissue,  $10^5$ – $10^7$  cells and upwards of 100 channels. Images of this size are not only difficult to process and share but they are also difficult to understand. We envision a key role for ‘interpretive guides’ or ‘digital docents’ that help walk users through a series of human-provided and machine-generated annotations about an image in much the same way that the results section of a published article guides users through a multi-panel figure. Genomic science faced an analogous need for efficient and intuitive visualization tools a decade ago, and this led to the development of the influential Integrative Genomics Viewer<sup>40</sup> and its many derivatives.

Interactive guides of pictures and images have proven highly successful in other scientific fields, as exemplified by Project Mirador (<https://projectmirador.org>). Project Mirador focuses on the development of web-based interpretive tours of cultural resources such as art museums, illuminated manuscripts and maps of culturally or historically significant cities. In these online tours, a series of waypoints (reference points used to help navigate the information displayed) and accompanying text direct users to areas of interest while also allowing free exploration (the user can then return to the narrative when needed).

This type of interactive narration—also known as digital storytelling or visual storytelling—mimics some of the benefits provided by tour guides and functions as a pedagogical tool that enhances comprehension<sup>41</sup> and memory formation<sup>42</sup>. Multiple studies have identified benefits associated with receiving complex information in a narrative manner, and digital storytelling has been applied to several areas of medicine and research, including oncology<sup>43</sup>, mental health<sup>44</sup>, health equity<sup>45</sup> and science communication<sup>46,47</sup>.

How might these lessons be applied to tissue images? At the very least, they suggest that simply making gigapixel-sized images available for download and analysis on desktop software is insufficient to optimally advance scientific discovery and to improve diagnostic outcomes. Instead, an easy approach that guides users through the salient features of an image with associated annotation and commentary would be beneficial to both specialists and non-specialists. Currently, pathologists share insights with colleagues using a multi-head microscope to pan across an image and to switch between high and low-power fields (magnifications), thereby studying cells in detail while also placing them in the context of the overall tissue. In this process, key features are often highlighted using a pointer. In the case of multiplex images, time-tested pathology practice must be built into software along with the means to toggle channels on and off (so that the contribution of specific antibodies to the final image can be ascertained) and display quantitative data arising from image analysis.

## Software-based interpretive guides

How might interpretive guides for tissue images be implemented? One possibility involves the use of an OMERO<sup>33</sup> client. OMERO is the most widely deployed open-source image-informatics system, and it is compatible with a range of software clients. However, an OMERO client requires access to an OMERO database and, as images get larger, the server becomes substantially loaded, which limits the number of concurrent sessions. Database-independent viewers are used by Project Mirador and are based on the open-source OpenSeadragon<sup>48</sup> platform. OpenSeadragon makes it easy to zoom and pan across images (in a manner similar to Google Maps<sup>49</sup>). Minerva, which is customized for guided viewing of multiplexed images, has taken this approach<sup>50</sup>. Minerva is a web application that uses a client-side JavaScript (and thus requires no downloaded software), and it is easily deployable on standard commercial clouds such as Amazon Web Services (AWS) (Figure 3) or on local computing servers that support the static site generator Jekyll<sup>51</sup>. OpenSeadragon has been previously used for displaying H&E images<sup>52</sup>, but in Minerva, it is paired with narrative features, interactive views of derived single-cell features within the image space, lightweight implementation and the ability to accommodate both bright-field (H&E and IHC) and multiplexed immunofluorescence images. Minerva can also display images and annotations on smartphones.

Minerva is OME-compatible and BioFormats-compatible and is therefore usable with images from virtually any existing microscope or slide scanner. There is no practical limit to the number of users who can concurrently access narrative guides developed in Minerva and there is no requirement for specialized servers or a relational database, thereby keeping complexity and costs low. Anyone familiar with GitHub and AWS (or similar



cloud services) can deploy a Minerva story in a few minutes, and new stories can then be generated by individuals with little expertise in software and computational biology. Developers can also build on the Minerva framework to develop narration tools for other applications (for example, for interpreting protein structures or complex high-dimensional datasets). Software viewers such as Minerva are not intended to be all-in-one solutions to the many computational challenges associated with processing and analysing tissue data; rather, they are specialized browsers that perform one task well. In the case of Minerva, it provides an intuitive and interpretive approach to images that have already been analysed using other tools. Genome browsers are similar: they do not perform alignment and data analysis; instead, they make it possible to interact with the processed sequences effectively.

Within a Minerva window on a standard web browser, a narration panel directs the user's attention to particular regions of an image and to specific channel groupings, and is accompanied by a text description (which Minerva can read aloud) and by image annotations involving overlaid geometric shapes or text (Figure 4). Each image can be associated with more than one narrated story and with different ways of viewing the same type of data. A fundamental aspect of the narrative guides is that individuals with expertise in a particular disease or tissue (in particular, a pathologist) can create stories for use by others to assist them in the understanding of the morphology of a specimen. Creating narrations requires an authoring tool; for Minerva, this is the Minerva Author. Minerva Author is a desktop application (in JavaScript React with a Python Flask backend) that converts images in OME-TIFF format<sup>53</sup> to pyramid images (layered images with different resolution, with the largest one at the bottom and the smallest one on the top) and that assists with the addition of waypoints and text annotations. Minerva Author supports RGB images, such as bright-field, H&E and IHC, as well as multichannel images such as immunofluorescence, CODEX and CyCIF. After specifying the rendering settings and writing the waypoints in Minerva Author, a user receives a configuration file and image pyramids to deploy to AWS S3 or to another web-based storage location (Figure 3). Stories can be as simple as a single panel with a short introduction or a multi-panel narration enriched with a series of views with detailed descriptions, changes in zoom level and associated data analysis. It takes users a few hours to learn the software and then from about 30 min to a few hours to create a story; this is about the same time required to create a good static figure panel for a published article, and much less time than that needed for data collection, image registration, segmentation and data analysis.

## Interactive guides for exploring human tumours

As a case study of an interactive tissue guide, we used Minerva Story to mark up a large specimen of human lung adenocarcinoma (<https://www.cycif.org/MinervaLungHistology>). This primary tumour measured  $\sim 5 \times 3.5$  mm and was imaged at subcellular resolution using 44-plex t-CyCIF<sup>4</sup>. Multiple fields were then stitched into a single image (described in ref. 54; the image is referred to within that paper as LUNG-3-PR). Because there is no single best way to analyse an image containing several hundred thousands of cells, we created two guides: a guide focused on the histological review of regions of interest and on specific types of immune cells and tumour cells, and a guide focused on the presentation of quantitative data analysis in the context of the original image (Figure 4a). A panel to the left of the

Minerva Story window (outlined in orange in Figure 4a) shows the name of the samples, the links to its related stories, a table of contents and the navigation tools. On the right of the window (outlined in pink), another navigation panel allows for the selection of channels of interest. Channels and channel groups, as well as the cell types that they define, can be pre-specified. Each protein (antigen) is linked to an explanatory source of information (such as the GeneCards<sup>55</sup> database; a more customized and tissue-specific annotation of markers is under development).

Stories progress from one waypoint to the next (analogous to the numbered system used by museum audio guides), and each waypoint can involve a different field of view, magnification and set of channels, as well as arrows and text describing specific features of interest (marked in grey). At any point, users can diverge from a story by panning and zooming around the image or by selecting different channels, and then return to the narrative by selecting the appropriate waypoint on the table of contents. In Figure 4a, ‘waypoint one’ shows pan-cytokeratin-positive tumour cells growing in chords and clusters at the tumour–stroma interface. This region of the tumour is characterized by an inflammatory microenvironment, as evidenced by the presence of a variety of lymphocyte and macrophage populations distinguishable by the expression of their cell surface markers. By using the panel on the right, users can toggle these markers on and off to explore the images and data, and to evaluate the accuracy of the classification. The remainder of the story explores the expression of PD-L1 and the localization of populations of lymphocytes and macrophages.

Narrative guides are also useful for showing the results of quantitative data analyses in the context of the original image (as described in the lung adenocarcinoma story). For example, the data analysis of high-plex tissue images typically involves measuring the co-expression of multiple cell type markers (such as immune lineage markers) for the identification of individual cell types. The analysis of cell states, morphologies and neighbourhood relationships are also common. Within Minerva Story, it is possible to link representations of quantitative data directly to the image space. For example, when data are captured in a 2D plot, such as a UMAP for dimension reduction (Figure 4b), selecting a data point takes the user directly to the corresponding position of the cell in the image (denoted with an arrow). This is a standard feature in desktop software such as histoCat<sup>56</sup> and Facetto<sup>57</sup>, and greatly enhances a user’s understanding of the relationship between images and image-derived features. The display of segmentation masks is similarly useful for troubleshooting and for assessing data quality (for example, in Figure 4b, middle, unwanted fusion and fission events are highlighted by arrows, both of which result from errors in segmentation). Additionally, users can interactively highlight areas of interest, add notes and generate sharable links that allow others to navigate to the same position in the image and to view any added annotations and text (for example, a high-magnification view with a labelled CD8+ Foxp3+ regulatory T cell in Figure 4b, right).

## Narrative guides for medical education

Narrative tissue guides can be used for teaching. Histology is challenging to teach in an undergraduate setting<sup>58</sup> and changes in curriculum have reduced the amount of time that medical students and residents spend in front of a microscope. Online collections of tissue



images are a frequent substitute. In fact, pairing classroom instruction with dynamic viewing and flexible interaction with image data is essential for learning<sup>59</sup>. We therefore created a narrative guide of H&E images obtained from specimens of heart tissue from a patient who experienced multiple episodes of myocardial infarction (Figure 5). An introductory panel depicts the overall structure of the heart and the positions from which various specimens were resected. These images display the histological hallmarks of ischaemic heart disease: severe coronary artery atherosclerosis, plaque rupture, stunned myocardium, reperfusion injury, as well as the early, intermediate and late features of myocardial tissue infarction. Rather than looking at snippets of histological data in a textbook or in poorly annotated online images, the interactive narrations of tissue-image data provide context that facilitates the more nuanced understanding of common cardiac pathophysiology.

Software such as Minerva Story should be useful for medical education and could enhance interactive textbooks. With relatively little effort, text can be hyperlinked to stories and waypoints, and students could use image-annotation features to take notes and to ask questions. Audio-based narration is advantageous in this context, as it could allow students to receive lecture content while simultaneously concentrating on a relevant image.

## Outlook

Multiplexed tissue-imaging methods generate images that are data-rich; however, their potential to inform basic, translational and clinical research remains largely untapped, which is in large part because the complexity and size of these images makes it difficult to process and share them. As tissue atlases become increasingly available, software that enables interactive data exploration with cloud-based storage of pre-rendered narrative guides could dramatically increase the number of people who can access a pathologist's expertise and view an image. Ideal systems combine intuitive panning, zooming and data-exploration interfaces, guidance from digital docents and low costs for the data providers. These possibilities are illustrated by the Minerva software and the OpenSeadragon platform on which it is based. Ideally, publication of high-plex images could in the future be tied to the release of the primary data and to software viewers having digital docents to make the data easier to interpret; this would meet the key goals of the JCB DataViewer project<sup>30</sup> and be a step towards realizing FAIR data standards.

With respect to data analysis, a long-term goal of computational biology is to have a single copy of the data in the cloud so that computational algorithms can be more easily used across cloud datasets. At present, however, the dominant approach of local data analysis remains, in particular for interactive tasks such as optimizing segmentation, training machine-learning algorithms, and performing and validating data clustering, which all require local access to full-resolution primary data. With commercial cloud services, data download has become the primary expense and can be substantial for multi-terabyte datasets. A solution could be the use of 'requester pays' buckets, available on cloud services from Amazon, Google and other providers of these services. These would allow a data generator to make even large datasets publicly available by requiring the requester to cover the cost of data transfer. In the short term, we envision an ecosystem in which web clients for databases such as OMERO provide interactive viewing of primary data in private

datasets for which demand on the OMERO server can be better managed. Lightweight viewers such as Minerva would provide unrestricted access to published and processed data, making the data easy to access and understand by non-experts, and requester-pays cloud buckets would allow experts (in particular, computational biologists) to access primary data and to perform their own analyses locally. In the longer term, these functions are likely to merge, with OMERO providing sophisticated cloud-based processing of data and Minerva serving as a means to access data virtually cost-free, either in OMERO or in a static file system.

An obvious extension of Minerva would be the addition of tools that facilitate supervised machine learning; in such an application, a Minerva variant would provide an efficient way of adding the needed labels to data to train classifiers and neural networks. The results of the machine-learning-based classification could also be checked in Minerva, as its lightweight implementation facilitates crowdsourcing. With these possibilities in mind, a key question is whether tools such as Minerva should also expand to include sophisticated image-processing functions. We do not think so; instead, we believe that narrative guides should be optimized for image review, publication and description, and other interoperable software should continue to be used for data analysis. These tools can be joined together into efficient workflows (such as MCMICRO<sup>60</sup>) that use software containers (such as Docker)<sup>61</sup> and pipeline frameworks (such as Nextflow<sup>62</sup>). This approach is not perfect<sup>63</sup> (Box 2), but it cannot reasonably be replaced by all-in-one commercial or academic software. Recent interest in single-cell tissue biology derives not only from advances in microscopy but also from the widespread adoption of single-cell sequencing<sup>64,65</sup>. Comprehensive characterization of normal and diseased tissues will almost certainly involve the integration of data from multiple analytical modalities, including imaging, spatial transcriptomic profiling<sup>66,67</sup>, mass spectrometry imaging of metabolites and drugs<sup>68,69</sup>, and computational registration of dissociated single-cell RNA sequencing<sup>70</sup> with spatial features. Minerva cannot currently handle all of these tasks, but it can be readily combined with other tools to create the multi-omic viewers that are needed for tissue atlases. An attractive possibility is to add narrative tissue guides to widely used genomics platforms such as cBioPortal for Cancer Genomics<sup>71</sup> to create environments in which genomic and multiplexed tissue histopathology can be viewed simultaneously. Better visualization could also help with the more conceptually challenging task of integrating spatiomolecular features in multiplex images with gene expression and mutational data. With regards to clinical applications, the biomedical community needs to ensure that digital-pathology systems do not become locked behind proprietary data formats based on non-interoperable software. OME and BioFormats for microscopy, and DICOM for radiology<sup>72</sup>, exemplify that it is possible for academic developers, commercial instrument manufactures and software vendors to work together for mutual benefit. Easy-to-use visualization software based on permissive open-source licensing could readily be deployed by research groups, instrument makers and scientific publishers to make high-plex image data widely accessible and easy to understand. This would help to realize a FAIR future for tissue imaging and digital pathology.

## Acknowledgements

This work was funded by NIH grants U54-CA225088 to P.K.S. and S.S., and by the Ludwig Center at Harvard. The Dana-Farber/Harvard Cancer Center is supported in part by NCI Cancer Center Support Grant P30-CA06516.

Competing interests

P.K.S. is a member of the SAB and BOD member of Applied Biomath, RareCyte Inc., and Glencoe Software, which distributes a commercial version of the OMERO database. P.K.S. is also a member of the NanoString SAB. In the past 5 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. The remaining authors declare no competing interests.

## Data availability

Source code for Minerva Story can be found at <https://github.com/labsyspharm/minerva-story> and detailed documentation and user guide at <https://github.com/labsyspharm/minerva-story/wiki>.

## References

- Giesen C et al. Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nat. Methods* 11, 417–422 (2014). [PubMed: 24584193]
- Angelo M et al. Multiplexed ion beam imaging of human breast tumors. *Nat. Med.* 20, 436–442 (2014). [PubMed: 24584119]
- Gerdes MJ et al. Highly multiplexed single-cell analysis of formalin-fixed, paraffin-embedded cancer tissue. *Proc. Natl Acad. Sci. USA* 110, 11982–11987 (2013). [PubMed: 23818604]
- Lin J-R et al. Highly multiplexed immunofluorescence imaging of human tissues and tumors using t-CyCIF and conventional optical microscopes. *eLife* 10.7554/eLife.31657 (2018).
- Bodenmiller B Multiplexed epitope-based tissue imaging for discovery and healthcare applications. *Cell Syst.* 2, 225–238 (2016). [PubMed: 27135535]
- Coy S et al. Multiplexed immunofluorescence reveals potential PD-1/ PD-L1 pathway vulnerabilities in craniopharyngioma. *Neuro Oncol.* 20, 1101–1112 (2018). [PubMed: 29509940]
- Goltsev Y et al. Deep profiling of mouse splenic architecture with CODEX multiplexed imaging. *Cell* 174, 968–981.e15 (2018). [PubMed: 30078711]
- Stack EC, Wang C, Roman KA & Hoyt CC Multiplexed immunohistochemistry, imaging, and quantitation: a review, with an assessment of tyramide signal amplification, multispectral imaging and multiplex analysis. *Methods* 70, 46–58 (2014). [PubMed: 25242720]
- Akturk G, Sweeney R, Remark R, Merad M & Gnjatic S Multiplexed immunohistochemical consecutive staining on single slide (MICSSS): multiplexed chromogenic IHC assay for high-dimensional tissue analysis. *Methods Mol. Biol.* 2055, 497–519 (2020).
- Saka SK et al. Immuno-SABER enables highly multiplexed and amplified protein imaging in tissues. *Nat. Biotechnol.* 37, 1080–1090 (2019). [PubMed: 31427819]
- Li W, Germain RN & Gerner MY Multiplex, quantitative cellular analysis in large tissue volumes with clearing-enhanced 3D microscopy (Ce3D). *Proc. Natl Acad. Sci. USA* 114, E7321–E7330 (2017). [PubMed: 28808033]
- Regev A et al. The Human Cell Atlas. *eLife* 10.7554/eLife. 27041 (2017).
- Consortium HuBMAP. The human body at cellular resolution: the NIH Human Biomolecular Atlas Program. *Nature* 574, 187–192 (2019). [PubMed: 31597973]
- Rozenblatt-Rosen O et al. The Human Tumor Atlas Network: charting tumor transitions across space and time at single-cell resolution. *Cell* 181, 236–249 (2020). [PubMed: 32302568]
- The Cancer Genome Atlas Research Network et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat. Genet.* 45, 1113–1120 (2013). [PubMed: 24071849]

16. Shin D et al. PathEdEx—uncovering high-explanatory visual diagnostics heuristics using digital pathology and multiscale gaze data. *J. Pathol. Inf.* 8, 29 (2017).
17. Wagner J et al. A single-cell atlas of the tumor and immune ecosystem of human breast cancer. *Cell* 177, 1330–1345.e18 (2019). [PubMed: 30982598]
18. Bui MM et al. Digital and computational pathology: bring the future into focus. *J. Pathol. Inform.* 10, 10 (2019).
19. Brown M & Wittwer C Flow cytometry: principles and clinical applications in hematology. *Clin. Chem.* 46, 1221–1229 (2000). [PubMed: 10926916]
20. Coons AH, Creech HJ, Jones RN & Berliner E The demonstration of pneumococcal antigen in tissues by the use of fluorescent antibody. *J. Immunol.* 45, 159–170 (1942).
21. Pantanowitz L et al. Twenty years of digital pathology: an overview of the road travelled, what is on the horizon, and the emergence of vendor-neutral archives. *J. Pathol. Inf.* 9, 40 (2018).
22. Gurcan MN et al. Histopathological image analysis: a review. *IEEE Rev. Biomed. Eng.* 2, 147–171 (2009). [PubMed: 20671804]
23. Technical Performance Assessment of Digital Pathology Whole Slide Imaging Devices (US Food and Drug Administration, 2016); <http://www.fda.gov/regulatory-information/search-fda-guidance-documents/technical-performance-assessment-digital-pathology-whole-slide-imaging-devices>
24. Ehteshami Bejnordi B et al. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA* 318, 2199–2210 (2017). [PubMed: 29234806]
25. Gulshan V et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 316, 2402–2410 (2016). [PubMed: 27898976]
26. Greenspan H, van Ginneken B & Summers RM Guest editorial deep learning in medical imaging: overview and future promise of an exciting new technique. *IEEE Trans. Med. Imaging* 35, 1153–1159 (2016).
27. Abels E et al. Computational pathology definitions, best practices, and recommendations for regulatory guidance: a white paper from the Digital Pathology Association. *J. Pathol.* 249, 286–294 (2019). [PubMed: 31355445]
28. Schapiro D et al. MITI Minimum Information guidelines for highly multiplexed tissue images. Preprint at <https://arxiv.org/abs/2108.09499> (2021).
29. Goldberg IG et al. The Open Microscopy Environment (OME) data model and XML file: open tools for informatics and quantitative analysis in biological imaging. *Genome Biol.* 6, R47 (2005). [PubMed: 15892875]
30. Hill E Announcing the JCB DataViewer, a browser-based application for viewing original image files. *J. Cell Biol.* 183, 969–970 (2008).
31. Swedlow JR, Goldberg I, Brauner E & Sorger PK Informatics and quantitative analysis in biological imaging. *Science* 300, 100–102 (2003). [PubMed: 12677061]
32. Singh J FigShare. *J. Pharm. Pharmacother.* 2, 138–139 (2011).
33. Allan C et al. OMERO: flexible, model-driven data management for experimental biology. *Nat. Methods* 9, 245–253 (2012). [PubMed: 22373911]
34. Williams E et al. The Image Data Resource: a bioimage data integration and publication platform. *Nat. Methods* 14, 775–781 (2017). [PubMed: 28775673]
35. Levit LA et al. Ethical framework for including research biopsies in oncology clinical trials: American Society of Clinical Oncology research statement. *J. Clin. Oncol.* 37, 2368–2377 (2019). [PubMed: 31343905]
36. Kaye J, Heeney C, Hawkins N, de Vries J & Boddington P Data sharing in genomics—reshaping scientific practice. *Nat. Rev. Genet.* 10, 331–335 (2009). [PubMed: 19308065]
37. Reardon J et al. Bermuda 2.0: reflections from Santa Cruz. *Gigascience* 5, 1–4 (2016).
38. Gutman DA et al. The Digital Slide Archive: a software platform for management, integration, and analysis of histology for cancer research. *Cancer Res.* 77, e75–e78 (2017). [PubMed: 29092945]
39. Wilkinson MD et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3, 160018 (2016). [PubMed: 26978244]

40. Robinson JT et al. Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–26 (2011). [PubMed: 21221095]
41. Stephens GJ, Silbert LJ & Hasson U Speaker–listener neural coupling underlies successful communication. *Proc. Natl Acad. Sci. USA* 107, 14425–14430 (2010).
42. Rieger KL et al. Digital storytelling as a method in health research: a systematic review protocol. *Syst. Rev.* 7, 41 (2018). [PubMed: 29506568]
43. Wilson DK, Hutson SP & Wyatt TH Exploring the role of digital storytelling in pediatric oncology patients' perspectives regarding diagnosis: a literature review. *SAGE Open* 10.1177/2158244015572099 (2015).
44. De Vecchi N, Kenny A, Dickson-Swift V & Kidd S How digital storytelling is used in mental health: a scoping review. *Int J. Ment. Health Nurs.* 25, 183–193 (2016). [PubMed: 26900000]
45. Lee H, Fawcett J & DeMarco R Storytelling/narrative theory to address health communication with minority populations. *Appl. Nurs. Res* 30, 58–60 (2016). [PubMed: 27091254]
46. Botsis T, Fairman JE, Moran MB & Anagnostou V Visual storytelling enhances knowledge dissemination in biomedical science. *J. Biomed. Inf.* 107, 103458 (2020).
47. ElShafie SJ Making science meaningful for broad audiences through stories. *Integr. Comp. Biol.* 58, 1213–1223 (2018). [PubMed: 30059998]
48. OpenSeadragon v.2.4.2 (OpenSeadragon contributors, 2013); <https://openseadragon.github.io/>
49. Jianu R & Laidlaw DH What Google Maps can do for biomedical data dissemination: examples and a design study. *BMC Res. Notes* 6, 179 (2013). [PubMed: 23642009]
50. Hoffer J et al. Minerva: a light-weight, narrative image browser for multiplexed tissue images. *J. Open Source Softw.* 5, 2579 (2020). [PubMed: 33768192]
51. Jekyll v.4.2.0 (The Jekyll Team; 2021); <https://jekyllrb.com/>
52. Aeffner F et al. Introduction to digital image analysis in whole-slide imaging: a white paper from the Digital Pathology Association. *J. Pathol. Inf.* 10, 9 (2019).
53. Hiner MC, Rueden CT & Eliceiri KW SCIFIO: an extensible framework to support scientific image formats. *BMC Bioinformatics* 17, 521 (2016). [PubMed: 27927161]
54. Rashid R et al. Highly multiplexed immunofluorescence images and single-cell data of immune markers in tonsil and lung cancer. *Sci. Data* 6, 323 (2019). [PubMed: 31848351]
55. Rebhan M, Chalifa-Caspi V, Prilusky J & Lancet D GeneCards: integrating information about genes, proteins and diseases. *Trends Genet.* 13, 163 (1997). [PubMed: 9097728]
56. Schapiro D et al. histoCAT: analysis of cell phenotypes and interactions in multiplex image cytometry data. *Nat. Methods* 14, 873–876 (2017). [PubMed: 28783155]
57. Krueger R et al. Facetto: combining unsupervised and supervised learning for hierarchical phenotype analysis in multi-channel image data. *IEEE Trans. Vis. Comput. Graph.* 26, 227–237 (2020). [PubMed: 31514138]
58. García M, Victory N, Navarro-Sempere A & Segovia Y Students' views on difficulties in learning histology. *Anat. Sci. Educ.* 12, 541–549 (2019). [PubMed: 30378295]
59. Mione S, Valcke M & Cornelissen M Remote histology learning from static versus dynamic microscopic images. *Anat. Sci. Educ.* 9, 222–230 (2016). [PubMed: 26441000]
60. Schapiro D et al. MCMICRO: a scalable, modular image-processing pipeline for multiplexed tissue imaging. Preprint at bioRxiv 10.1101/2021.03.15.435473 (2021).
61. O'Connor BD et al. The Dockstore: enabling modular, community-focused sharing of Docker-based genomics tools and workflows. *F1000Res* 6, 52 (2017). [PubMed: 28344774]
62. Di Tommaso P et al. Nextflow enables reproducible computational workflows. *Nat. Biotechnol.* 35, 316–319 (2017). [PubMed: 28398311]
63. Siepel A Challenges in funding and developing genomic software: roots and remedies. *Genome Biol.* 20, 147 (2019). [PubMed: 31358028]
64. Macaulay IC, Ponting CP & Voet T Single-cell multiomics: multiple measurements from single cells. *Trends Genet.* 33, 155–168 (2017). [PubMed: 28089370]
65. Trapnell C Defining cell types and states with single-cell genomics. *Genome Res.* 25, 1491–1498 (2015). [PubMed: 26430159]

66. Ståhl PL et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* 353, 78–82 (2016). [PubMed: 27365449]
67. Schulz D et al. Simultaneous multiplexed imaging of mRNA and proteins with subcellular resolution in breast cancer tissue samples by mass cytometry. *Cell Syst.* 6, 25–36.e5 (2018). [PubMed: 29289569]
68. Van de Plas R, Yang J, Spraggins J & Caprioli RM Image fusion of mass spectrometry and microscopy: a multimodality paradigm for molecular tissue mapping. *Nat. Methods* 12, 366–372 (2015). [PubMed: 25707028]
69. Liu X et al. Molecular imaging of drug transit through the blood–brain barrier with MALDI mass spectrometry imaging. *Sci. Rep.* 3, 2859 (2013). [PubMed: 24091529]
70. Moncada R et al. Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. *Nat. Biotechnol.* 38, 333–342 (2020). [PubMed: 31932730]
71. Cerami E et al. The cBio Cancer Genomics Portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* 2, 401–404 (2012). [PubMed: 22588877]
72. Mildenerberger P, Eichelberg M & Martin E Introduction to the DICOM standard. *Eur. Radio.* 12, 920–927 (2002).
73. Samuel AL Some studies in machine learning using the game of checkers. *IBM J. Res. Dev.* 3, 210–229 (1959).
74. Inoué S & Spring K Video Microscopy: The Fundamentals (Springer US, 1997).
75. Bacus JV & Bacus JW Method and apparatus for acquiring and reconstructing magnified specimen images from a computer-controlled microscope. US patent 6,101,265 (2000).
76. HANDEL-P (Pancreatlas, 2021); <https://pancreatlas.org/>
77. Rubin DL, Greenspan H & Brinkley JF in *Biomedical Informatics: Computer Applications in Health Care and Biomedicine* (eds Shortliffe EH & Cimino JJ) 285–327 (Springer, 2014).
78. caMicroscope (GitHub, 2021); <https://github.com/camicroscope>
79. Clark K et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository. *J. Digit. Imaging* 26, 1045–1057 (2013). [PubMed: 23884657]
80. PathPresenter (Singh R et al., 2021); <https://public.pathpresenter.net/#/login>
81. Olson AH Image Analysis Using the Aperio ScanScope (Quorum Technologies, 2006).
82. McQuin C et al. CellProfiler 3.0: next-generation image processing for biology. *PLoS Biol.* 16, e2005970 (2018). [PubMed: 29969450]
83. Bankhead P et al. QuPath: open source software for digital pathology image analysis. *Sci. Rep.* 7, 16878 (2017). [PubMed: 29203879]
84. Stritt M, Stalder AK & Vezzali E Orbit Image Analysis: an open-source whole slide image analysis tool. *PLoS Comput. Biol.* 16, e1007313 (2020). [PubMed: 32023239]
85. Mantis Viewer (Parker Institute for Cancer Immunotherapy, 2021); <https://mantis.parkerici.org/>
86. ASAP—Automated Slide Analysis Platform (Computation Pathology Group—Radboud University Medical Center, 2021); <https://computationalpathologygroup.github.io/ASAP/#home>
87. Maaten Lvander & Hinton G Visualizing data using t-SNE. *J. Mach. Learn. Res.* 9, 2579–2605 (2008).
88. McInnes L, Healy J & Melville J UMAP: uniform manifold approximation and projection for dimension reduction. Preprint at <https://arxiv.org/abs/1802.03426> (2018).
89. Arnol D, Schapiro D, Bodenmiller B, Saez-Rodriguez J & Stegle O Modeling cell–cell interactions from spatial molecular data with spatial variance component analysis. *Cell Rep.* 29, 202–211.e6 (2019). [PubMed: 31577949]
90. Berg S et al. ilastik: interactive machine learning for (bio)image analysis. *Nat. Methods* 16, 1226–1232 (2019). [PubMed: 31570887]
91. Chen J et al. The Allen Cell Structure Segmenter: a new open source toolkit for segmenting 3D intracellular structures in fluorescence microscopy images. Preprint at bioRxiv 10.1101/491035 (2018).
92. Sofroniew N et al. napari/napari: 0.3.6rc2. Zenodo 10.5281/zenodo.3951241 (2020).



93. Schindelin J, Rueden CT, Hiner MC & Eliceiri KW The ImageJ ecosystem: an open platform for biomedical image analysis. *Mol. Reprod. Dev.* 82, 518–529 (2015). [PubMed: 26153368]

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**Box 1:****Software for managing and visualizing image data**

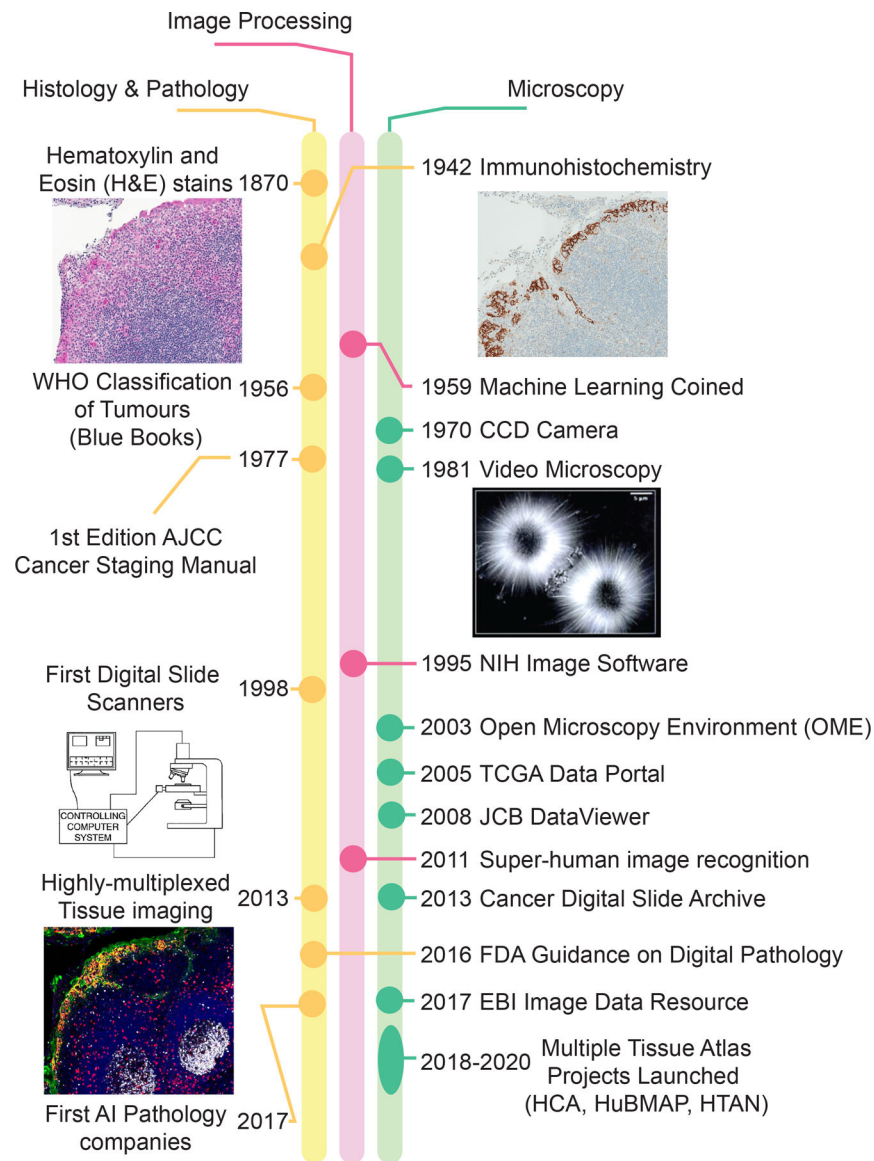
The OME-based OMERO<sup>33</sup> server is the most widely used image-informatics system for microscopy data in a research setting. It is the foundation of the IDR-based resource at the EBI<sup>34</sup>, a prominent large-scale publicly accessible image repository, and is part of more specialized repositories such as Pancreatlas<sup>76</sup>. OMERO has a client-server three-tier architecture involving a relational database, an image server and one or more interoperable user interfaces. OMERO is well suited to managing image data and metadata and to organizing images so that they can be queried using a visual index or via search<sup>77</sup> (Figure 2a). In its current form, it does not perform sophisticated image analysis and it is not specialized for the creation of narrative guides.

A range of other software tools are available for the static visualization or partially interactive visualization of H&E and IHC images in a web browser. In particular, caMicroscope<sup>78</sup> is used to organize IHC and H&E images for TCGA<sup>15</sup>. Other publicly available image interfaces include the Cancer Imaging Archive<sup>79</sup>, the Digital Slide Archive<sup>38</sup>, PathPresenter<sup>80</sup> and Aperio<sup>81</sup> (Figure 2b). A white paper from the Digital Pathology Association provides a description of the tools being developed to view bright-field digital-pathology data<sup>52</sup>. However, such H&E and IHC viewers are not generally compatible with multichannel images or with the integration of different types of omics data.

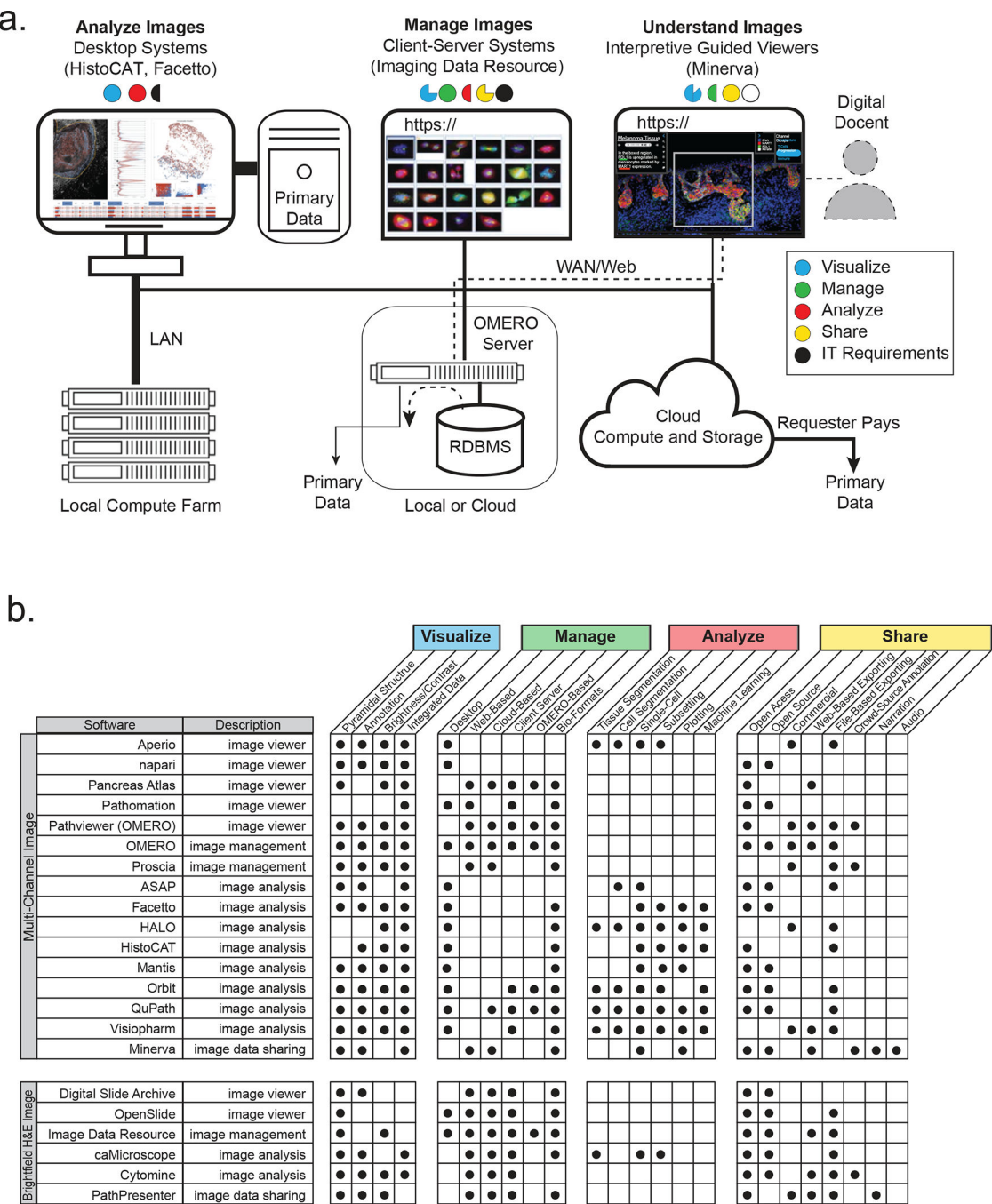
**Box 2:****Software for the analysis of image data**

Multiple software suites—CellProfiler<sup>82</sup>, histoCAT<sup>56</sup>, Facetto<sup>57</sup>, QuPath<sup>83</sup>, Orbit<sup>84</sup>, Mantis<sup>85</sup> and ASAP<sup>86</sup>, among others—have been developed for the analysis of high-dimensional image data from tissues or cultured cells. Many of these software systems are desktop-based and run locally (Figure 2b). They generally perform image segmentation to identify individual cells or tissue-level features, determine cell centroids and shape parameters (such as area and eccentricity), and compute staining intensities in designated regions of an image and across all channels. The resulting vectors can then be processed using standard tools for high-dimensional data analysis such as supervised and unsupervised clustering, t-SNE<sup>87</sup> or uniform manifold approximation and projection (UMAP)<sup>88</sup> to identify cell types and to study cell–cell interactions<sup>89</sup>. Although some tools support segmentation, others require pre-generated segmentation masks and single-cell feature tables. As an alternative, other approaches can analyse tissues at the level of individual pixels via the use of machine-learning algorithms and convolutional neural networks<sup>90</sup>. This approach potentially bypasses the need to segment individual cells from densely packed tissues, in which cells can dramatically vary in size. A key feature of software such as histoCat<sup>56</sup> or Facetto<sup>57</sup> is the integration of an image viewer with feature-based representations of the same data. This is essential for training and testing classifiers, for the quality control of image-processing routines and for obtaining insight into spatial characteristics. Pipelines that integrate and simplify the processing of high-plex images have started to appear<sup>60</sup> and are based on technologies used in existing pipelines for the processing of sequencing data.

Several cloud-based computing platforms for digital pathology are being commercially developed, such as HALO (Indica Labs), Visiopharm (Visiopharm) and PathViewer (Glencoe Software, the commercial developer of OMERO). Academic efforts such as the Allen Cell Explorer<sup>91</sup> and napari<sup>92</sup> build on highly successful open-source software platforms such as ImageJ<sup>93</sup>. Napari is particularly attractive to computational biologists because it is written entirely using the programming language Python and has user-interface elements and a console. Commercial tools often strive for an all-in-one approach to analysis and visualization, but this comes at the cost of complexity, proprietary implementations, and licensing fees. It also ignores one of the primary lessons from genomics: progress in rapidly developing research fields rarely involves the use of a single integrated software suite; rather, it relies on an ecosystem of interoperable tools specialized for specific tasks.



**Figure 1. Milestones in the development of histopathology, image processing and microscopy.** Methods used for histology and anatomical pathology have evolved towards digitalization in parallel with the growth of image-processing capabilities, multiplexed imaging modalities and the launch of multiple tissue-atlas projects. The first use of the term “machine learning”, in 1959, is widely credited to A. Samuel (ref. 73). Publicly available atlases of human tissues, tumours and cell types are listed in Table 1. Video microscopy image reproduced with permission from ref. 74. Slide scanner diagram reproduced with permission from ref. 75. AI, artificial intelligence; AJCC, American Joint Committee on Cancer; CCD, charge-coupled device; WHO, World Health Organization.

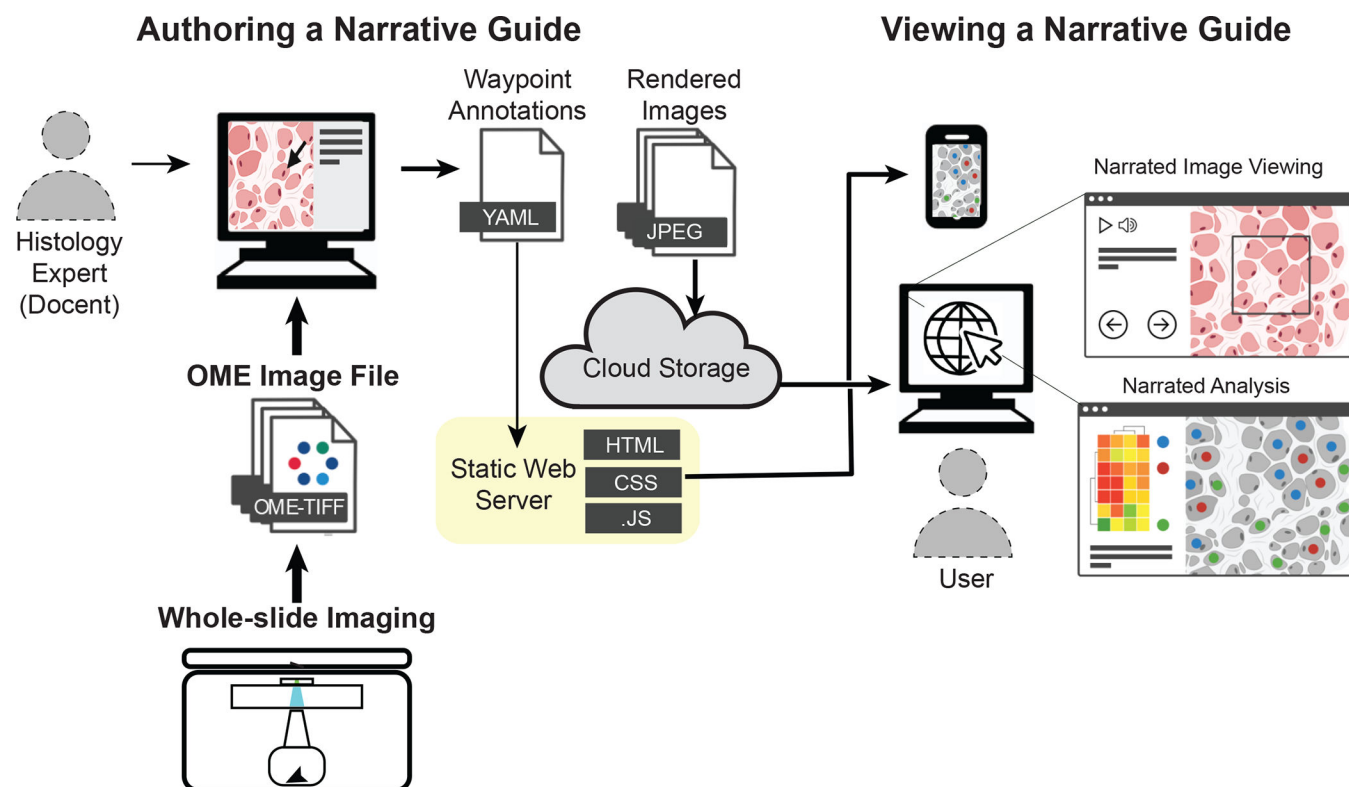


**Figure 2. Software used to visualize, analyse, manage and share tissue images.** Software systems for processing tissue images are typically optimized to perform a select number of tasks and to co-evolve to complement each other’s strengths and weaknesses as part of an ecosystem held together by common data standards and interoperable application programming interfaces. a, Desktop applications such as histoCat provide sophisticated tools for the interactive quantification and analysis of primary image data, and effectively exploit embedded image-rendering capabilities. Many desktop applications can be run in a ‘headless’ configuration on computer farms to accelerate the analysis of large images.

They can also fetch images from three-tier client–server systems such as OMERO. OMERO involves a RDBMS, an image server and one or more browser-based clients and desktop clients; such systems provide a full-featured approach to organizing, visualizing and sharing large numbers of images, but they typically support limited data processing. The Minerva application discussed in this Perspective is optimized for data sharing and interpretive image viewing and enables a detailed narration of image features and derived data. Pre-rendering of stories with multiple waypoints in an image mimics the insight provided by a human expert or guide (a ‘docent’). Public access to large-scale primary data is typically limited with all of these systems because of the costs of data transfer. However, in some OMERO configurations (in particular, the IDR), primary data can be downloaded from the same file system. Another approach to primary-data access being used by HTAN involves ‘requester pays’ cloud-based storage (for instance, an AWS S3 bucket). Minerva Author may also be extended to enable direct access to primary data stored in OMERO or these S3 buckets.

b, The key features of commercial and academic software tools suitable for viewing image data; each tool has its strengths and weaknesses, but none comprehensively satisfies all of the functions needed for complete image analysis. Data generators and consumers rely on a suite of interoperable desktop-and-server or cloud-based software systems. Panel a reproduced with permission from ref. 57, IEEE (left) and ref. 33 (right).

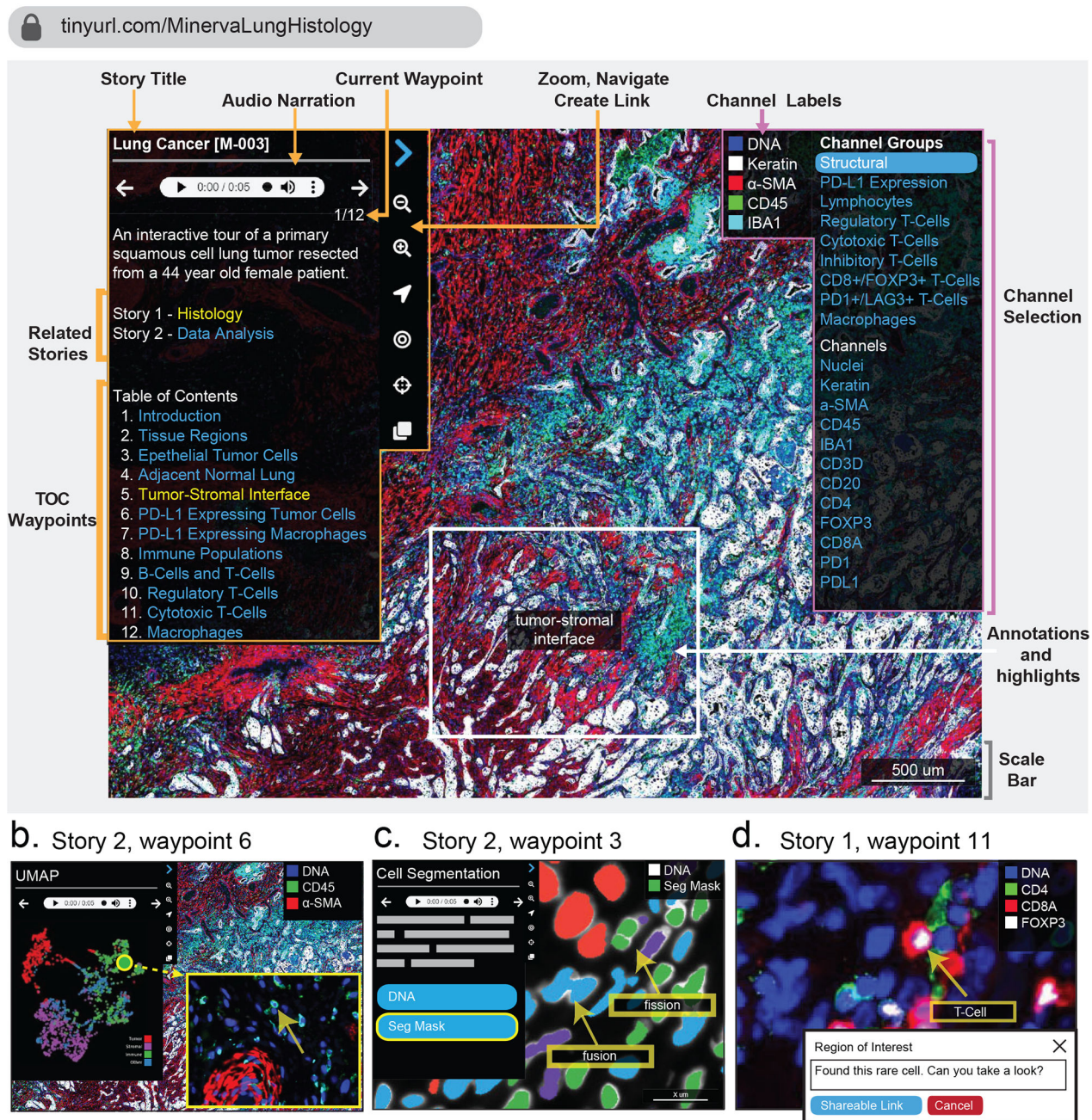




**Figure 3. A system for generating and viewing online narrative guides for histopathology tissue images.**

The illustrated architecture is based on the OpenSeadragon viewer Minerva Story and the narration tool Minerva Author. The thickness of the arrows indicates the amount of data transferred. Whole-slide images, including bright-field and multi-channel microscopy images, in the standard OME-TIFF format are imported into Minerva Author by a user with expertise in tissue biology and image interpretation (the 'docent'); in many cases this individual is a pathologist, histologist or cell biologist. The docent uses tools in Minerva Author to pan across the image and to then set channel-rendering values (background levels and intensity scale), specify waypoints and add text and graphical annotations. Minerva Author then renders image pyramids for all channel groupings as JPEG files and generates a YAML configuration file that specifies waypoints and associated information. The rendered images are stored on a cloud host (such as AWS S3) and accessed via a static web server supporting Jekyll (on GitHub, for example). A user simply opens a web browser, selects a link, and Minerva Story launches the necessary client-side JavaScript (.JS), making it possible to follow the story and to freely explore the image by panning, zooming and selecting channels. Because interactivity is handled on the client side, no special backend software or server is needed. Images rendered by Minerva Story are compatible with multiple devices, including cell phones. Source code for the Minerva Story can be found at <https://github.com/labsyspharm/minerva-story>, and detailed documentation and user guide at <https://github.com/labsyspharm/minerva-story/wiki>.

a.

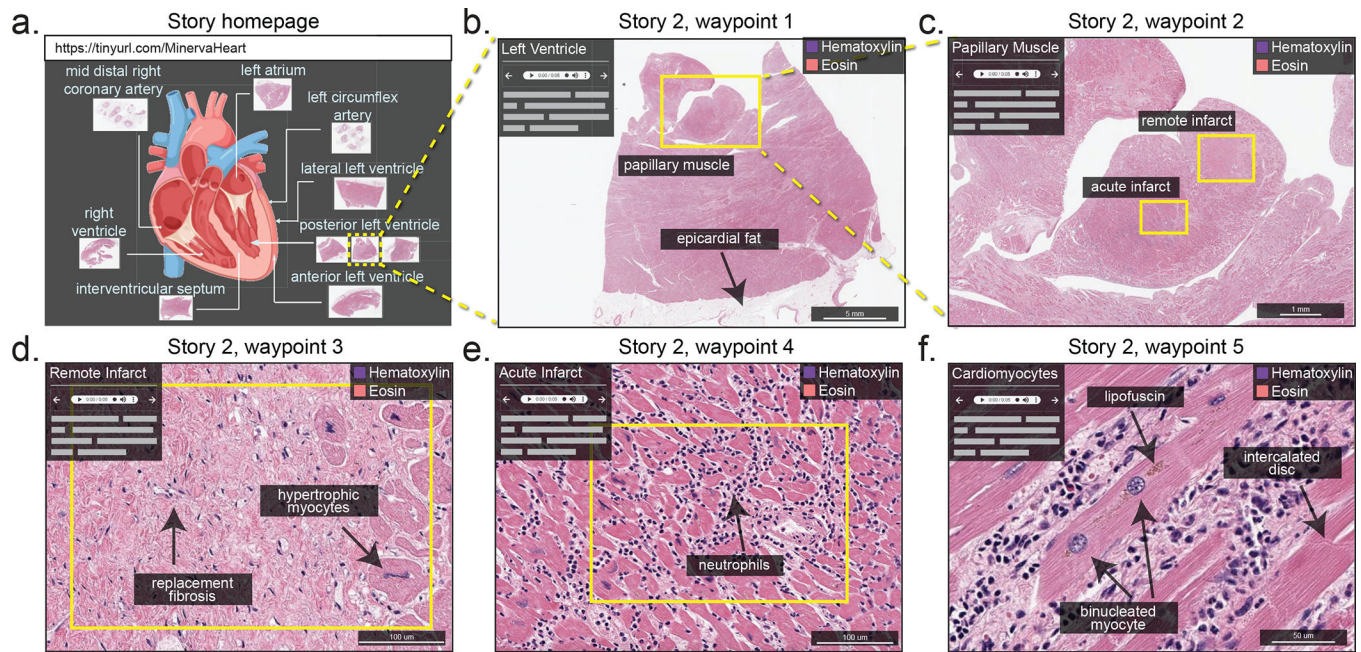


**Figure 4. The key features of the user interface of Minerva Story.**

Two stories exemplify the use of Minerva for a multiplexed image of a specimen of lung adenocarcinoma. Story 1 focuses on histology and immune populations and can be accessed at <https://www.cycif.org/MinervaLungHistology>. Story 2 focuses on data generation and analysis and can be accessed at <https://www.cycif.org/MinervaLungData>. a, The home screen for story 1. A narration panel on the left-hand side shows the title and narrative text, which can be read aloud by the software (through the use of the audio panel). Also, the left panel has links to related stories for the specimen and a table of contents (TOC) listing

each waypoint in the story (outlined in orange). Users can select the navigation arrows to step through each waypoint or skip to a waypoint of interest by using the table of contents. On the right-hand side, the channel-selection panel allows users to change which channels are rendered (outlined in pink). At any point, users can depart from a story and freely pan and zoom in or out of the image by using a mouse, trackpad or the magnification icons in the navigation panel. Arrowheads, concentric circles and other graphical elements can be used to annotate images and to generate web URLs specific to the current rendering. The viewer also contains a scale bar that grows and shrinks as a user navigates across zoom levels (grey). The screenshot shown includes only a subset of the waypoints and channel groups. Extensive information on these features is available at <https://github.com/labsyspharm/minerva-story/wiki>. b, Left: waypoint 6 from story 2 displays an interactive plot of a UMAP performed on a random sample of 2,000 cells from a tissue. Users can select any data point in the plot (each data point represents a single cell), and the browser will zoom to the position of that cell in the image and place an arrow on it. Middle: waypoint 3 from story 2 shows each cell in the tissue overlaid with a segmentation mask in which the colour denotes the cell type, as determined by quantitative k-means clustering. This makes it possible for users to assess the accuracy and the effects of segmentation on downstream cell-type-calling. The user can toggle image data and segmentation masks for different immunofluorescence channels on and off, and also overlay the clustering results and the image data in the same view. Waypoint 3 annotates two types of common segmentation errors: fission and fusion. Right: on-the-fly annotation tool. Users can add annotations and enter text and then hit the blue button to generate a URL that allows anyone to render the same image and location, preserving the panning and zoom settings, the marker group and the annotations. The screenshots displayed in this figure contain text of size and rendering that are optimized for presentation in this figure panel.





**Figure 5. Minerva Story for medical education.**

Minerva Story can be used for the guided viewing of conventional H&E-stained sections. For example, a Minerva story was created to guide students through tissue specimens from different anatomical regions of the heart of a patient who suffered multiple myocardial infarcts. The Story homepage shows an anatomical schematic of the heart, indicating the regions from which tissue specimens in the Minerva story were collected. Waypoint 1 shows a specimen sliced with a posterior view of the left ventricle, with box, arrow and text annotations of a few histological features. Waypoint 2 of the story shows a zoomed-in view of a papillary muscle, characteristic of the left ventricle, with annotations indicating regions of tissue showing acute (~12 days old) and remote (6 weeks to years old) myocardial infarction. Waypoint 3 and Waypoint 4 guide the user to the regions of remote infarct and acute infarct in the papillary muscle, describing histological hallmarks to distinguish the two types of infarcts. Waypoint 5 depicts microscopic cellular structures of cardiomyocytes at high resolution in a region of the tissue with a late acute infarct. This Minerva story is available at <https://www.cycif.org/MinervaHeart>.

**Table 1.**

Available atlases of normal and diseased human tissue

Atlas	Year Established	Microscopy	Lead Country	Atlas Link	Description/Goals
Human Genome Project	1990	–	International	<a href="https://www.genome.gov/human-genome-project">https://www.genome.gov/human-genome-project</a>	Map all genes of the human genome
FANTOM (Functional ANnotation Of the Mammalian genome)	2000	–	International	<a href="https://fantom.gsc.riken.jp/">https://fantom.gsc.riken.jp/</a>	Assign functional annotations to full-length cDNAs collected in Mouse Encyclopedia Project at RIKEN
ENCODE (Encyclopedia of DNA Elements)	2003	+	US	<a href="https://www.encodeproject.org/">https://www.encodeproject.org/</a>	Identify functional elements in the human genome
Human Protein Atlas (HPA)	2003	++	Sweden	<a href="https://www.proteinatlas.org/">https://www.proteinatlas.org/</a>	Map all human proteins in cells, tissues, and organs
Wellcome Trust Case Control Consortium (WTCCC)	2005	+	UK	<a href="https://www.wtccc.org.uk/">https://www.wtccc.org.uk/</a>	Understand patterns of human genome sequence variation to explore the utility, design and analyses of genome-wide association studies (GWAS)
The Cancer Genome Atlas (TCGA)	2005	+	US	<a href="https://portal.gdc.cancer.gov/">https://portal.gdc.cancer.gov/</a>	Obtain a comprehensive understanding of the genomic alterations that underlie all major cancers
International Cancer Genome Consortium (ICGC)	2007	–	International	<a href="https://icgc.org/">https://icgc.org/</a>	Unravel genomic changes present in many forms of cancer that contribute to the burden of disease
European Bioinformatics Institute (EMBL-EBI) Expression Atlas	2009	–	International/ EBI	<a href="https://www.ebi.ac.uk/gxa/home">https://www.ebi.ac.uk/gxa/home</a>	Database for querying differential gene expression across tissues, cell types and cell lines under various biological conditions
The Genotype-Tissue Expression (GTEx) project	2010	+	US	<a href="https://gtexportal.org/home/">https://gtexportal.org/home/</a>	Study how inherited changes in genes lead to common diseases
LungMAP	2010	+++	US	<a href="https://www.lungmap.net/">https://www.lungmap.net/</a>	Generate structural and molecular data on perinatal and postnatal lung development
BLUPRINT	2011	–	Netherlands/E U	<a href="https://www.blueprint-epigenome.eu/">https://www.blueprint-epigenome.eu/</a>	Understand how genes are activated or repressed in both healthy and diseased haematopoietic cells
Cancer Cell Line Encyclopedia (CCLE)	2012	–	US	<a href="https://depmap.org/portal/ccle/">https://depmap.org/portal/ccle/</a>	Compile gene expression, copy number and sequencing data from 947 human cancer cell lines
GenitoUrinary Development Molecular Anatomy Project (GUDMAP)	2015	++	US	<a href="https://www.gudmap.org/">https://www.gudmap.org/</a>	Develop tools to facilitate research on the genito-urinary (GU) tract
Accelerating Medicine Partnership (AMP)	2014	+	US	<a href="https://www.nih.gov/research-training/accelerating-medicines-partnership-amp">https://www.nih.gov/research-training/accelerating-medicines-partnership-amp</a>	Identify and validate biological targets for diagnostics & drug development

Atlas	Year Established	Microscopy	Lead Country	Atlas Link	Description/Goals
Human Cell Atlas	2016	++	International	<a href="https://www.humancellatlas.org/">https://www.humancellatlas.org/</a>	Develop reference maps of all human cells as a basis for diagnosing, monitoring, treating disease
Stimulating Peripheral Activity to Relieve Conditions (SPARC)	2016	++	US	<a href="https://sparc.science/">https://sparc.science/</a>	Accelerate development of devices that modulate neuronal activity to improve organ function
Pancreatlas	2016	+++	US	<a href="https://pancreatlas.org/">https://pancreatlas.org/</a>	Generate comprehensive images of the human pancreas
Human Tumor Atlas Network (HTAN)	2017	++	US	<a href="https://humantumoratlas.org/">https://humantumoratlas.org/</a>	Multidimensional molecular, cellular, and morphological mapping of human cancers
4D Nucleome (4DN)	2017	++	US	<a href="https://www.4dnucleome.org/">https://www.4dnucleome.org/</a>	Study the three-dimensional organization of the nucleus in space and time (the 4th dimension)
Kidney Precision Medicine Project/ Kidney Tissue Atlas	2017	++	US	<a href="https://atlas.kpmp.org/repository">https://atlas.kpmp.org/repository</a>	Study kidney biopsies to map heterogeneity and molecular pathways for drug discovery
Human BioMolecular Atlas Program (HuBMAP)	2017	++	US	<a href="https://hubmapconsortium.org/">https://hubmapconsortium.org/</a>	Develop an open and global platform to map healthy cells in the human body
Single Cell Expression Atlas	2018	–	International/ EBI	<a href="https://www.ebi.ac.uk/gxa/sc/home">https://www.ebi.ac.uk/gxa/sc/home</a>	Display gene expression in single cells from 12 species
BRAIN Initiative Cell Census Network (BICCN)	2018	+++	US	<a href="https://biccn.org/">https://biccn.org/</a>	Identify and provide experimental access to different brain cell types to determine their roles in health and disease
Oncobox Atlas of Normal Tissue Expression (ANTE)	2019	–	Russia	<a href="https://www.nature.com/articles/s41597-019-0043-4">https://www.nature.com/articles/s41597-019-0043-4</a>	Atlas of RNA sequencing profiles for normal human tissues
LifeTime Initiative	2019	+++	International/ EU	<a href="https://lifetime-initiative.eu/">https://lifetime-initiative.eu/</a>	Single-cell multi-omics & imaging, AI and patient-derived experimental disease models of health & disease
Pediatric Cell Atlas	2019	++	US	<a href="https://www.humancellatlas.org/pca/">https://www.humancellatlas.org/pca/</a>	Cytogenomic framework for study of pediatric health and disease.
Japanese cancer genome atlas (JCGA)	2019	–	Japan	<a href="https://pubmed.ncbi.nlm.nih.gov/31863614/">https://pubmed.ncbi.nlm.nih.gov/31863614/</a>	The Japanese TCGA
Gut Cell Atlas	2019	++	International	<a href="https://www.gutcellatlas.helmsleytrust.org/">https://www.gutcellatlas.helmsleytrust.org/</a>	Catalogue cell types in small, large intestines