Edinburgh Research Explorer

The Comparative Pathology Workbench: Interactive Visual Analytics for Biomedical Data

Journal Pre-proof

The Comparative Pathology Workbench: Interactive visual analytics for biomedical data

Michael N. Wicks, Michael Glinka, Bill Hill, Derek Houghton, Mehran Sharghi, Ingrid Ferreira, David Adams, Shahida Din, Irene Papatheodorou, Kathryn Kirkwood, Michael Cheeseman, Albert Burger, Richard A. Baldock, Mark J. Arends

Please cite this article as: M.N. Wicks, M. Glinka, B. Hill, et al., The Comparative Pathology Workbench: Interactive visual analytics for biomedical data, (2023), https://doi.org/10.1016/j.jpi.2023.100328

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

The Comparative Pathology Workbench: Interactive Visual Analytics for Biomedical Data

Michael N Wicks[1], Michael Glinka[1], Bill Hill[2], Derek Houghton[2], Mehran Sharghi[2], Ingrid Ferreira[1,3], David Adams[3], Shahida Din[4], Irene Papatheodorou[5], Kathryn Kirkwood[6], Michael Cheeseman[1], Albert Burger[2], Richard A Baldock[1], and Mark J Arends[1].

1) Edinburgh Pathology & Centre for Comparative Pathology, Institute of Genetics & Cancer, University of Edinburgh, Crewe Road, Edinburgh, EH4 2XR, UK.
2) Department of Computer Science, School of Mathematical and Computer Sciences, Heriot-Watt University, Edinburgh, UK;
3) Experimental Cancer Genetics, Wellcome Sanger Institute, Hinxton, Cambridge, UK;
4) Edinburgh IBD Unit Western General Hospital, NHS Lothian, Edinburgh, UK;
5) European Molecular Biology Laboratory - European Bioinformatics Institute (EMBL-EBI), Hinxton, Cambridge, UK;
6) Pathology Department, Western General Hospital, NHS Lothian, Edinburgh, UK.

Abstract

Pathologists need to compare histopathological images of normal and diseased tissues between different samples, cases and species. We have designed an interactive system, termed Comparative Pathology Workbench (CPW), that allows direct and dynamic comparison of images at a variety of magnifications, selected regions of interest, as well as the results of image analysis or other data analyses such as scRNA-seq. This allows pathologists to indicate key diagnostic features, with a mechanism to allow discussion threads amongst expert groups of pathologists and other disciplines. The data and associated discussions can be accessed online from anywhere in the world. The Comparative Pathology Workbench (CPW) is a web-browser based visual analytics platform providing shared access to an interactive "spreadsheet" style presentation of image and associated analysis data. The CPW provides a grid layout of rows and columns so that images that correspond to matching data can be organised in the form of an image-enabled "spreadsheet". An individual workbench can be shared with other users with read-only or full edit access as required. In addition, each workbench element or the whole bench itself has an associated discussion thread to allow collaborative analysis and consensual interpretation of the data.

The CPW is a Django-based web-application that hosts the workbench data, manages users and user-preferences. All image data are hosted by other resource applications such as OMERO or the Digital Slide Archive. Further resources can be added as required. The discussion threads are managed using WordPress and include additional graphical and image data. The CPW has been developed to allow integration of image analysis outputs from

systems such as QuPath or ImageJ. All software is open-source and available from a GitHub repository.

Keywords: Image visualisation, shared workspace, image spreadsheet, visual comparison, visual analytics, embedded discussion

Abbreviations

API	Application Programming Interface
CPW	Comparative Pathology Workbench [1]
DERMATLAS	The Dermatlas Project [2]
EBI	European Bioinformatics Institute [3]
IDR	Image Data Resource [4]
JSON	Javascript Object Notation
OMERO	Open Microscopy Environment Related Objects [5]
RESTful	Representational State Transfer API
SCEA	Single Cell Expression Atlas [6]
scRNA-seq	Single Cell RNA-sequencing
tSNE	t-distributed stochastic neighbour embedding
UMAP	Uniform Manifold Approximation and Projection for Dimension Reduction

1 Introduction

Pathologists need to be able to compare histopathological images of diseased organs and tissues, for inflammatory and immune disorders, tumours and many other conditions. Such comparisons may involve large numbers of cases of a specific disease or tumours, to permit comparison of the variation of appearance within human populations. Comparisons are also made between the same or highly similar disease processes in difference species, for comparative pathology studies. Currently, pathologists view different cases serially by examining the glass slide sections using a microscope and either retain the diagnostic features in memory while comparing a large series of cases, or take a series of photomicrographs of cases at fixed magnifications and compare these static images. Both of these approaches are limiting, often unsatisfactory and not conducive to easy sharing and discussion of observations. To aid such comparisons in a more dynamic and interactive way, we devised a computer-based approach allowing the pathologist to upload digital images of the scanned histopathological data to be compared within a computerised grid of images, with the images arranged in rows and columns in a way that is meaningful for the study (see Figure 1). This arrangement allows the user to directly compare images of different cases or different lesions in a variety of subjects or species, matching appropriate magnifications and selected fields of view as required for the study. Requirements gathered from a range of pathologists indicated that this new comparative system should allow regions of interest (ROI) to be easily marked for further discussion between users. Such mechanisms for discussions coupled with remote access allows for convenient discussion by a group of pathologists at different times or in different places.

Figure 1. A bench in the Comparative Pathology Workbench system.
This shows a typical bench within the CPW system with the tabular arrangement of rows and columns of cells, and cells populated with images from various sources including graphical presentation of image analysis results.

Developing and using a visual layout to support the analysis of large-scale data by domain experts has been long researched in the field of *visual analytics*[7]. Visual analytics emerged from the earlier studies of "sensemaking" [8,9], which is described as a process of initial data capture or "foraging" followed by an expert-driven process of "schema" development to find an optimal representation of the data for the required analysis. Visual analytics was identified as a *Grand Challenge*[10] to enable and accelerate the analysis of visually presented data (images, spatial and temporal visualisations, graphical layouts, etc.) coupled with real-time updating and high-performance computing. Most of the research effort responding to this challenge has focussed on large-scale and diverse data-types that require novel visualisation techniques and high-performance computing with a number of tools providing "dashboard" style views including automatic updating. Design patterns for visual analytics have been described and systems have emerged using aspects of these patterns for layout and processes, including discussion[7,11].

There are a number of commercial applications (e.g. 'Tableau: Business Intelligence and Analytics Software') that incorporate these ideas and various open-access tools such as Google-Sheets, but these are not domain-expert led and tuned to a specific application. In contrast, the underlying design of the Comparative Pathology Workbench (CPW) has evolved from the sketching of ideas by *domain experts* for an image analytics system useful to pathology research with a particular interest in comparative (cross-species) pathology. The underlying design pattern is a presentation of rows and columns as in a database table or a classical "spreadsheet", which we term a *workbench* (or *bench* for short). The design criteria include controlled shared access across the internet, simple web-browser-based application

with a threaded shared discussion for all data levels including image annotation and markup. The image and genomic/transcriptomic data of interest for this application already have open-access tools for archiving and sophisticated visualisation, so the requirement for this interface was to incorporate these within the workbench in a simple fashion linking out to the full capabilities as required for more detailed visualisation and analysis

The visual analytics design patterns include a standard row-column schema[7], but none of the tools previously available provided the required combination of simplicity with internet wide sharing, annotation and discussion threads. A spreadsheet design based on Unix tools and systems was developed in 1994 to provide a grid interface for image manipulation[13], but this does not appear to have been adopted and adapted to a collaborative working environment. In the next section we describe the primary purpose and design criteria for the CPW, and the following sections present results with real world practical applications, and a discussion of the application capabilities.

2 Material and Methods

2.1 Comparative Pathology Workbench Requirements

The practice of clinical and research pathology includes a collaborative process between experts to enable discussions leading to a deeper understanding of the pathological processes involved. For comparative research this is often a collaboration between internationally distributed groups, making meetings difficult in terms of physical distance and different time-zones. The requirement is therefore a simple interface to organise and view pathology image and other data within an environment that allows easy sharing together with discussion or commentary. It is important that the system is accessible for collaboration on any computer system without the need for special application installation.

In addition to the collaborative working environment, the pathologist-led design process enumerated a number of specific requirements:

1. A simple structured interface that allows comparison of images and analysis charts in 2 dimensions of rows and columns (e.g. image type against experimental sample, or *in situ* hybridisation gene expression stain image against anatomical location).
2. Options to display any visual data (e.g. results of genomic or transcriptomic sequence analysis, or the quantitative image analysis output from a machine learning analysis of histopathology images).
3. Maximum use and re-use of existing systems for image data management and data repositories, such as OMERO[5], IDR[4] or similar image servers.
4. Integration of images under investigation with the visual outputs from analysis tools (e.g. quantitative histopathology analysis using ImageJ, QuPath[14] or similar software, or transcriptomic analysis with cell-type mapping tools).
5. Lightweight web-based application allowing the user easy and internet-wide access.
6. Sharing model allowing either fully-public, or controlled sharing or fully private, as determined by the user.
7. Collaborative tools for capturing discussion threads at each level within the system (e.g. commentary or discussion on individual images, or on images with image annotations or regions of interest, or on a set of images, or at the level of the whole workbench collection of images).

2.2   Implementation

With these considerations we have developed a 2-D grid style interface analogous to the "spreadsheet" model for numerical data, but specifically for image and other visual data outputs (see Figure 1). In the first instance we have focussed on providing integration with three key resources. The first is the image management and archiving system *OMERO*[5], which has emerged as the lead system for very large-scale repositories of biomedical image data, and is an open-source system for researchers and institutions to set up their own repositories. In particular, OMERO provides the archiving server for the Image Data Resource[15] currently holding about 335 TB of publicly accessible image data. The second is the open-source histopathology image-analysis system *QuPath*[14] which delivers very-high quality manual and automated analysis of pathology images, including a range of machine-learning options for training and applying image classifiers for detailed numerical analyses of tissues and cell-types, across the very large images now generated by automated slide scanning. The third is the *Single Cell Expression Atlas (SCEA)*[16] at the EMBL European Bioinformatics Institute (EBI) which is an open archive of single cell RNA sequencing (scRNA-seq) data of direct importance to the research for which CPW has been developed, and represents the type of data that is important to be visualised alongside histological image data.

Although OMERO, QuPath and SCEA are the initial focus, the CPW system has been designed to allow any open resource with a suitable web service interface to be incorporated. For example, CPW can operate with the Digital Slide Archive[17] and the CZI CellXGene system[18] which are adopted for archiving data within the NIH-funded HuBMAP programme for human single cell analysis[19].

The CPW has been developed using open-access tools to ensure license-free distribution with the front-end visualisation delivered via a web application written in Python, using the Django web development framework, and incorporating HTML, CSS and JavaScript to provide the image display, user-interface components and a RESTful API, as well as user and permissions management. The backend SQL database is implemented using Postgres (version: 11) and the discussion threads and management provided use a WordPress instance[20]. Figure 2 shows an architectural overview of the CPW software with links through to external services such as OMERO.

Figure 2. Architectural overview of the CPW software.
This overview represents the data flow and exchange in CPW. Minimal data is stored in the CPW, instead only references to supported data and metadata are stored along with the arrangement of the bench as determined by the user. The CPW uses a small PostgreSQL database (DB) to maintain the state of all the benches and collections. It has a comprehensive REST API which allows for import and export of benches and collections of images. An integration with the metadata held in EBI's Single Cell Expression Atlas is being developed. This allows for the inclusion of t-SNE or UMAP plots into bench cells. QuPath analysis results can be stored within OMERO, to allow the referencing of these within the CPW. The clouds represent software systems. Normal arrows represent direct connection and ability to manipulate the data within the systems. Dashed arrows represent links or references to other systems.

3   Results

We have implemented the CPW using the "spreadsheet" grid structure to allow the user to arrange data in a tabular format. Each "spreadsheet" instance created by the user is termed a "workbench" or "bench" and each "cell" within a bench can store a URL reference to an image held in OMERO archives or any web-addressable image source. Benches may be shared with other CPW users, for viewing or editing, with a facility provided to allow analysis and discussion between collaborators. In addition, collections of images and other data links can also be created, managed, maintained and shared with other CPW users.

Minimal original data is stored in the CPW, but derived data from particular analyses and the associated metadata can be stored along with data sharing information and provenance. The CPW uses a Postgres database to maintain the state and associated sharing permissions of all the benches and data collections.

In the context of the CPW we have defined a set of elements:

- **Cell:** an individual component of a bench that can display an image or analysis results.
- **Workbench/Bench:** a user configurable tabular array of cells with the addition of column and row headers that provides an interface for cell editing, copying and moving using "drag and drop".
- **Source:** database archives of image and other data which provide a Uniform Resource Locator (URL) reference to the data.

- **Image:** in the CPW context any image is linked by an URL plus associated metadata. A thumbnail can be stored locally.
- **Collection:** a named set of images and other URLs which can be shared with other users.
- **Link:** images and analysis results can be formally linked which can include linking metadata.
- **Commentary:** a discussion thread associated with a cell, row or column within a bench or the whole bench.

These bench components are described in more detail below followed by descriptions of specific applications of the CPW as a series of exemplars.

3.1   Bench Elements and functions

3.1.1   Cells

A cell is the basic component of a bench and displays the image reference or analysis content that the user intends for comparison within the bench. The cell holds the link to the original visualisation of that data using the interface provided by the data resource or archive. For example, a thumbnail image of a whole slide high-resolution image of histopathology will link when selected, to the original image displayed in say, the OMERO image viewer. Each cell holds, in addition to the URL link, a user-defined title, description and discussion thread link. The cell menu provides a details page to allow editing of all cell content.

3.1.2   Benches

A bench is a grid of cells, which can each display a thumbnail of the referenced content (Figure 3). On bench creation, the user specifies the initial number of rows and columns (up to a maximum of 10 each, with the possibility of adding further rows and columns later), the display dimensions of the cells in the grid and a user-defined bench title and description and each new bench has a unique identifier. The row and column headers can also be given titles and descriptions, as the user requires (Figure 3: A and C). The number of rows and columns can be changed either by adding or removing individual rows and columns via the column or row menu functions, or by dragging and dropping ordinary cells onto the grey column or row footer cells (Figure 3: E and F).

The Bench page consists of two panels: the currently selected image collection is displayed to the left of the page (Figure 3: G); and the bench cell-array to the right (Figure 3: B). The image collection provides a "drag and drop" mechanism to populate any give cell with an image.

There are four types of cells (Figure 3) that make up a bench array: *master cell*, *row and column header cells*, *row and column footer cells*, and *ordinary cells*. At the top-left there is a "master" cell (Figure 3: B) providing a menu to the main control functions for the bench, including editing the bench detail, bench discussion thread and bench dimensions. Row and column header cells (Figure 3: A and C) provide row and column level functions including user-defined labels, title and description. Row and Column Footer cells are simply place holder cells, that provide "targets" for the dropping of dragged ordinary cells, resulting in adding extra columns or rows to the bench (Figure 3: E and F).

An ordinary cell in a bench has a menu offering the functions: amend cell; clear cell; view & edit commentary (Figure 3: D). The ordinary cell is analogous to the data-cells in a numerical spreadsheet and is the target for displaying selected images or data visualisations.

Figure 3. Structure of a cell and bench.
The screenshot shows an example bench with all the functions highlighted: (A) Column Header Cell – provides functions to add title to the column for better organisation, move the column around, etc.; (B) Master Cell – provides options to Edit the bench and the commentary associated with it; (C) Row Header Cell – similar to Column Header Cell, provides functions to add title to the whole row for better organisation, move the row around, etc.; (D) Ordinary Cell – this is the cell which gets populated with images or graphs from different sources; (E) Row Footer Cell – a placeholder cells, provides ability to expand the bench by adding more columns; (F) Column Footer Cell – a placeholder cell, provides ability to expand the bench by adding more rows; (G) Collection Cell – cell with the option to pick specific collections to add them to the bench.

3.1.3 Sources

In the current version there are 4 types of data source that the CPW can handle:

1. OMERO image archive
2. WordPress Media Libraries
3. The EBI Single Cell Expression Atlas (SCEA)
4. Web-based images – a URL plus user-supplied metadata

OMERO image servers[5] provide a JSON interface that allows the resources that they manage to be exposed, via unique URLs. The CPW uses this interface to provide a set of screens that enable navigation through the image archive and provides an interface to drill down through the available *groups*, *projects* and *datasets* stored on an OMERO server. If the datasets are not open to the public, then the user will be prompted to login to the server to allow access. The images within a dataset are then displayed with any defined regions of interest (ROIs) for further processing within the CPW.

The SCEA is a freely available community archive for single-cell RNA sequencing data across all species. Users can query and browse the data and visualise specific analyses such as "t-SNE" and "UMAP" charts derived from selected experiments. Once a specific analysis has been identified the user can "cut and paste" the URL to the CPW, download the generated image of the chart, and upload these to the CPW. These images are then stored locally within the CPW itself for display within a cell as required.

A WordPress[20] server is included to provide a discussion facility and also as a supplementary mechanism to store miscellaneous images for use associated with a bench. In addition to the discussion threads shown by the CPW, the WordPress server allows an extended "blog" style interface than can be made public for additional input from invited experts or from the community.

3.1.4 Images

The CPW does not store any large-scale histopathological images locally within its system; instead, it stores URL references to images that are held on other systems. Images showing analysis outputs often as charts or from web references may be stored locally within the CPW for convenience and to ensure robust access to avoid re-processing. The URL references are to images held on the sources discussed earlier and examples are shown in Figure 4. In addition to the URL the CPW stores metadata, associated sharing permissions and a user-input name.

Figure 4. Image Collection.

A Collection of images stored in the CPW, showing the image thumbnail, title, originating server, which collections hold this image, and which Benches display this image; these example images are held on an OMERO server referenced by the CPW.

The CPW web interface is designed to be viewed in any browser running on any standard computing platform. There are no size limits for images when viewed via the CPW, as the CPW system itself does not store any images, instead storing only references to images stored on another platform, in this case OMERO. Therefore, the image size limitation is determined by the OMERO system which has no theoretical limit – we have tested whole slide images up to 5GB, but a more typical WSI size is around 1GB. The CPW server requires a small hardware footprint (e.g. 8GB RAM, 80GB hard disc), and this was tested on MacOS (11.7.8), and Linux (Centos 7).

3.1.5 Collections

A *collection* is a set of images that the CPW user has gathered into a single, named group. When browsing the image sources, the user is given the option to add an image to a collection (see Figure 4). The images in a collection can then be used by the user to populate individual cells within a bench. Any collection can be shared with editing or viewing privileges with any other CPW user and can be visualised in a browsing interface to check details, metadata and the image usage within benches and other collections. A collection can be selected for visualisation as a compact list adjacent to a bench to allow images to be copied into individual cells. These interfaces are shown in Figure 5; collections may consist of images from many different sources.

Figure 5. Choosing a Collection.

As shown previously in Figure 3: G, a collection can be chosen and the images from the collection moved by Drag & Drop into other cells to occupy them. In fact, the bench can be populated with images and graphs from multiple different collections that can be chosen from the drop-down list, as long as the user has the privileges to view those collections.

3.1.6 Links

In the CPW, images can be "linked" together, to indicate simple relationships between images. For example, if an image has been generated from another image, using an analysis tool such as QuPath, it is useful to describe such a relationship. The CPW can record simple relationships between pairs of images, along with a comment or a description of the reason for the link. Supporting evidence, for example a detailed description of a segmentation process or a processing script, is provided in a zip file, which must be uploaded at the time the link is created. Whenever an image is displayed in the CPW that has linked images associated with it, a "chain" icon is displayed. Clicking on the chain icon displays the link details, which includes the images linked together, the comment on the link and the supporting evidence file.

For example, the integration of QuPath to the CPW is indirect, via the linking of 2 images. Images can be created in QuPath and both these derived images and original images from QuPath can be independently uploaded to the CPW and OMERO system and the link between the original and derived images can be defined. This link must have suitable comments and may have a zip file of QuPath outputs (scripts, annotations, masks or text) attached to the link as well. It is possible to upload single-cell masks and other QuPath or ImageJ masks (or from any other image processing system) derived from image segmentation protocols.

3.1.7   Sharing and Discussion

In addition to these key CPW components we have built a sharing model to provide users with the capability to share benches and resources with other users of the workbench. Any resource (bench, image or collection) can be fully private, shared read-only, or shared with full editing permission, although only the owner of a resource can edit the permission status. This sharing within the workbench enables direct collaboration within the set of users registered with a particular instantiation of the CPW. To allow sharing between multiple workbench servers we have developed and implemented a mechanism to "dump" and "reload" any given workbench, via the RESTful JSON API (see below). This facility is also useful for saving a workbench, for example, as an analysis snapshot associated with a publication, as well as for secure backup.

For visual analysis of histopathology images, the research process includes discussion between domain experts, as shown in Figure 6. For this we have used WordPress to provide a web-interface to capture the discussion thread or commentary which is analogous to "blog" posts and responses. The WordPress interface can also be used to extend the thread with images and other media for wider sharing. This discussion also serves as a backed-up record of the discussion and decision-making process in the analysis of individual images or sets of images. The discussion can be viewed in the aggregated comments page where both Bench and Cell comments can be seen and addressed.

Figure 6. Bench Commentary.
A screenshot visualising the possible commentary for a bench.

3.1.8   Search

There are 3 search facilities provided within the CPW: a simple search facility provided in the Header Bar on every page in the system; a Bench search facility on the List Benches page; and a Collection Search facility on the List Collections page.

Simple Search is a single search box in the System Header Bar, returning a set of Benches whose Title or Description contain this string (Figure 7: A).

Bench Search allows the user to provide up to 8 different search criteria to find subsets of benches within the system, including the Title, Description, Owner, Authority, Prior to a Creation Date, Post a Creation Date, Prior to a last Modified Date, and Post a Last Modified Date. One, some or all of these criteria may be supplied by the searcher to narrow the set of returned benches (Figure 7: B).

Collection Search allows the user to provide up to 4 different search criteria to find subsets of collections within the system, including the Title, Description, Owner, and Authority. One, some or all of these criteria may be supplied to narrow the set of returned collections (Figure 7: C).

Figure 7. Search Functions.
A screenshot visualising the search functions: (A) the Simple Search, (B) the Bench Search with the 8 different search criteria, (C) the 4 Collections Search options, and (D) the 7 Image Search options.

3.1.9 RESTful JSON API

The CPW has a comprehensive REST API in addition to the usual Web HTML interface. Benches, Collections and Images can be manipulated using this API, using the HTTP commands, GET (read), POST (create), PUT (update) and DELETE; PATCH (partial update) and GET to list benches is not supported. This API allows benches, collections and images to be represented in JSON format (Javascript Object Notation); and this format is used to retrieve, create and update benches, collections and images. The prime motivation behind this interface is to provide a "Dump and Restore" mechanism between instances of the Workbench; and to allow users to copy an existing bench for the purposes of sharing or copying to a different system, should this be required.

3.2   Bench Exemplars

Here we illustrate usage of the CPW via three examples of "real-world" applications. Two examples are part of ongoing research programmes and one example is clinically based.

3.2.1   DERMATLAS Expert Diagnosis Review

The CPW workbench has been a highly valuable tool for the DERMATLAS project (http://www.dermatlasproject.org/), which aims to collect 70 types of uncommon and rare skin tumours (including sebaceous neoplasms, sweat gland tumours, other skin appendage tumours, vascular tumours, etc) collected from a number of different countries from around the world (UK, Belgium, France, Canada, USA, Mexico) with collection of around 50 cases for each tumour type (over a thousand scanned slide images have been uploaded to date) (Figure 8). The CPW has been used to build a series of *workbenches* in order to extensively review this wide range of skin tumours in a practical and pathologist user-friendly way, that is easy to use for any histopathologist.

To be consistent, a workbench has been allocated to each skin tumour type. The grid of rows and columns allowed the user to design the bench in the most efficient and convenient manner for the user (in our case, each column corresponds to a collaborator or source of tumours), as well as direct comparison of skin tumours previously given the same or closely related histopathological diagnoses. As the CPW is accessible from anywhere in the world and from any type of internet-linked computer, all skin tumour diagnoses were reviewed by a panel of 2-4 expert dermatopathologists based in Belgium, UK and Canada, and specific discussions were possible due to the use of the commentary and annotation functions, from which several diagnoses were modified or changed to new diagnoses. Indeed, using the OMERO viewer, the CPW allows the user to identify and annotate regions of interest (ROI) for discussion, including making some comments on the scanned slide images in the bench. In addition, OMERO supports a large number of scanned slide formats. Scanned images of immunohistochemistry (IHC) stains were imported to the *workbench* for comparative analysis and discussion by a panel of expert consultant pathology diagnosticians with appropriate expertise in skin tumour IHC interpretation, sometimes as a key component of making the primary diagnosis and at other times IHC was used as a critical molecular pathological test, such as use of mismatch repair IHC for identification of defective mismatch repair in sebaceous skin neoplasms[21,22,23]. Discussions between expert dermatopathologists as part of the DERMATLAS project took place to ensure consistency of histopathological diagnosis and agreement on small regions of tissue identified histologically for targeted high-quality sampling of FFPE (formalin-fixed paraffin-embedded) tumour material and adjacent normal skin FFPE samples for DNA and RNA extraction for comparative tumour genome and transcriptome sequencing studies. This approach provided unique biological and histopathological insights into comparative analyses of the component parts and of the overall diagnoses of a range of uncommon and rare skin tumours, including analysis and expert discussion of the critical diagnostic criteria for diagnosis, molecular pathological testing and for tumour genomic / transcriptomic analysis[23–25].

Figure 8. DERMATLAS comparative histopathological analysis workbenches.

(A) A typical DERMATLAS workbench for reviewing the diagnoses of a specific skin tumour type. The column headers represent collaborators ("Path1", "Path2", etc). Within each column, each cell contains a different skin tumour case (eg. "PD42477a" = anonymised case number, etc). This approach allows all tumours from the same tumour subtype to be compared across the different collaborators for easy review, comparison of key diagnostic features and discussion of challenging or discrepant cases. (B) A different DERMATLAS workbench for reviewing immunohistochemistry stains performed to aid diagnosis of a particular skin tumour subtype. In this workbench, the column headers represent the different stains used on the samples (eg. "HE" = Haematoxylin & Eosin, "MSH2" = immunohistochemical stain for the mismatch repair protein MSH2, etc), with the row headers indicating specific skin tumour anonymised case numbers (eg. "PD46980a" etc). This approach allows the individual cases to be easily reviewed, compared and discussed by different expert pathologists.

Crohn's Disease Fibrostenosing Lesion Comparative Analysis

The functionality of the CPW allowed for visualisation of the differences in the morphology of a set of 30 terminal ileal fibrostenosing lesions surgically resected from different Crohn's disease patients. In addition, a set of 30 normal terminal ileum controls were collected from non-Crohn's patients (Figures 5 and 9) collected in the context of the Human Gut Cell Atlas[26–28] and Human Cell Atlas programmes[19]. Standard histological H&E stained sections, scanned as whole slide images, were uploaded to the bench for these Crohn's fibrostenotic lesions.

Further sections were taken from the same tissue blocks for both histochemical analysis of the amount of collagenous fibrous tissue determined by Picrosirius Red staining and also for a set of immunohistochemical (IHC) stains for a variety of different cell types (B-cells, T-cells, macrophages, smooth muscle cells, etc) for comparison of the histopathological changes between these different cases of Crohn's Disease fibrostenosing lesions with normal ileum controls.

A shared discussion between researchers and pathologists allowed for identification of additional regions of interest which in turn generated a further focus on in-depth discussions and quantitative analyses of the amount of collagen using QuPath[14]. The additional regions of interest were then uploaded to OMERO and linked in the CPW (Figure 9: A and B). The quantified data from QuPath were then uploaded to CPW itself and a summary row was generated (Figure 9: C), which allowed for better visualisation of the differences between the cases. The functionality of the CPW has allowed images to be shared and provided a discussion platform for experts to gain further insights into the development, locations and consequences of excessive fibrosis in Crohn's Disease (Figure 9: D).

Figure 9. Crohn's Disease Fibrostenosing Lesion Comparative Analysis Workbench.
(A) Visualisation of the workbench displaying images of the histological sections. Rows represent specific samples; columns represent specific stains and/or data related to all samples. (B) Chain link icons representing links that were created between sections to visualise and notify the user that there is a relationship between such sections. (C) Red-dashed line delineated box represents accumulation of all of the data points quantified from annotations. These data represent the surface area size of each region of the ileum as well as the collagen area quantified using QuPath. The summary data can be found in the last row as shown in "Data Summary". (D) The CPW commentary function allowed for discussion about the workbench images and data to further increase understanding of the integrated morphological, fibrosis-histochemical and immunohistochemical stains with quantitative evaluation and analysis.

3.2.2 Coeliac Disease Duodenal Biopsy Histopathology Audit

A CPW workbench was set up to use as the platform for a histopathological audit study of a large series of scanned whole slide images of duodenal biopsies showing either normal appearances, or Coeliac disease diagnostic features, including total or subtotal villous atrophy, crypt hyperplasia and intraepithelial lymphocytosis. A group of 13-17 pathologists including expert gastrointestinal consultant histopathologists, were asked to use the CPW platform to view a series of scanned whole slide images and make a diagnosis of either normal appearances, Coeliac disease or indeterminate/uncertain conclusion[29,30]. This allowed an audit study to be rapidly and efficiently performed by this group of pathologists using remote access of the CPW, with the initial round of diagnoses made by the panel of pathologists without knowledge of any other pathologists' diagnoses or opinions. A small number of discrepant-diagnosis cases were identified that had produced significantly differing diagnoses or opinions amongst the pathologists and the CPW platform was used for review of the histopathological features of these cases by a smaller group of national experts with discussion of their features, using the CPW as a discursive

Figure 10. Visualisation of the Coeliac Disease Histopathology Audit. The infographic represents how the audit was performed. A number of pathologists simultaneously accessed the workbench from many different locations, edited it and exchanged their observations (solid arrows), allowing for a rapid exchange of diagnoses and opinions (dashed arrows).

and interactive platform in real-time, aimed at reaching a consensus diagnosis or exclusion of the case from the audit study (Figure 10).

4 Discussion

Having identified a clear need for pathologists to be able to directly compare histopathological images of diseases both between different cases and different species, we designed an interactive system, the CPW, that allows direct and dynamic comparison of scanned whole slide histopathological images at a variety of magnifications. This approach also allowed use of selected regions of interest, to permit pathologists to indicate and discuss key diagnostic features, with a discursive mechanism, allowing discussions amongst a group of expert and non-expert pathologists and other users, that can be accessed remotely from anywhere in the world. The CPW not only allows comparison of standard H&E stained sections as scanned histopathological images, but images of any other type that can be

uploaded to the workbench grid, including histochemical special stains (such as the collagen stain illustrated above), immunohistochemical stains, immunofluorescence stains, *in situ* RNA or DNA hybridisation stains, macroscopic pathological images, radiological images and images of molecular pathology investigations, such as single cell transcriptomics or genomics chart images, along with other graphs and charts relevant to molecular analysis of the case. As such, the CPW is very flexible in its use of a wide variety of images. Some of these functions have been illustrated here in the three use cases included.

Images can be arranged in rows and columns in a way that is meaningful to the user or group of users involved in analysing the grid of images. For example, multiple images from a single patient's disease process can be arranged in a column, allowing comparison of multiple patients across the rows in a grid. The user has the flexibility to link selected images as part of an analysis as well as keep notes about such linkage. The selection of regions of interest allows the pathologist or other user to focus on critical diagnostic features of a series of cases, along with writing a commentary about these selected features to promote discussion amongst a group of users invited to join analysis of a particular workbench.

The CPW platform can be accessed from most computers that can run the most frequently used internet browsers (Chrome, Firefox, Safari, etc). Remote access allows pathologists and other users to view the images within a bench at a time of their choosing from anywhere in the world and together with the simplicity of the grid arrangement and the ease of access (once invited with the appropriate weblinks) permits a group of busy expert pathologists to share their expertise in order to confirm diagnoses or to generate new insights following analysis of a range of similar cases in a workbench.

The CPW system also has the potential for use as a histology or histopathology teaching tool, by allowing the teacher or demonstrator to upload a set of histological or other images and directing the students to the appropriate weblink with instructions to view the images, identify particular features (marked by ROIs if necessary), invite students to indicate certain structures by the drawing of their own ROIs (with different coloured ROI lines for several features), provide descriptions, diagnoses, or answers to questions, using the commentary function. This system can also allow real-time online virtual interactions between the teacher and students even if they are not together in the same venue. This potential use as a teaching tool will be explored by us in the coming months, along with follow-up studies on the usability of the CPW by others with collection of responses in answer to questionnaires.

The CPW has some limitations, including the design feature that it works mostly with scanned microscopic whole slide images (stored on OMERO), although it can also display any image type, such as macroscopic pathology images (JPEG, TIFF or other image formats), amongst others. The CPW has a custom interface to the EBI Single Cell Expression Atlas (SCEA) in order to display SCEA chart outputs (UMAP plots or t-SNE plots), whereas other data sources would require custom interfaces to be constructed and can be developed as required. The CPW can sometimes be slow to render the web page for a large workbench that contains a large number of images (>100) from an OMERO server. We are currently implementing a software upgrade for thumbnail image caching that would speed up web page rendering. The CPW only usefully displays on a desktop or laptop machine and is not suitable with constrained devices with small screens.

The workbench interface has been implemented and currently is in use with a number of pathology research programmes[22,24,30] and an open test version is available from the Centre

for Comparative Pathology (CCP) website[31]. In addition, all software is open-source and available from a GitHub repository[32]. A detailed user-manual describing the detailed structure and functions of the interface has been written and is also available from both the GitHub repository and the CCP website.

5   Conclusions

The Comparative Pathology Workbench (CPW) is a tool designed and developed to enable more rapid and effective pathology analysis via a novel visual interface coupled with sharing and discussion capabilities for widely distributed collaborative teams[33]. The effectiveness has been demonstrated with two research-based exemplars from which further publications are emerging. In addition, a use-case for pathology auditing, in this instance for coeliac disease in duodenal biopsies has been presented to illustrate the effectiveness of the approach. The initial prototype for the CPW was used for student teaching and examination and we envisage significant benefit for this visual analytic design pattern not just in pathology practice, research and teaching but across many fields that use and compare related sets of images and especially if the analysis involves multiple collaborators contributing different expertise.

The user needs access to the internet to be able to use the CPW and in cases where the network bandwidth is limiting, the loading time for CPW might be prolonged due to the necessity of accessing the image sources.

Future development of the CPW will streamline the existing capabilities and allow interaction with other sources of data, in particular we will bring in the GCA and HCA adopted tools such as the Digital Slide Archive[17] and CELLxGENE[18] system and these will be available as an HCA "Analysis Portal" particularly in the context of the Human Gut Cell Atlas[26,28]. Other planned developments will be supporting clinical and research pathology in understanding disease progression and comparative analysis across different species for disease modelling.

6   References

1. Comparative Pathology Workbench. The University of Edinburgh. Published November 29, 2021. Accessed February 21, 2023. https://www.ed.ac.uk/comparative-pathology/the-gut-cell-atlas-project/detail/comparative-pathology-workbench

2. Dermatlas. Accessed February 21, 2023. https://www.dermatlasproject.org/

3. Institute EB. EMBL-EBI homepage. Accessed February 21, 2023. https://www.ebi.ac.uk/

4. IDR: Image Data Resource. Accessed February 21, 2023. https://idr.openmicroscopy.org/

5. Allan C, Burel JM, Moore J, et al. OMERO: flexible, model-driven data management for experimental biology. *Nat Methods*. 2012;9(3):245-253. doi:10.1038/nmeth.1896

6. Single Cell Expression Atlas. Accessed February 21, 2023. https://www.ebi.ac.uk/gxa/sc/home

7. Heer J, Agrawala M. Design Considerations for Collaborative Visual Analytics. *Inf Vis*. 2008;7(1):49-62. doi:10.1057/palgrave.ivs.9500167

8. Russell DM, Stefik MJ, Pirolli P, Card SK. The cost structure of sensemaking. In: *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems*. CHI '93. Association for Computing Machinery; 1993:269-276. doi:10.1145/169059.169209

9. Pirolli P, Card S. The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. In: ; 2005.

10. Thomas JJ, Cook KA. A visual analytics agenda. *IEEE Comput Graph Appl*. 2006;26(1):10-13. doi:10.1109/MCG.2006.5

11. Zhang L, Stoffel A, Behrisch M, et al. Visual analytics for the big data era — A comparative review of state-of-the-art commercial systems. In: ; 2012:173-182. doi:10.1109/VAST.2012.6400554

12. Tableau: Business intelligence and analytics software. Tableau. Accessed January 19, 2023. https://www.tableau.com/en-gb/node/62770

13. Levoy M. Spreadsheets for images. In: *Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques. SIGGRAPH '94*. ACM Press; 1994:139-146. doi:10.1145/192161.192190

14. Bankhead P, Loughrey MB, Fernández JA, et al. QuPath: Open source software for digital pathology image analysis. *Sci Rep*. 2017;7(1):16878. doi:10.1038/s41598-017-17204-5

15. Williams E, Moore J, Li SW, et al. Image Data Resource: a bioimage data integration and publication platform. *Nat Methods*. 2017;14(8):775-781. doi:10.1038/nmeth.4326

16. Moreno P, Fexova S, George N, et al. Expression Atlas update: gene and protein expression in multiple species. *Nucleic Acids Res*. 2022;50(D1):D129-D140. doi:10.1093/nar/gkab1030

17. Gutman DA, Khalilia M, Lee S, et al. The Digital Slide Archive: A Software Platform for Management, Integration, and Analysis of Histology for Cancer Research. *Cancer Res*. 2017;77(21):e75-e78. doi:10.1158/0008-5472.CAN-17-0629

18. Chan Zuckerberg CELLxGENE Discover. Cellxgene Data Portal. Published 2023. Accessed January 15, 2023. https://cellxgene.cziscience.com/

19. Snyder MP, Lin S, Posgai A, et al. The human body at cellular resolution: the NIH Human Biomolecular Atlas Program. *Nature*. 2019;574(7777):187-192. doi:10.1038/s41586-019-1629-x

20. WordPress.com. WordPress.com. Accessed January 19, 2023. https://wordpress.com/

21. Rashid M, van der Horst M, Mentzel T, et al. ALPK1 hotspot mutation as a driver of human spiradenoma and spiradenocarcinoma. *Nat Commun*. 2019;10(1):2213. doi:10.1038/s41467-019-09979-0

22. Ferreira I, Wiedemeyer K, Demetter P, Adams DJ, Arends MJ, Brenn T. Update on the pathology, genetics and somatic landscape of sebaceous tumours. *Histopathology*. 2020;76(5):640-649. doi:10.1111/his.14044

23. Ferreira I, Arends MJ, van der Weyden L, Adams DJ, Brenn T. Primary de-differentiated, trans-differentiated and undifferentiated melanomas: overview of the clinicopathological, immunohistochemical and molecular spectrum. *Histopathology*. 2022;80(1):135-149. doi:10.1111/his.14545

24. Ferreira I, Droop A, Edwards O, et al. The clinicopathologic spectrum and genomic landscape of de-/trans-differentiated melanoma. *Mod Pathol*. 2021;34(11):2009-2019. doi:10.1038/s41379-021-00857-z

25. Wong K, van der Weyden L, Schott CR, et al. Cross species genomic landscape comparison of human mucosal melanoma with canine oral and equine melanoma. *Nat Commun*. 2019;10(1):353. doi:10.1038/s41467-018-08081-1

26. Burger A, Baldock RA, Adams DJ, et al. Towards a clinically-based common coordinate framework for the human gut cell atlas: the gut models. *BMC Med Inform Decis Mak*. 2023;23(1):36. doi:10.1186/s12911-023-02111-9

27. Helmsley Gut Cell Atlas Roadmap. Helmsley Trust. Accessed February 21, 2023. https://helmsleytrust.org/wp-content/uploads/2023/03/Gut-Cell-Atlas-_-Broad-Roadmap.pdf

28. Zilbauer M, James K, Kaur M, et al. A roadmap for the human gut cell atlas. *Nat Rev Gastroenterol Hepatol*. 2023;Accepted Mar 2023.

29. Denholm J, Schreiber B, Jaeckle F, Wicks M, Benbow E, Bracey T, Yiu J, Chan H, Farkas L, Fryer E, Gopalakrishnan K, Hughes C, Kirkwood K, Langman G, Mahler–Araujo B, McMahon R, Win Myint KL, Natu S, Robinson A, Sanduka A, Sheppard K, Tsang Y-W, Arends MJ, Soilleux E. Digital pathology reporting of Coeliac Disease: an inter-observer agreement study. (Manuscript in preparation).

30. Denholm J, Schreiber BA, Evans SC, et al. Multiple-instance-learning-based detection of coeliac disease in histological whole-slide images. *J Pathol Inform*. 2022;13:100151. doi:10.1016/j.jpi.2022.100151

31. Centre for Comparative Pathology. The University of Edinburgh. Published June 4, 2020. Accessed February 21, 2023. https://www.ed.ac.uk/comparative-pathology

32. Comparative Pathology Workbench. Published online December 21, 2021. Accessed February 21, 2023. https://github.com/Comparative-Pathology/comparativepathologyworkbench

33. McInnes EF, Meyerholz DK, Arends MJ. Concerns about pathology expertise and data quality. *J Pathol*. 2023;259(4):468-468. doi:10.1002/path.6058

7   Acknowledgements

8   Funding

9   Availability of data and materials

All software developed for this project is freely available from the GitHub repository https://github.com/Comparative-Pathology. The exemplar benches and screen snapshots all use version 1.0.0 of the available software as an archive tar- or zip-file from https://github.com/Comparative-Pathology/comparativepathologyworkbench/releases/tag/Release-Candidate-1.0.0.

10 Ethics approval and consent to participate

Histopathological and radiological image data used for the comparative pathology workbenches was fully anonymised with accordance with the appropriate UK ethical and legal framework. Ethical approval for anonymised tissue section use was obtained from Lothian NRS Bioresource Research Tissue Bank.

11 Competing interest

The authors declare that they have no competing interests.

12 Author contributions

All – manuscript preparation and review; MNW - primary software development and implementation; MG, IF, SD, KK – application testing and review, bench applications; BH, DH, MS – software review and application testing; DA, IP - PI support and data sources; AB – PI support and Comp. Sci. input; MC – original design concepts; RAB, MJA – PI support, original design concepts, applications testing and bench development.

Declaration of interests

☒The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: