Interpretative guides for interacting with tissue atlas and digital pathology data using the Minerva browser

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

Harvard Tissue Atlas Program

[1]Laboratory of Systems Pharmacology, Harvard Medical School, Boston, MA

[2]Ludwig Center for Cancer Research at Harvard, Harvard Medical School, Boston, MA

[3]Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA

[4]School of Engineering and Applied Sciences, Harvard University, Cambridge, MA

[5]Department of Systems Biology, Harvard Medical School, Boston, MA

* Co-corresponding authors:

Peter K. Sorger, Harvard Medical School, 200 Longwood Avenue, Warren Alpert Building, Room 440, Boston, MA 02115; Telephone: 617-432-6901; email: 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; email: ssantagata@bics.bwh.harvard.edu

Running Title: A Web-based viewer for tissue images

Competing interests

P.K.S. is on the SAB of RareCyte, Inc., whose product was used to acquire some of the image data in this perspective, and Glencoe Software, Inc., whose product was used to visualize some of the image data in this perspective. S.S. is a consultant for RareCyte, Inc.

ABSTRACT

The recent introduction of highly multiplexed imaging of human tissues and tumors promises to fundamentally advance research in tissue biology and human disease. At the same time, histopathology in the clinical setting is undergoing a rapid transition to digital methods. Thus, repositories of imaging data from research and clinical specimens will soon join genomic databases as a means to systematically explore the molecular basis of disease. Even with recent advances in machine learning, experience in anatomic pathology has shown that there is no substitute for expert visual review, annotation, and description of image data. We review the ecosystem of software available for atlas and histopathology images and introduce a new Web-based software tool, *Minerva Story*, that addresses a critical unmet need. *Minerva* is an interpretative and interactive guide to complex images organized around guided analysis. We discuss how *Minerva* and similar software will be integrated into multi-omic browsers for data dissemination of future atlases.

MAIN TEXT

Light microscopy is a traditional, powerful and widely used method for studying tissues at single-cell resolution and for diagnostic pathology. The relatively recent introduction of a diversity of methods for highly multiplexed tissue imaging[1–4] now makes it possible to conduct deep molecular profiling of tissues and tumors at subcellular resolution while preserving native tissue architecture.[5] This greatly facilitates the study of cell states, cell-cell interactions, and tissue architecture in normal and disease conditions. A common application of tissue imaging in oncology, for example, is identifying and enumerating immune cell types and mapping their locations relative to tumor and stromal cells[5]. Such spatially-resolved data is pertinent to understanding the mechanisms of action of immunotherapies (e.g., immune checkpoint inhibitors)[6] that function by blocking juxtracrine signaling between immune and tumor cells. Whole slide tissue images are complex however: when data are acquired using a fluorescence microscope, structures ranging in size by over five orders of magnitude are visible, from subcellular vesicles and nuclear granules to configurations of multiple cells such as blood vessels.

A wide variety of multiplexed tissue imaging methods have been described over the last few years. These include Imaging Mass Cytometry (IMC)[1] and Multiplexed Ion Beam Imaging (MIBI)[2], which detect antigens using metal isotope-labeled antibodies, tissue ablation, and atomic mass spectrometry. In contrast, methods such as MxIF[3], CODEX[7], t-CyCIF[4], multiplexed IHC[8], and immuno-SABER[9], use fluorescently-labelled (or enzyme-linked) antibodies followed by imaging on a fluorescence microscope. These methods differ in the number of antigens they can routinely detect on a single tissue section (currently ~12 in the case of multiplexed IHC to ~40-60 in the case of IMC or t-CyCIF); some methods are restricted to selected fields of view (typically a ~500 µm square) whereas others can perform whole slide imaging (WSI) on an area ~400-1,000 times larger. Most of these methods are in active development and their strengths and limitations with respect to speed, sensitivity, resolution etc. remain to be determined. However, they all share the feature of generating multiplexed 2D images of cells and supporting structures *in situ*. Methods that use high resolution microscopes can also generate 3D images, by optical sectioning for example.

Tissue and tumor atlases

Multiple international projects are currently underway to generate atlases of normal human cells, tissues, and tumors (**Figure 1**). These projects seek to combine the genetic and molecular precision of dissociative single-cell methods such as single-cell RNA sequencing with morphological and spatial

information obtained from tissue imaging and spatial transcriptomics[10]. At their inception, the imaging components of these atlases are likely to contain data acquired from one or a few individuals but they will evolve to merge multiple specimens into common reference systems. Conceptually, integration across samples and data types is easiest to imagine at the level of *derived features*, such as a census of cell types and positions from imaging data and transcript levels from scRNA-Seq. Adding meso-scale information from images such as the number and arrangement of supporting stroma, capsules and membranes, blood and lymphatic vessels, etc. is a greater challenge. It is not yet known how best to capture this information computationally or which features represent meaningful spatial interactions that mediate important biological processes. Studies are underway to better understand how pathologists make diagnoses from tissue specimens[11] and to quantify connections between features computed from cellular neighborhoods and clinical outcome[12]. However, human inspection of tissue images is almost certain to remain essential for biological understanding, judging the quality of image processing algorithms and training classifiers.

Tissue imaging in a clinical setting

The introduction of multiplexed tissue imaging in a research setting is occurring concurrently with an accelerating transition to digital technologies[13] in clinical histopathology (**Figure 1**). While genetics is pertinent to disease diagnosis, particularly in cancer and inherited diseases, histology and cytology remain the central pillars of routine clinical work and are the primary diagnostic modalities for most diseases. In current practice, tissue samples recovered by biopsy or surgical resection are formaldehyde-fixed and paraffin embedded (FFPE), sliced into 5μm sections, mounted on glass slides, and stained with hematoxylin and eosin (H&E). Samples that need to be examined immediately, when a patient is still undergoing surgery for example, are frozen, sectioned, and stained (these are often called OCT samples based on the: optimal cutting temperature" medium in which they are embedded). Blood cells are spread on a slide to create blood smears and then stained with colorimetric dyes (Romanowsky–Giemsa staining). H&E staining imparts a characteristic pink to blue color on cells and other structures that enables review by anatomical pathologists using a bright-field microscope. Some clinical samples are also subjected to immunohistochemistry (IHC) to obtain information on the expression of one or a small number of protein biomarkers per slide[14]. While cost-effective and widely established, many histopathology methods were developed well over a century ago, and IHC is itself 75 years old[15]. 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 genome sequencing (in a clinical setting, this usually involves exome sequencing).

Recently, pathologists have started to leverage machine learning to accelerate pattern recognition from histologic data and potentially extract deeper diagnostic insight. Digital analysis of histological specimens first became possible with the introduction of bright-field WSI instruments twenty years ago[16,17], but it was not until 2016 that the FDA released guidance on the technical requirements for use of digital imaging in diagnosis[18](**Figure 1**). Digital pathology has experienced dramatic growth in the past few years fueled in large part by the development of machine learning algorithms capable of assisting in the interpretation of H&E stained slides, which histopathology services must process in very high volume (often >1 million slides per year in a single hospital). Machine learning on images has proven successful in several areas of medicine[19,20] and promises to assist practitioners by increasing the efficiency and reproducibility of pathologic diagnosis[21]. On-line tools and datasets for digital pathology focus on H&E-stained specimens and bright field imaging, whereas atlas construction for research purposes is focused primarily on highly-multiplexed imaging. It is nonetheless highly desirable that software developed for clinical and research purposes be compatible and interoperable.

Accessing and sharing imaging data

Tissue atlases are envisioned to be spatially resolved counterparts of well-established genome-focused atlases such as The Cancer Genome Atlas (TCGA)[22] and Encyclopedia of DNA Elements (ENCODE)[23]. Unfortunately, algorithms, software, and standards for high-dimensional image data[24] remain under-developed relative to almost all types of genomic information. Moreover, with sequencing data, almost all of the information present in primary data files (e.g., FASTQ files) is retained (or enhanced) when reads are aligned and count tables are generated; it is rarely necessary to re-access the primary fastq files. In contrast, methods to extract features from tissue images are immature, and visual inspection by knowledgeable viewers remains essential to ensure the integrity of quantitative analysis and subsequent interpretation. It is therefore essential that tissue atlases provide ready access to image data at native resolution, even for casual use.

As far back as 2008, the *Journal of Cell Biology* worked with the founders of the Open Microscopy Environment (OME)[25] to deploy a JCB DataViewer[26] that 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[27], emphasizing 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; this format typically provides access to a few selected fields of view at a single resolution. In the best case these data might be available via Figshare[28]. The importance of sharing image data is particularly pronounced in the case of the research biopsies being used to assemble tumor atlases. These biopsies are obtained to advance scientific knowledge rather than inform the treatment of a specific patient and there is an ethical obligation for the resulting data (appropriately anonymized) to be made available in the most open and useful form possible for scientific discovery[29]. This principle is widely recognized in genomics[30,31] but has only recently been addressed in the area of tissue imaging[32]. More generally, digital pathology and tissue imaging are disciplines in which the goal of making research Findable, Accessible, Interoperable, and Reusable (FAIR)[33] is highly relevant, but the necessary computational infrastructure is deficient. For the vast majority of users, it is not reasonable to ask that they download terabyte-size files and install special software simply to view images.

Software for image analysis, management, and interpretation

To a first approximation, the wide variety of available academic and commercial microscopy software currently available has been implemented either as a desktop system focused on data analysis or, like OMERO, as a client-server (database) system focused on image management (**Figure 2a**; see **Box 1** and **Box 2** for details). Desktop software is particularly good for interactive image analysis because it exploits graphics cards for rapid image rendering and high-bandwidth connections to local data for computation. Client-server systems are ideal for data management because they enable sophisticated relational queries, ensure data integrity, and can access large-scale local and Cloud-based compute resources (a more sophisticated comparison of some of the available software is provided in **Figure 2b**).

As the first high-plex, whole-slide tissue images have become available, it has become clear that a new type of software will be required to guide users through the extraordinary complexity of images that encompass multiple square centimeters of tissue, 10^5-10^6 cells, and upwards of 100 channels. These "interpretative guides" walk users through a series of human and machine-generated annotations about an image in much the same way that the results section of paper guides users through a multi-panel figure. Genomic science faced an analogous need for efficient and intuitive visualization tools and this

lead to the development of the highly influential Integrative Genomics Viewer[34] and its many derivatives.

Minerva: a lightweight, narrative approach for sharing complex digital images

To facilitate interpretation and narration of complex tissue images, we developed a light-weight, Web-based, and open-source image viewer. Inspiration for *Minerva* came from observing how a Pathologist discusses a specimen with a colleague. This typically involves two or more people seated around a multi-headed optical microscope with the leader alternating the level of magnification between high and low power fields, panning around the specimen and highlighting key features with an LED pointer. *Minerva* aims to capitalize on this well-established, didactic approach. Further inspiration for story-based image viewing was provided by Project Mirador[35] which develops Web-based interpretative tours of cultural resources such as art museums, illuminated manuscripts, and even cities. In such online tours, a series of waypoints and accompanying text direct users to areas of interest while also allowing free exploration and a return to the narrative.

Minerva is designed for interpreted viewing of multiplexed, H&E, and IHC tissue images before and after publication. *Minerva* is based on OpenSeadragon[36]and displays zoomable images in a manner similar to Google Maps, making the interface familiar and intuitive. *Minerva* is OME- and BioFormats-compatible and therefore usable with images from virtually any existing microscope or scanner. *Minerva* is also simple to deploy, relying for its functionality on client-side JavaScript and standard commercial cloud storage (e.g. Amazon S3; **Figure 3**). Research teams can therefore share their own image data without the need for complex back-end software. *Minerva* is nonetheless inter-operable with client-server OMERO systems. Within a Minerva window on a standard Web browser, a narration panel directs a user's attention to specific regions of an image and specific channel groupings, accompanied by text description (which *Minerva* can read aloud) as well as image annotations involving overlaid rectangles, polygons, arrows, text, etc. (**Figure 4**). Each image can be associated with more than one narrated story and with different ways of viewing the same type of data.

Minerva is not an all-in-one tool. Instead it performs one task well: providing an intuitive interpretative guide to complex image data and them from 'postage-stamp' representations in journals. Genome browsers are similar: they do not perform alignment and data analysis, but they make it possible to interact with data effectively. OpenSeadragon has been used previously for displaying H&E images[37], but in *Minerva* it is paired with narrative features, a lightweight implementation, the ability to

accommodate both bright-field and multiplexed immunofluorescence images, and easy deployment. Anyone familiar with GitHub and AWS can get *Minerva* running in a few hours. *Minerva* currently comprises both the primary user-facing interface (*Minerva Story*) and *Minerva Author*, which converts images in the BioFormats-compatible OME-TIFF format[38] into pyramidal form and assists with the addition of waypoints and text annotations. After specifying rendering settings and writing the waypoints in *Minerva Author*, a user receives a configuration file and an image pyramid to deploy to AWS S3 or another Web-based storage location (**Figure 3**).

Using Minerva Story to explore a human tumor

As a use case for *Minerva Story* we generated two narrated stories about human lung adenocarcinoma imaged with t-CyCIF. This primary tumor of ~ 5 mm x 3.5 mm in dimension was imaged using 44-plex t-CyCIF[39] and the properties of resulting data have been described in detail a recent publication (it is referred to "LUNG-3-PR" in reference 40). The two stories associated with this image focus on (i) histologic regions of interest and specific immune and tumor cell types (**Figure 4a** left) (ii) quantitative data analysis. The name of the samples, links to related stories, a table of contents, and navigation tools are found in a panel to the left of the *Minerva Story* window (outlined in yellow in **Figure 4b**). Another navigation panel, focused on channel selection, is found on the right of the window (outlined in pink); channels and channels groups, as well as the cell types they define, can be pre-specified. We typically link each protein (antigen) name to an explanatory such as Genecards, but a more customized annotation of markers would be preferable and is under development. Users can interactively highlight areas of interest, add notes and recover sharable link that allows others to return to the same position in the image and also view any added annotations and explanatory text (**Figure 4a right;** in this case a high magnification view is shown with a T-cell labelled).

Stories progress from one waypoint to the next (the analogy is to numbers 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 gray). At any point users can diverge from the story by panning around the image or selecting different magnifications and channels, and then return to the narrative by clicking on the appropriate waypoint in the table of contents. In the image field shown in **Figure 4b**, the story is at waypoint one and shows pan-cytokeratin positive tumor cells growing in chords and clusters at the tumor-stromal interface. This interface is characterized by an inflammatory microenvironment as evidenced by the presence of a variety of

lymphocyte and macrophage populations distinguishable by their expression of cell surface markers; to the right of the viewing panels users can toggle these markers on and off to judge for themselves the accuracy of classification. The remainder of the story explores the expression of PD-L1 (Programmed death-ligand 1), an immune checkpoint protein and drug target.

Minerva can also be used to display the results of quantitative data analysis performed using other tools (this constitutes the second story about this lung adenocarcinoma). Data analysis typically quantifies the expression cell type markers (e.g. immune linear markers) and uses this to identify cell types. Analysis of cell state, morphology, and neighborhood relationships are also performed. Within *Minerva Story* it is possible to link representations of quantitative data directly to relevant image features. For example, in a heat map representing *k*-means clustering of keratin, αSMA, and CD45 levels in single cells, three canonical cell types can be identified (tumor, stromal, immune and other; this is waypoint 2 in story 2). Clicking on the heat map brings up a mask for the cells belonging to that group, overlaid on an immunofluorescence image (**Figure 5a**). As in story one, users can zoom and pan to inspect specific cell types and toggle markers or groups of markers on and off to explore their relationship to this cell type. When data are captured in a scatterplot, such as a UMAP (**Figure 5b**) clicking on a data point takes the user directly to the corresponding position in the image with the cell denoted by an arrow. This is a standard feature in desktop software such as HistoCat and Facetto and greatly enhances a user's understanding of the relationship between images and image-derived features. Display of segmentation masks as overlays is similarly useful for troubleshooting and assessing data quality; in Figure unwanted fusion and fission events are highlighted, both of which results from errors in segmentation (**Figure 5c**).

Teaching is another application for *Minerva.* Histology is challenging to teach in an undergraduate setting[41] and, in the case of medical students and residents, changes in curriculum have resulted in much less time available in front of a microscope. On-line collections of tissue images are a frequent substitute. However, studies have shown that pairing classroom instruction with dynamic viewing and interaction with image data are essential for learning[42]. We demonstrate an application of *Minerva Story* in narrating H&E images of specimens from the heart of a patient who experienced multiple episodes of myocardial infarction prior to death (**Figure 6**). An introductory panel depicts the overall structure of the heart and the positions from which various specimens were resected. These images reveal the histologic hallmarks of ischemic heart disease such as severe coronary artery atherosclerosis, plaque rupture, stunned myocardium, reperfusion injury, and the early, intermediate, and

late features of myocardial tissue infarction. The interactive narration of this common clinical syndrome provides a context for developing a more nuanced understanding of cardiac pathophysiology than looking at snippets in a textbook or a vaguely annotated sets of on-line images.

Future directions for Minerva and tissue atlas viewers

Recent interest in single-cell tissue biology derives less from the advances in microscopy described above and more from the introduction of single cell sequencing[43,44]. Comprehensive characterization of normal and diseased tissue will almost certainly involve the integration of data from multiple analytical modalities, including imaging, spatial transcript profiling,[45,46] mass spectrometry imaging of metabolites and drugs[47,48], and computational registration of dissociated scRNA-seq[49] with spatial features. *Minerva* does not address all of these visualization channels but because it is built around widely used open-source software tools and libraries (e.g. OpenseaDragon and JavaScript), it can easily be combined with other tools to create the multi-omic viewers that will be needed for tissue atlases. It is also compatible with any image repository that uses the OME-TIFF standard. As currently configured, *Minerva* does not require a backend server or specialized infrastructure. This makes it very easy to deploy but also limits its ability to efficiently access large numbers of images. *Minerva* is OMERO-compatible however, and we are working on variants that will provide search and indexing functionality. *Minerva* could also be added to powerful genomics tools such as cBio Portal.[50] With respect to clinical applications, the broad biomedical community needs to work to ensure that digital pathology systems do not become locked behind proprietary data formats and non-interoperable software. The OME and BioFormats demonstrate that is possible for academic developers and commercial instrument and software vendors to work together for mutual benefit. The availability of software such as *Minerva* with a non-restrictive open-source license is one step in realizing an interoperable and FAIR future for digital pathology.

DATA AVAILABILITY

Source code for *Minerva Story* can be found at https://github.com/labsyspharm/minerva-story and detailed documentation at https://labsyspharm.github.io/minerva-story-wiki/.

Box 1: Software for Managing and Visualizing Image Data

The OME-based *OMERO*[51] server remains the most widely used image informatics system for microscopy data in a research setting; it is the foundation of the European Bioinformatics Institute (EBI)-based Image Data Resource (IDR)[52], a prominent, large-scale publicly accessible image repository, as well as more specialized repositories such as Pancreatlas[53]. *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 organizing images so that they can be queried using a visual index or via search[54](**Figure 2a**). In its current form, it does not perform sophisticated image analysis.

A range of other software is available for static or partially interactive visualization of H&E and IHC images in a Web browser, including caMicroscope[55] which is used to organize IHC and H&E images for The Cancer Genome Atlas (TCGA)[22], the Cancer Imaging Archive (TCIA),[56] Digital Slide Archive[32], PathPresenter[57], and Aperio[58] (**Figure 2b**). Interested readers are referred to a recent White Paper from the Digital Pathology Association that discusses tools being developed to view bright-field digital pathology data[37]. However, such H&E-IHC viewers are not generally designed to display multi-channel images and integrate different types of omic data.

Box 2: Software for Analysis of Image Data

Multiple software suites have been developed for analysis of high-dimensional image data including *CellProfiler*,[59] *histoCAT*[60], *Facetto*[61], *QuPath*[62], *Orbit*[63], Mantis[64], and *ASAP*[65]; many but not all run locally on the desktop (**Figure 2b**) . These software systems generally perform image segmentation to identify individual cells or tissue-level features, determine cell centroids and shape parameters (e.g., 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[66], or UMAP[67] to identify cell types and study cell-cell interactions[68]. While some tools support segmentation, others require pre-generated segmentation masks and single-cell feature tables. As an alternative, emerging approaches analyze tissues at the level of individual pixels using machine learning and CNNs[69]. This approach potentially bypasses the need to segment individual cells from densely-packed tissues in which cells can vary dramatically in size. A key feature of software such as *histoCat*[60] or *Facetto*[70] is

integration of an image viewer and feature-based representations of the same data. This is essential for training and testing classifiers, quality-controlling image processing routines, and obtaining insight into spatial characteristics.

Multiple companies are developing Cloud-based computing platforms for digital pathology, including *HALO* (Indica Labs), *Visiopharm* (Visiopharm), and *PathViewer* (Glencoe Software, the commercial developer of *OMERO*). Public domain academic efforts include the *Allen Cell Explorer*[71] and the Chan Zuckerberg Initiative's *Napari*[72] and build on highly successful open source software platforms such as *ImageJ*[73]. Commercial tools often strive for an all-in-one approach to analysis and visualization, but this comes at the cost of complexity, proprietary implementation, and licensing fees. It also ignores one of the primary lessons from genomics: progress has rarely been based on a single integrated software suite, but instead has involved an ecosystem of interoperable tools specialized to specific tasks. This approach is not perfect[74], but it cannot reasonably be replaced by a single piece of commercial code.

FIGURE LEGENDS

Figure 1. Timeline of milestones in development of human atlases, histopathology, and multiplexed tissue imaging technologies. The upper timeline depicts the establishment of publicly available atlases of human tissues, tumors and cell types (gray denotes genome-focused atlases and blue denotes atlases with a substantial imaging component). Each dot represents the year that one new large-scale atlas was established. The number of atlases is increasing as is the emphasis on imaging data. The lower timeline depicts the evolution of diagnostic methods used for anatomical pathology in a clinical setting, with an accelerating trend towards digital methods occurring in parallel with growth of multiplexed imaging and atlases. AI: Artificial intelligences; IHC: immunohistochemistry; FDA: Food and Drug Administration; GPU: graphics processing unit; TCGA the Cancer Genome Atlas.

Figure 2. Software used to visualize, analyze, manage, and share tissue images. A variety of software is are available for processing tissue image data, each with strengths and weaknesses. We expect them to co-evolve as part of an ecosystem held together by common data standards and interoperable application programming interfaces (APIs). **(a)** Desktop applications such as *histoCat* provide sophisticated tools for interactive quantification and analysis of image data and effectively

exploit embedded image rendering capabilities. Many desktop applications can be run in a "headless" configuration on compute farms to accelerate analysis of large images. They can also fetch images from *OMERO*. Three-tier client server systems such as *OMERO* involve a relational database management system (RDMS), an image server, and one or more browser-based and desktop clients; such systems provide a full-featured approach to organizing, visualizing, and sharing large numbers of images but they typically support only limited data processing. The *Minerva* application described in this perspective is designed for simplicity and optimized for data sharing and interpretative image vieing: within *Minerva* detailed narration of images and derived data is possible. Access to primary data (bold arrows) is primarily to Cloud-based storage (e.g. an Amazon Web Service S3 bucket) but *Minerva* is being developed to also function as an OMERO client. Public access to large-scale primary data is a typically limited with all of these systems because of the costs of data transfer. **(b)** The range of commercial and academic software suitable for viewing image data with key features described. While each tool has its strengths and weaknesses, none of these tools comprehensively satisfy all of the range of functions needed for image analysis.

Figure 3. System architecture of *Minerva*. Whole slide images including multi-channel microscopy images in the standard OME-TIFF format are imported into *Minerva Author* in which the authoring user interacts with the primary data to set channel rendering settings, specify waypoints in stories, and add text and graphical annotation. *Minerva Author* then renders image pyramids for all channel groupings and a generates a configuration file that specifies waypoints and their associated information. The rendered images are stored on a cloud host (e.g. Amazon Web Services S3) which *Minerva Story* reads to render to the client and renders using client-side JavaScript (.JS). Because rendering and interaction is performed on the client, no special backend software is needed. Images rendered by *Minerva* story (in JPEG) are compatible with multiple devices, including cell phones. The thickness of the arrows indicates the size of data transferred.

Figure 4. *Minerva Story*. Screenshots of the *Minerva Story* user interface showing key features. Note that the size and rendering of the text has been altered from the original to make the interface easier to understand in a figure panel. The original story can be accessed at https://tinyurl.com/MinervaLungHistology. (a) The panel on the left depicts the two *Minerva* stories available for a single multiplexed image of a lung adenocarcinoma specimen; note that the screenshot

has been manipulated to emphasize specific features. On the right is a field of interest that a user has highlighted; the user can enter text and then hit the red button to generate a URL that allows anyone to render the same image and location with pan, zoom, marker group, and annotations preserved. (b) The home screen for *Minerva* story one. A narration panel on the left hand side shows the title, narrative text that can also be read aloud using the audio panel, links to related stories for the specimen, and a table of contents outlining each waypoint in the story (outlined in yellow). Users can click the navigation arrows to step through each waypoint or skip to a waypoint of interest using the table of contents. On the right side is channel selection panel that users can use to change which channels is rendered (outlined in pink). At any point, users can depart from a story and freely pan and zoom in or out of the image using the mouse or the magnification icons in the navigation panel. Arrowheads, concentric circles, and other graphical elements can be used to annotate images and 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). Note that this synthetic screenshot includes only a subset of the waypoints and channel groups; the full tool can be accessed at https://tinyurl.com/MinervaLungHistology.

Figure 5. *Minerva Story* of the data analysis of a lung tumor. In a second Minerva story (Story 2) generated for same specimen as in Figure 4, users are guided through the data generation and analysis for this tissue. Story 2 can be accessed at https://tinyurl.com/MinervaLungData. Note that all of the analysis was performed prior to creation of the interpretative guide by the authoring users. (a) Waypoint 4 of the story depicts a k-means clustergram and heat map for all cells in the tissue based on the quantification of staining for Keratin, a-SMA, and CD45. Users can click on any cluster, and the browser will display the appropriate immunofluorescence channel with an overlaid mask highlighting cells that belong to that cluster. (b) Waypoint 6 displays an interactive plot of a UMAP (Uniform Manifold Approximation and Projection for Dimension Reduction) performed on a random sample of 2,000 cells from the tissue. Users can click on 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. (c) Waypoint 3 shows each cell in the tissue overlaid with a segmentation mask in which the color denotes the cell type as determined by quantitative k-means clustering (as showing waypoint 4). This makes it possible for users to assess the accuracy and effects of segmentation on downstream cell type calling analysis. The reader can toggle the mask with immunofluorescence channels to view the quantitative and image data in the same view. The waypoint annotates two types of common segmentation errors.

Figure 6. An application of *Minerva Story* in medical education. *Minerva Story* can be used for multiple applications, including guided viewing of conventional H&E specimens. In this example available at https://tinyurl.com/MinervaHeart, a Minerva story has been created to guide students through tissue specimens from different anatomical regions of a heart from a patient who suffered multiple myocardial infarcts. (a) An anatomic schematic of the heart depicts the regions from which tissue specimens in the *Minerva* story were collected. (b) Waypoint 1 shows a specimen sliced with a posterior view of the left ventricle which box, arrow, and text annotations of few histologic features. (c) Waypoint 2 of the story in (b) showing 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 infarct. (d) Waypoint 3 and (e) Waypoint 4 guide the user to the remote infarct and acute infarct regions of the papillary muscle describe histologic hallmarks to distinguish the two types of infarcts. (f) 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://tinyurl.com/MinervaHeart.

ACKNOWLEDGEMENTS

This work was funded by NIH grants U54-CA225088 and U2C-CA233262 to P.K.S. and S.S., by U2C-CA233280 to P.K.S., and by the Ludwig Cancer Center at Harvard. The Dana-Farber/Harvard Cancer Center is supported in part by an NCI Cancer Center Support Grant P30-CA06516.

OUTSIDE INTERESTS

PKS is a founder, SAB member and equity holder in Merrimack Pharmaceutical Glencoe Software; he is a member of the Board of Directors of Applied Biomath and Glencoe and SAB of RareCyte Inc. In the last five years the Sorger lab has received research funding from Novartis and Merck. Sorger declares that none of these relationships have influenced the content of this manuscript. SS is a consultant for RareCyte Inc. The other authors declare no outside interests.

REFERENCES

1. Giesen, C. *et al.* Highly multiplexed imaging of tumor tissues with subcellular resolution by mass cytometry. *Nat. Methods* **11**, 417–422 (2014).
2. Angelo, M. *et al.* Multiplexed ion beam imaging of human breast tumors. *Nat. Med.* **20**, 436–442 (2014).
3. Gerdes, M. J. *et al.* Highly multiplexed single-cell analysis of formalin-fixed, paraffin-embedded cancer tissue. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 11982–11987 (2013).
4. Lin, J.-R. *et al.* Highly multiplexed immunofluorescence imaging of human tissues and tumors using t-CyCIF and conventional optical microscopes. *eLife* (2018) doi:10.7554/eLife.31657.
5. Bodenmiller, B. Multiplexed Epitope-Based Tissue Imaging for Discovery and Healthcare Applications. *Cell Syst* **2**, 225–238 (2016).
6. Coy, S. *et al.* Multiplexed Immunofluorescence Reveals Potential PD-1/PD-L1 Pathway Vulnerabilities in Craniopharyngioma. *Neuro-oncology* (2018) doi:10.1093/neuonc/noy035.
7. Goltsev, Y. *et al.* Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed Imaging. *Cell* **174**, 968-981.e15 (2018).
8. Stack, E. C., Wang, C., Roman, K. A. & Hoyt, C. C. Multiplexed immunohistochemistry, imaging, and quantitation: A review, with an assessment of Tyramide signal amplification, multispectral imaging and multiplex analysis. *Methods* **70**, 46–58 (2014).
9. Saka, S. K. *et al.* Immuno-SABER enables highly multiplexed and amplified protein imaging in tissues. *Nat. Biotechnol.* **37**, 1080–1090 (2019).
10. HuBMAP Consortium. The human body at cellular resolution: the NIH Human Biomolecular Atlas Program. *Nature* **574**, 187–192 (2019).
11. Shin, D. *et al.* PathEdEx - Uncovering High-explanatory Visual Diagnostics Heuristics Using Digital Pathology and Multiscale Gaze Data. *J Pathol Inform* **8**, 29 (2017).

12. Wagner, J. *et al.* A Single-Cell Atlas of the Tumor and Immune Ecosystem of Human Breast Cancer. *Cell* **177**, 1330-1345.e18 (2019).

13. Digital and Computational Pathology: Bring the Future into Focus. *Journal of Pathology Informatics* **10**, 10 (2019).

14. Brown, M. & Wittwer, C. Flow cytometry: principles and clinical applications in hematology. *Clin. Chem.* **46**, 1221–1229 (2000).

15. Immunologists, A. A. of. The Demonstration of Pneumococcal Antigen in Tissues by the Use of Fluorescent Antibody. *The Journal of Immunology* **45**, 159–170 (1942).

16. 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 Inform* **9**, 40 (2018).

17. Gurcan, M. N. *et al.* Histopathological image analysis: a review. *IEEE Rev Biomed Eng* **2**, 147–171 (2009).

18. Health, C. for D. and R. Technical Performance Assessment of Digital Pathology Whole Slide Imaging Devices. *U.S. Food and Drug Administration* http://www.fda.gov/regulatory-information/search-fda-guidance-documents/technical-performance-assessment-digital-pathology-whole-slide-imaging-devices (2019).

19. 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).

20. Greenspan, H., van Ginneken, B. & Summers, R. M. Guest Editorial Deep Learning in Medical Imaging: Overview and Future Promise of an Exciting New Technique. *IEEE Transactions on Medical Imaging* **35**, 1153–1159 (2016).

21. 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).

22. The Cancer Genome Atlas Research Network *et al.* The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* **45**, 1113–1120 (2013).

23. Consortium, T. E. P. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **306**, 636–640 (2004).

24. 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).

25. Goldberg, I. G. *et al.* The Open Microscopy Environment (OME) Data Model and XML file: open tools for informatics and quantitative analysis in biological imaging. *Genome Biology* **6**, R47 (2005).

26. Hill, E. Announcing the JCB DataViewer, a browser-based application for viewing original image files. *J Cell Biol* **183**, 969–970 (2008).

27. Swedlow, J. R., Goldberg, I., Brauner, E. & Sorger, P. K. Informatics and quantitative analysis in biological imaging. *Science* **300**, 100–102 (2003).

28. Singh, J. FigShare. *J Pharmacol Pharmacother* **2**, 138–139 (2011).

29. Levit, L. A. *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).

30. Kaye, J., Heeney, C., Hawkins, N., de Vries, J. & Boddington, P. Data sharing in genomics--re-shaping scientific practice. *Nat. Rev. Genet.* **10**, 331–335 (2009).

31. Reardon, J. *et al.* Bermuda 2.0: reflections from Santa Cruz. *Gigascience* **5**, 1–4 (2016).

32. Gutman, D. A. *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).

33. Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3**, 160018 (2016).

34. Robinson, J. T. *et al.* Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).

35. Mirador — Home. https://projectmirador.org/.

36. OpenSeadragon. https://openseadragon.github.io/.

37. Aeffner, F. *et al.* Introduction to Digital Image Analysis in Whole-slide Imaging: A White Paper from the Digital Pathology Association. *J Pathol Inform* **10**, 9 (2019).

38. Hiner, M. C., Rueden, C. T. & Eliceiri, K. W. SCIFIO: an extensible framework to support scientific image formats. *BMC Bioinformatics* **17**, 521 (2016).

39. Lin, J.-R. *et al.* Highly multiplexed immunofluorescence imaging of human tissues and tumors using t-CyCIF and conventional optical microscopes. *eLife Sciences* **7**, e31657 (2018).

40. 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).

41. García, M., Victory, N., Navarro-Sempere, A. & Segovia, Y. Students' Views on Difficulties in Learning Histology. *Anat Sci Educ* **12**, 541–549 (2019).

42. Mione, S., Valcke, M. & Cornelissen, M. Remote histology learning from static versus dynamic microscopic images. *Anat Sci Educ* **9**, 222–230 (2016).

43. Macaulay, I. C., Ponting, C. P. & Voet, T. Single-Cell Multiomics: Multiple Measurements from Single Cells. *Trends Genet* **33**, 155–168 (2017).

44. Trapnell, C. Defining cell types and states with single-cell genomics. *Genome Res* **25**, 1491–1498 (2015).

45. Ståhl, P. L. *et al.* Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016).

46. 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**, 531 (2018).

47. Fusion of mass spectrometry and microscopy: a multi-modality paradigm for molecular tissue mapping. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4382398/.

48. Molecular imaging of drug transit through the blood-brain barrier with MALDI mass spectrometry imaging. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3790202/.

49. Moncada, R. *et al.* Integrating single-cell RNA-Seq with spatial transcriptomics in pancreatic ductal adenocarcinoma using multimodal intersection analysis. *bioRxiv* 254375 (2019) doi:10.1101/254375.

50. Cerami, E. *et al.* The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data. *Cancer Discov* **2**, 401–404 (2012).

51. Allan, C. *et al.* OMERO: flexible, model-driven data management for experimental biology. *Nat. Methods* **9**, 245–253 (2012).

52. Williams, E. *et al.* The Image Data Resource: A Bioimage Data Integration and Publication Platform. *Nat. Methods* **14**, 775–781 (2017).

53. Browse & Filter Dataset -- Pancreatlas / HANDEL-P. https://pancreatlas.org/.

54. Rubin, D. L., Greenspan, H. & Brinkley, J. F. Biomedical Imaging Informatics. in *Biomedical Informatics: Computer Applications in Health Care and Biomedicine* (eds. Shortliffe, E. H. & Cimino, J. J.) 285–327 (Springer, 2014). doi:10.1007/978-1-4471-4474-8_9.

55. caMicroscope. *GitHub* https://github.com/camicroscope.

56. Clark, K. *et al.* The Cancer Imaging Archive (TCIA): Maintaining and Operating a Public Information Repository. *J Digit Imaging* **26**, 1045–1057 (2013).

57. PathPresenter. https://public.pathpresenter.net/#/login.

58. Olson, A. H. Using the Aperio ScanScope ®. in (2006).

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

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

61. Krueger, R. *et al.* Facetto: Combining Unsupervised and Supervised Learning for Hierarchical Phenotype Analysis in Multi-Channel Image Data. *IEEE Transactions on Visualization and Computer Graphics* **26**, 227–237 (2020).

62. Bankhead, P. *et al.* QuPath: Open source software for digital pathology image analysis. *Sci Rep* **7**, 1–7 (2017).

63. Stritt, M., Stalder, A. K. & Vezzali, E. Orbit Image Analysis: An open-source whole slide image analysis tool. *bioRxiv* 731000 (2019) doi:10.1101/731000.

64. Mantis Viewer. *Mantis Viewer* https://mantis.parkerici.org/.

65. ASAP - Automated Slide Analysis Platform. https://computationalpathologygroup.github.io/ASAP/#home.

66. Maaten, L. van der & Hinton, G. Visualizing Data using t-SNE. *Journal of Machine Learning Research* **9**, 2579–2605 (2008).

67. McInnes, L., Healy, J. & Melville, J. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv:1802.03426 [cs, stat]* (2018).

68. Arnol, D., Schapiro, D., Bodenmiller, B., Saez-Rodriguez, J. & Stegle, O. Modelling cell-cell interactions from spatial molecular data with spatial variance component analysis. *bioRxiv* 265256 (2018) doi:10.1101/265256.

69. Berg, S. *et al.* ilastik: interactive machine learning for (bio)image analysis. *Nat. Methods* (2019) doi:10.1038/s41592-019-0582-9.

70. Krueger, R. *et al.* Facetto: Combining Unsupervised and Supervised Learning for Hierarchical Phenotype Analysis in Multi-Channel Image Data. *bioRxiv* 722918 (2019) doi:10.1101/722918.

71. The Allen Cell Structure Segmenter: a new open source toolkit for segmenting 3D intracellular structures in fluorescence microscopy images | bioRxiv. https://www.biorxiv.org/content/10.1101/491035v1.

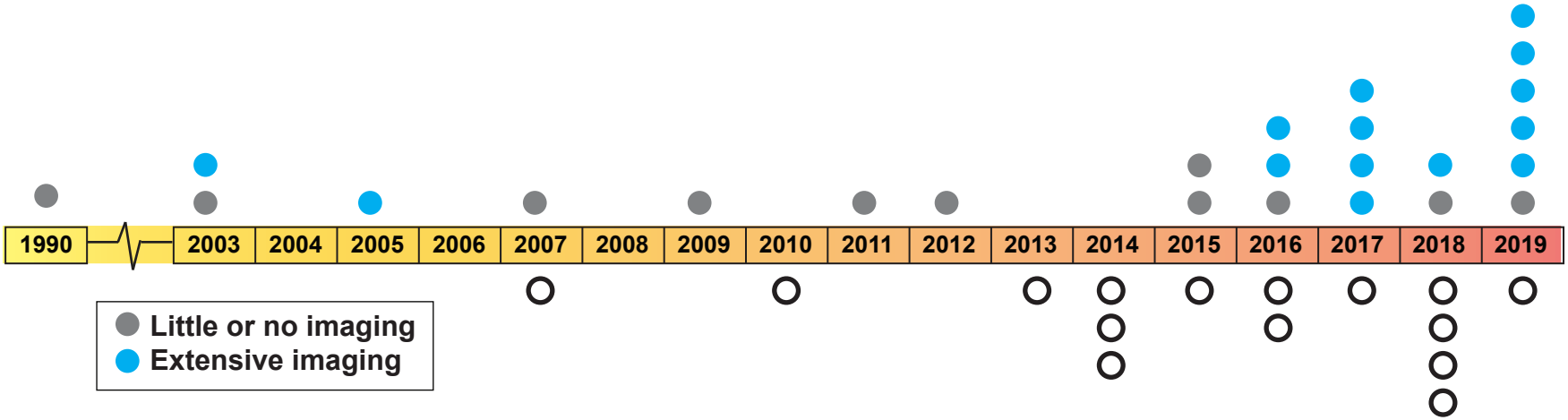
72. *napari/napari*. (napari, 2019).

73. Schindelin, J., Rueden, C. T., Hiner, M. C. & Eliceiri, K. W. The ImageJ ecosystem: An open platform for biomedical image analysis. *Molecular Reproduction and Development* **82**, 518–529 (2015).

74. Siepel, A. Challenges in funding and developing genomic software: roots and remedies. *Genome Biol.* **20**, 147 (2019).

Figure 1

Timeline of tumor and tissue atlases (above) and multiplexed tissue imaging technologies (below)



Milestones in histopathology and AI-visual recognition

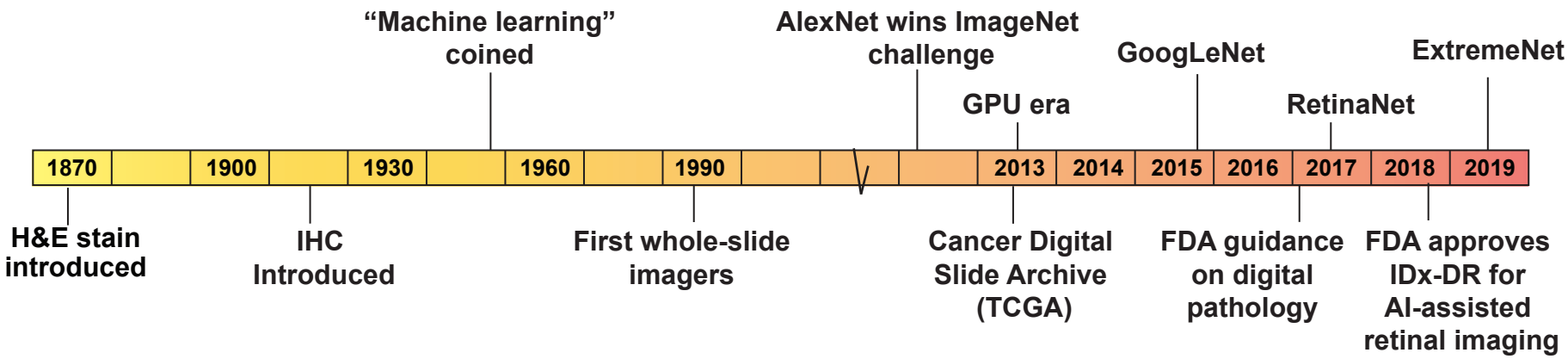
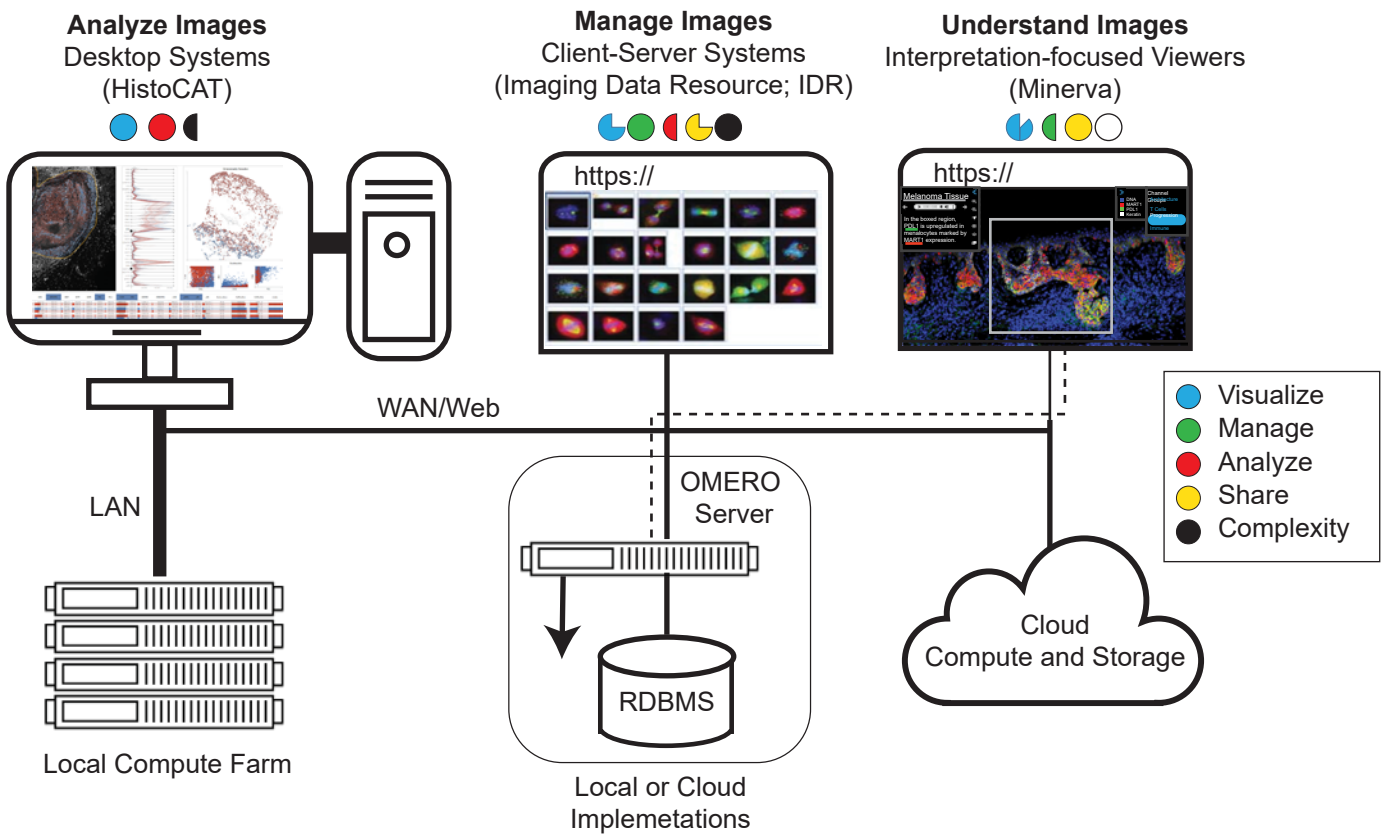


Figure 2

a.



b.

	Software	Description	Visualize: Pyramidal Structrue	Annotation	Brightness/Contrast	Integrated Data	Manage: Desktop	Web-Based	Cloud-Based	Client Server	OMERO-Based	Bio-Formats	Analyze: Tissue Segmentation	Cell Segmentation	Single-Cell	Subsetting	Plotting	Machine Learning	Share: Open Acess	Open Source	Commercial	Web-Based Exporting	File-Based Exporting	Crowd-Source Annotation	Narration	Audio
Multi-Channel Image	Aperio	image viewer	●	●	●	●	●						●	●	●	●					●		●			
	napari	image viewer	●	●	●	●	●												●	●						
	Pancreas Atlas	image viewer	●		●	●		●	●	●	●	●							●			●				
	Pathomation	image viewer				●	●	●		●		●							●	●						
	Pathviewer (OMERO)	image viewer	●	●	●	●		●	●	●	●	●							●		●	●	●	●		
	OMERO	image management	●	●	●	●	●	●	●	●	●	●							●	●	●	●	●			
	Proscia	image management	●	●	●	●		●	●			●									●		●	●		
	ASAP	image analysis	●	●		●	●							●	●				●	●			●			
	Facetto	image analysis	●	●	●	●	●					●			●	●	●	●	●	●						
	HALO	image analysis			●	●	●					●	●	●	●	●	●	●			●		●			
	HistoCAT	image analysis		●	●	●	●					●			●	●	●	●	●				●			
	Mantis	image analysis	●	●	●	●	●					●			●	●	●		●	●						
	Orbit	image analysis	●	●	●	●	●			●	●	●	●	●	●	●		●	●	●			●			
	QuPath	image analysis	●	●	●	●	●		●	●	●	●	●	●	●	●	●	●	●	●			●			
	Visiopharm	image analysis	●	●	●	●	●			●		●	●	●	●	●	●	●			●	●	●			
	Minerva	image data sharing	●	●		●		●	●			●			●		●		●	●		●		●	●	●
Brightfield H&E Image	Digital Slide Archive	image viewer	●	●				●	●	●		●							●	●						
	OpenSlide	image viewer	●				●	●	●	●		●							●	●			●			
	Image Data Resource	image management	●		●		●	●	●	●	●	●							●	●		●	●			
	caMicroscope	image analysis	●	●		●		●	●	●		●	●		●	●			●	●			●			
	Cytomine	image analysis	●	●	●	●		●	●	●									●	●		●	●	●		
	PathPresenter	image data sharing	●	●	●			●	●	●		●							●		●	●	●		●	

Figure 3

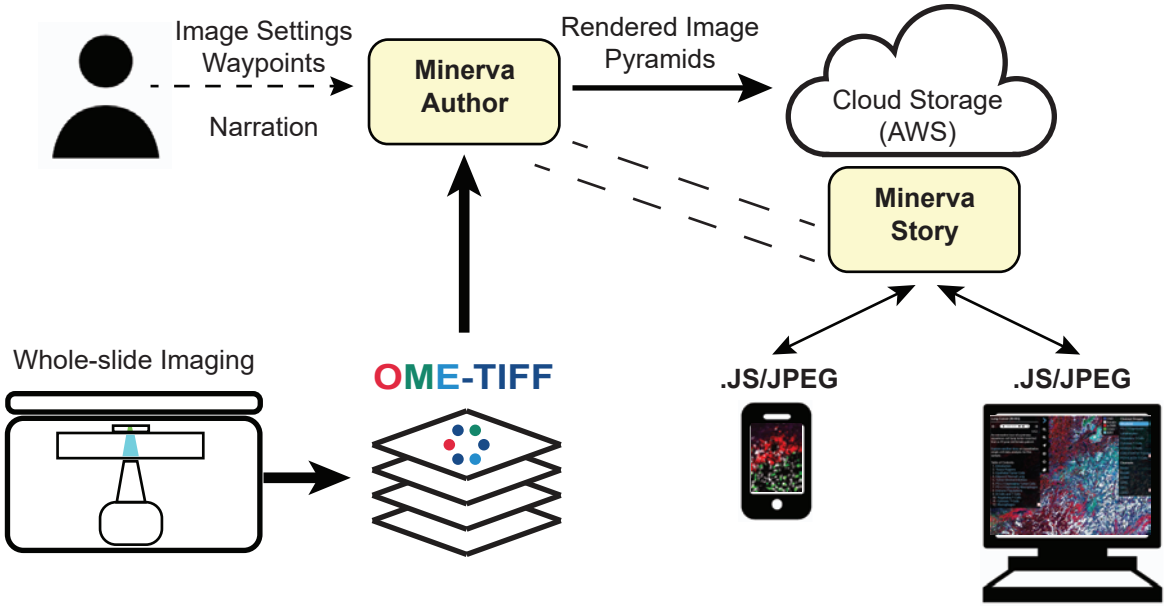
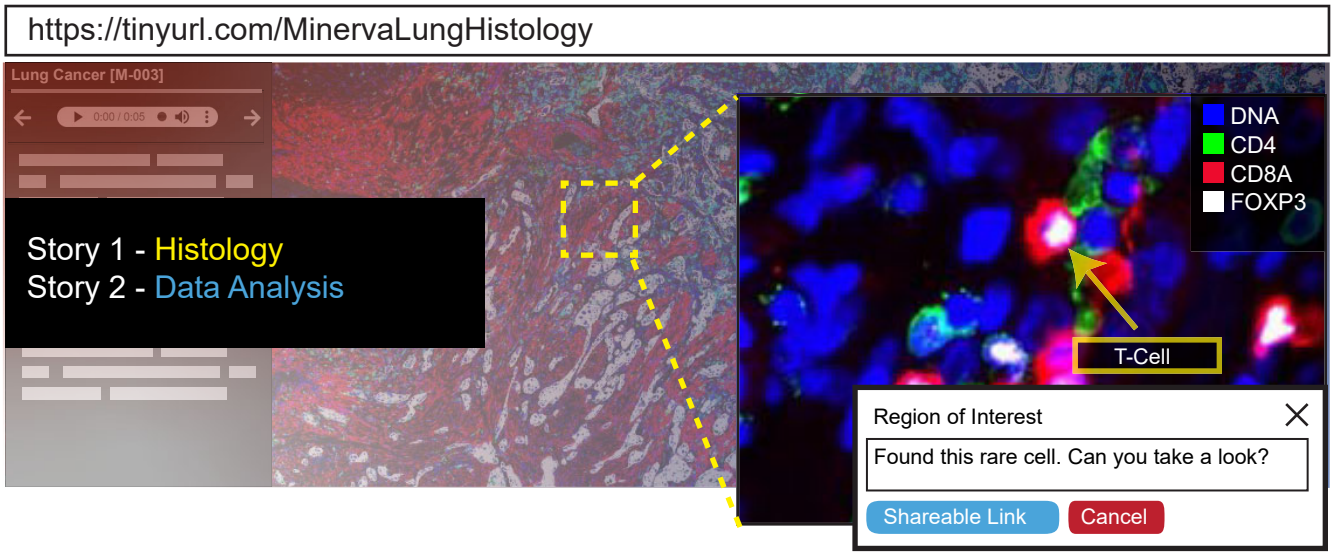


Image Settings
Waypoints
Narration
Minerva Author
Rendered Image Pyramids
Cloud Storage (AWS)
Minerva Story
Whole-slide Imaging
OME-TIFF
.JS/JPEG
.JS/JPEG

Figure 4

a.


https://tinyurl.com/MinervaLungHistology
Lung Cancer [M-003]
0:00 / 0:05
Story 1 - Histology
Story 2 - Data Analysis
DNA
CD4
CD8A
FOXP3
T-Cell
Region of Interest
Found this rare cell. Can you take a look?
Shareable Link
Cancel

b.

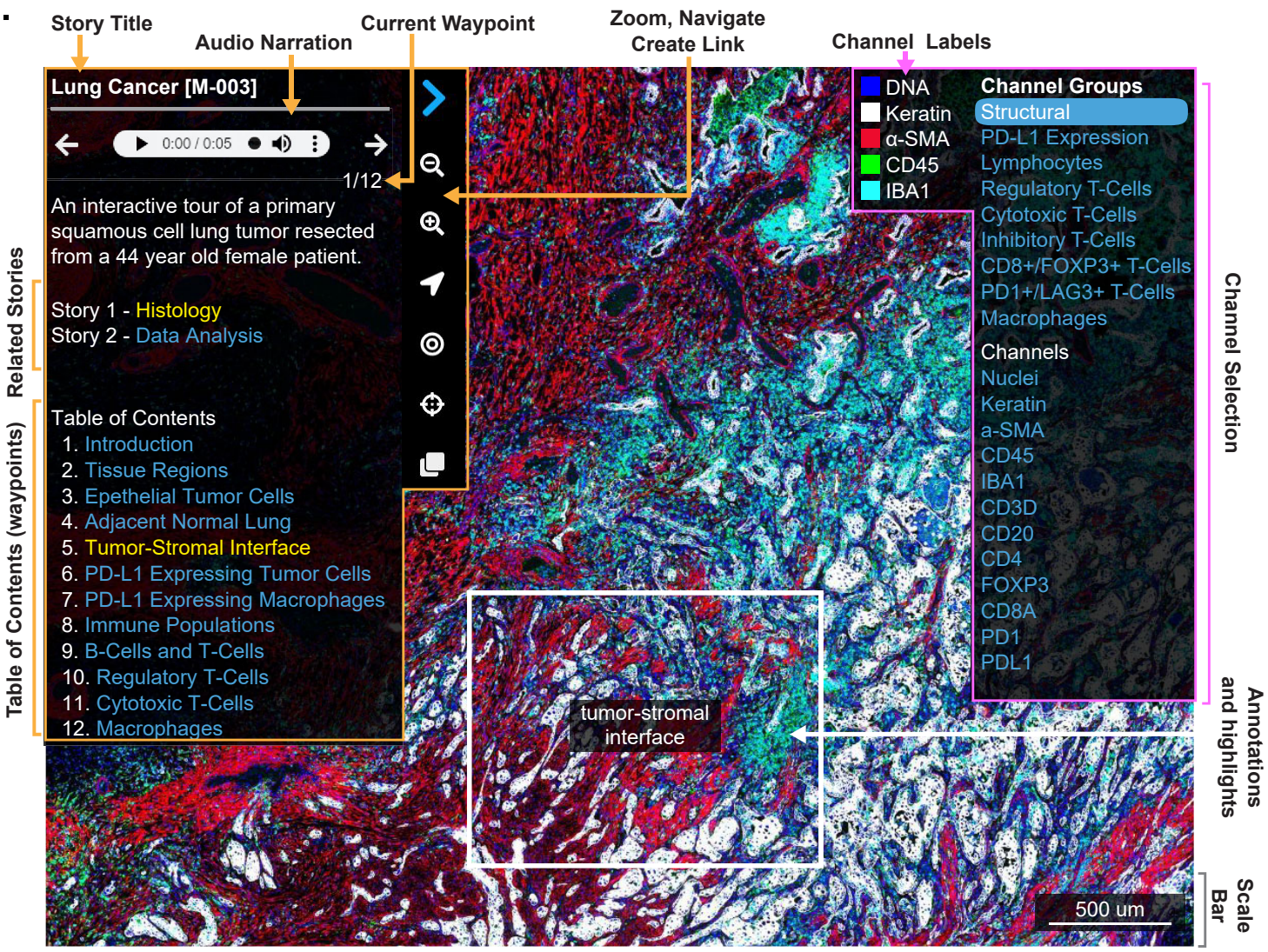
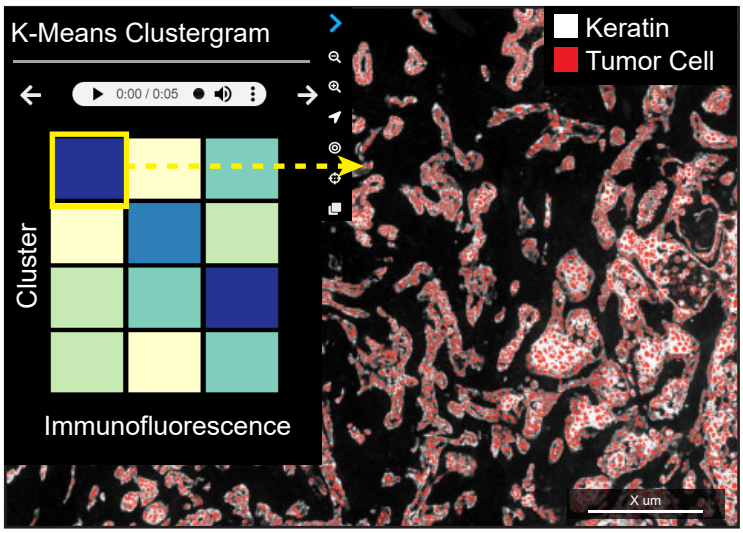
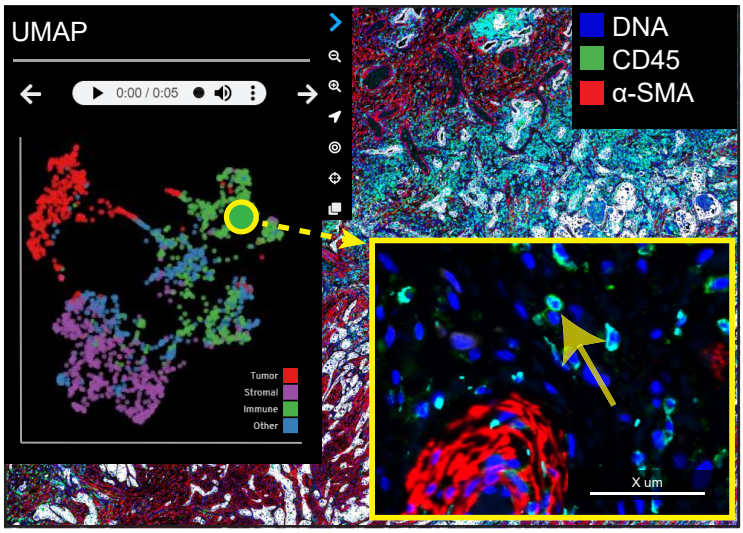

Story Title
Audio Narration
Current Waypoint
Zoom, Navigate
Create Link
Channel Labels
Lung Cancer [M-003]
0:00 / 0:05
1/12
An interactive tour of a primary squamous cell lung tumor resected from a 44 year old female patient.
Story 1 - Histology
Story 2 - Data Analysis
Related Stories
Table of Contents
1. Introduction
2. Tissue Regions
3. Epethelial Tumor Cells
4. Adjacent Normal Lung
5. Tumor-Stromal Interface
6. PD-L1 Expressing Tumor Cells
7. PD-L1 Expressing Macrophages
8. Immune Populations
9. B-Cells and T-Cells
10. Regulatory T-Cells
11. Cytotoxic T-Cells
12. Macrophages
Table of Contents (waypoints)
DNA
Keratin
α-SMA
CD45
IBA1
Channel Groups
Structural
PD-L1 Expression
Lymphocytes
Regulatory T-Cells
Cytotoxic T-Cells
Inhibitory T-Cells
CD8+/FOXP3+ T-Cells
PD1+/LAG3+ T-Cells
Macrophages
Channels
Nuclei
Keratin
a-SMA
CD45
IBA1
CD3D
CD20
CD4
FOXP3
CD8A
PD1
PDL1
Channel Selection
tumor-stromal interface
Annotations and highlights
500 um
Scale Bar

Figure 5

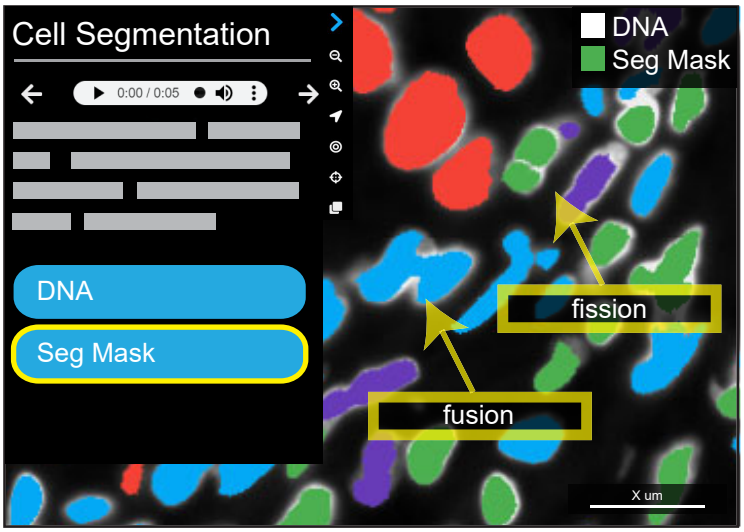
a. Story 2, waypoint 4



b. Story 2, waypoint 6

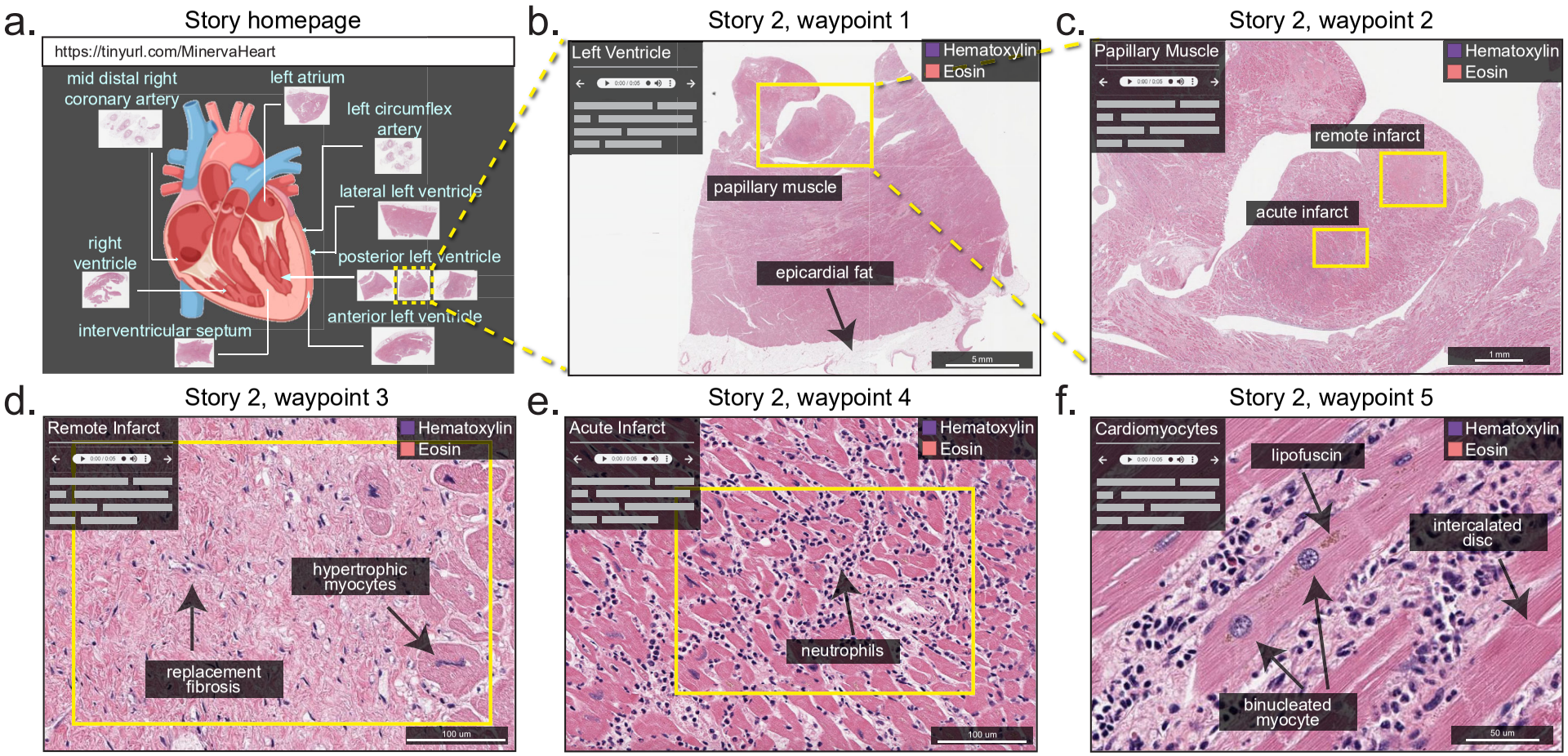


c. Story 2, waypoint 3



https://tinyurl.com/MinervaLungData

Figure 6


a.
Story homepage
https://tinyurl.com/MinervaHeart
mid distal right coronary artery
left atrium
left circumflex artery
lateral left ventricle
right ventricle
posterior left ventricle
anterior left ventricle
interventricular septum
b.
Story 2, waypoint 1
Left Ventricle
Hematoxylin
Eosin
papillary muscle
epicardial fat
5 mm
c.
Story 2, waypoint 2
Papillary Muscle
Hematoxylin
Eosin
remote infarct
acute infarct
1 mm
d.
Story 2, waypoint 3
Remote Infarct
Hematoxylin
Eosin
hypertrophic myocytes
replacement fibrosis
100 um
e.
Story 2, waypoint 4
Acute Infarct
Hematoxylin
Eosin
neutrophils
100 um
f.
Story 2, waypoint 5
Cardiomyocytes
Hematoxylin
Eosin
lipofuscin
intercalated disc
binucleated myocyte
50 um