Neuroinformatics

ISSN 1539-2791/03/309–326/$25.00

Original Article

EMAP and EMAGE

A Framework for Understanding Spatially Organized Data

Richard A. Baldock,[*] Jonathan B. L. Bard,[1] Albert Burger, Nicolas Burton, Jeff Christiansen, Guanjie Feng, Bill Hill, Derek Houghton, Matthew Kaufman,[1] Jianguo Rao, James Sharpe, Allyson Ross, Peter Stevenson, Shanmugasundaram Venkataraman, Andrew Waterhouse, Yiya Yang, and Duncan R. Davidson

MRC Human Genetics Unit, Western General Hospital, Crewe Road, Edinburgh, UK; [1]Division of Biomedical Sciences, University of Edinburgh, Hugh Robson Building, George Square, Edinburgh, UK

Abstract

The Edinburgh Mouse Atlas Project (EMAP) is a time-series of mouse-embryo volumetric models. The models provide a context-free spatial framework onto which structural interpretations and experimental data can be mapped. This enables collation, comparison, and query of complex spatial patterns with respect to each other and with respect to known or hypothesized structure. The atlas also includes a time-dependent anatomical ontology and mapping between the ontology and the spatial models in the form of delineated anatomical regions or tissues. The models provide a natural, graphical context for browsing and visualizing complex data.

The Edinburgh Mouse Atlas Gene-Expression Database (EMAGE) is one of the first applications of the EMAP framework and provides a spatially mapped gene-expression database with associated tools for data mapping, submission, and query. In this article, we describe the underlying principles of the Atlas and the gene-expression database, and provide a practical introduction to the use of the EMAP and EMAGE tools, including use of new techniques for whole body gene-expression data capture and mapping.

Index Entries: Mouse development; developmental atlas; gene-expression database; ontology; 3D image viewer; Java interfaces.

[*]Address to which all correspondence and reprint requests should be sent. E-mail: Richard.Baldock@hgu.mrc.ac.uk

Introduction

The Edinburgh Mouse Atlas Project (EMAP) has implemented a spatio-temporal framework for capturing spatially organized and mapped data (Baldock et al., 1992; Davidson et al., 1997; Bard et al., 1998; Brune et al., 1999), The framework consists of a time-series of 3D models for most postimplantation developmental stages of the mouse. The staging is based on Theiler (1989) with supplementation from Downs and Davies (1993), which overall provides stages at roughly half-day intervals. In addition to the models, the framework includes a standard nomenclature or ontology of the anatomical tissues present at each stage. The nomenclature is linked to the models by delineation of the principal named anatomical components within the same coordinate frame.

This basic design of a volumetric, voxel-based space coupled to an ontology of conceptual terms via spatial mappings is the de facto standard for this type of database in biomedical research. The voxel coordinates represent the resolution of the given models but can be considered as grid-lines for higher-resolution models if required. A key property of such systems is a mechanism for spatial transformation or *mapping*. This can be a simple linear affine transform (scale, rotate, translate) or arbitrarily complex non-linear warps to bring the experimental data into alignment with the standard model. The ontology provides interoperability with other resources and allows inference in terms of the conceptual meaning of data; the spatial mapping provides spatial query and inference in geometric terms.

This work complements other tools and applications in this special issue. Specifically the embryo atlas discussed here is directly related in form to both the Mouse Brain Atlas of Rosen et al. (2003) and the Mouse Atlas Project reported by MacKenzie-Graham et al. (2003). These atlases are primarily of the adult mouse brain, but the data is volumetric and presented as standard section views or directly as voxel images. Martone et al. (2003) describe access to very high-resolution data down to subcellular structure and organization localized within a standard atlas.

The general message is clear: Atlases are a necessary mechanism for managing data and tools for manipulation, browsing, and querying by spatial localization. The challenge is to provide efficient interoperability across different atlases at the spatial level as well as at the conceptual level (ontologies).

The model, ontology, and domains provide a basic framework for mapping, analyzing, collating, and querying spatial data. In other words, it is the underlying basis for a database, analogous to a geographic information system (GIS) with the possibility of mapping arbitrarily complex spatial data. In contrast to many GIS systems, the underlying space is three-dimensional and changes with time. Furthermore, the data captured is always from a new embryo, which introduces significant difficulties for the mapping process. This framework, the tools and resources available to access and to map onto the framework, and an example of its use for a spatial gene-expression database (EMAGE), are described in the following pages.

The EMAP Spatio-Temporal Framework

The EMAP framework is comprised of three key components: 3D grey-level (voxel) models of the embryos at each stage, an anatomical ontology and a mapping between the spatial context of the digital models, and the textual context of the anatomy.

The first component is the set of EMAP model embryos, which are 3D reconstructions made from images of stained histological sections (Brune et al., 1999). These split into two groups: those digitized from the wax-embedded sections Professor Kaufman used for his

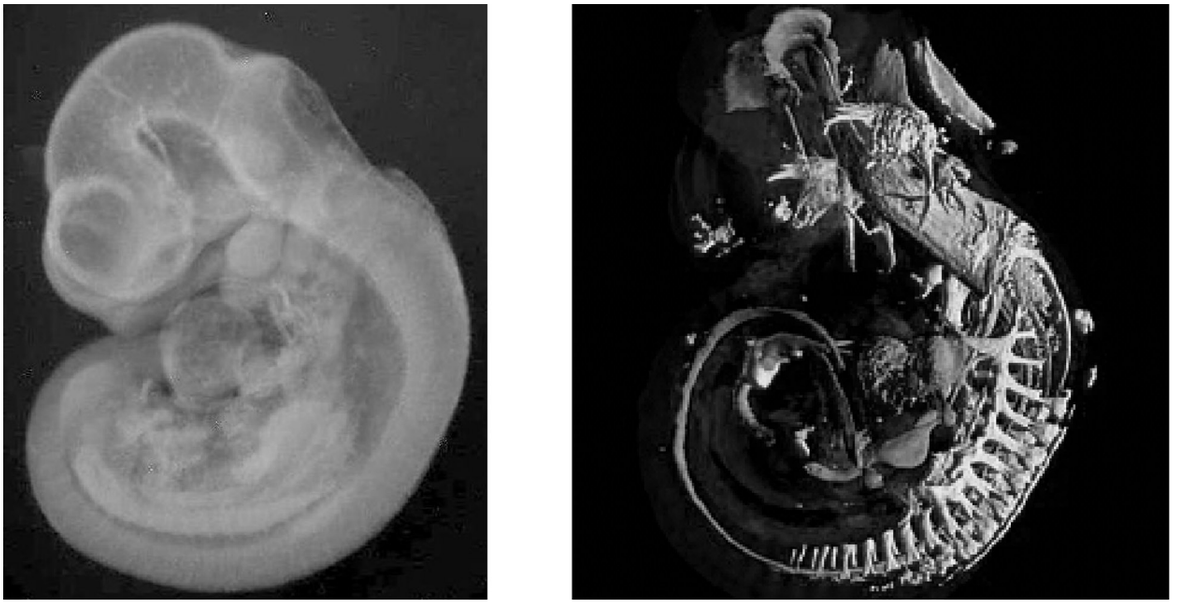

Fig. 1. Example images from OPT-captured 3D data. The left picture shows a volume rendering of a grey-level voxel image. The right picture shows a rendering of the expression pattern of two genes against the background of the autofluorescence of the embryonic tissue. The authors thank Dr. Ulf Ahlgren for the data for these figures.

paper atlas (Kaufman, 1994), and those digitized from plastic embedded sections.

The wax sections provide a model embryo with resolution $4 \times 4 \times 7$ microns and the plastic embedded $2 \times 2 \times 2$ microns. The next generation of models will be first captured using optical projection tomography (OPT): a technique which can capture full 3D data without physical sectioning over the scale and resolution range required for embryo studies (Sharpe et al., 2002); an example of OPT images is shown in Fig. 1. The OPT-scanned embryos are then sectioned and digitized at high resolution.

This will therefore combine the geometric fidelity of true 3D imaging with the high spatial and tissue contrast of the histology. Each model is from a single animal with the intention of providing a representative framework for analysis, and to provide multiple models at each stage to allow more accurate mapping from experimental data. It is intended that there will be a spatial mapping between stages so that a full spatio-temporal framework is available. The models provide a context-free, digital, 3D mouse that can be visualized and virtually sectioned in any plane to show internal tissue histology onto which structural and experimental information can be mapped.

The second component is the anatomical ontology, which is a set of anatomical and tissue terms for each developmental stage organized as a part-of hierarchy (Bard et al., 1998). Each named tissue is a node in the hierarchy. Except for the end or "leaf" nodes, which represent the smallest tissue components in the ontology, each node has component "child" tissues. Each node can in principle be a child of multiple "parent" nodes; which allows for different grouping of terms, so that, technically speaking, the hierarchy is a directed acyclic

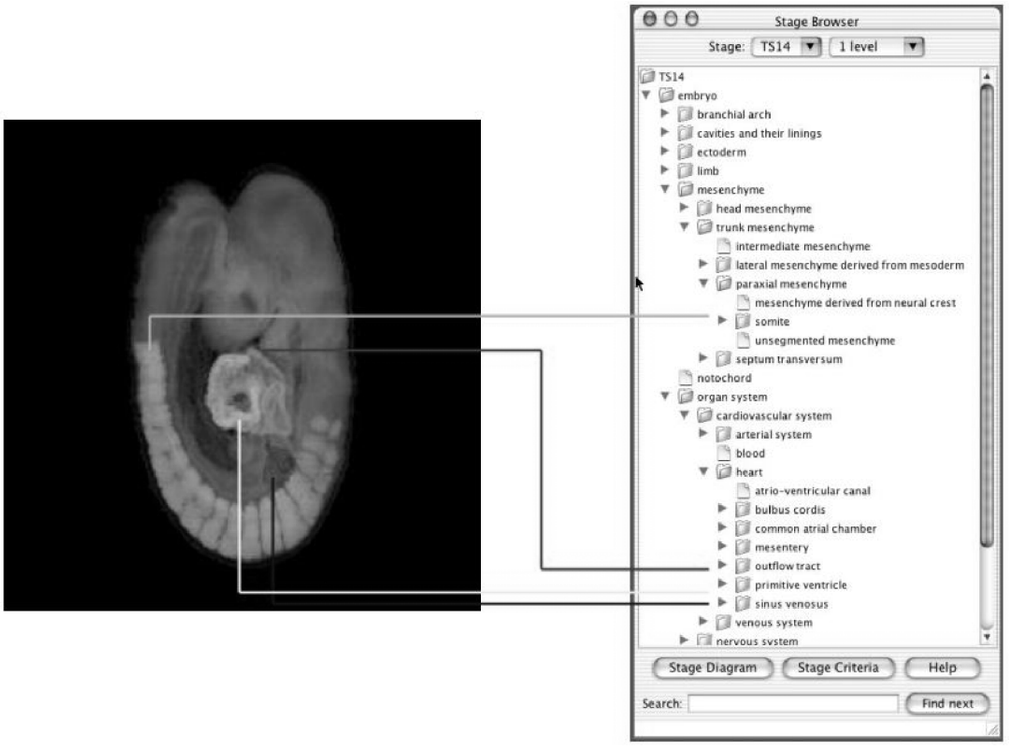

Fig. 2. LHS a three-dimensional rendering of the Theiler stage (TS) -14 voxel model with a number of anatomical domains displayed as 3D colored regions. The RHS picture is a screenshot of the anatomy browser application showing some of the anatomical hierarchy for TS14. The colored lines illustrate how the ontology terms are linked to the spatial domains within the embryo model. (Color figure available online.)

graph (DAG). In addition to the part-of links within a given Theiler stage, there are also "lineage" links between stages. These provide the temporal linking through the ontology. This ontology is available directly from a database or can be downloaded in a number of forms.

The third key component in the EMAP framework is the set of 3D spatial regions, defined within the embryo model coordinates and linked to specific terms in the ontology. We use the term domain for a 3D spatial region, and these anatomical domains provide the translation from text to space and vice versa. The anatomical nomenclature can be considered a representation of embryo space that is explicit but non-geometric, the models provide an iconic representation with implicit spatial relationships. This basic organization is illustrated in Fig. 2.

The anatomical domains within the 3D embryo models have been defined by a combination of image processing and manual editing. Some of the boundaries and tissue divisions are most easily defined by selecting section views that cut the boundary perpendicularly or by defining the plane dividing two components. For these operations, we developed a computer program, MAPaint, that can also be used to map data.

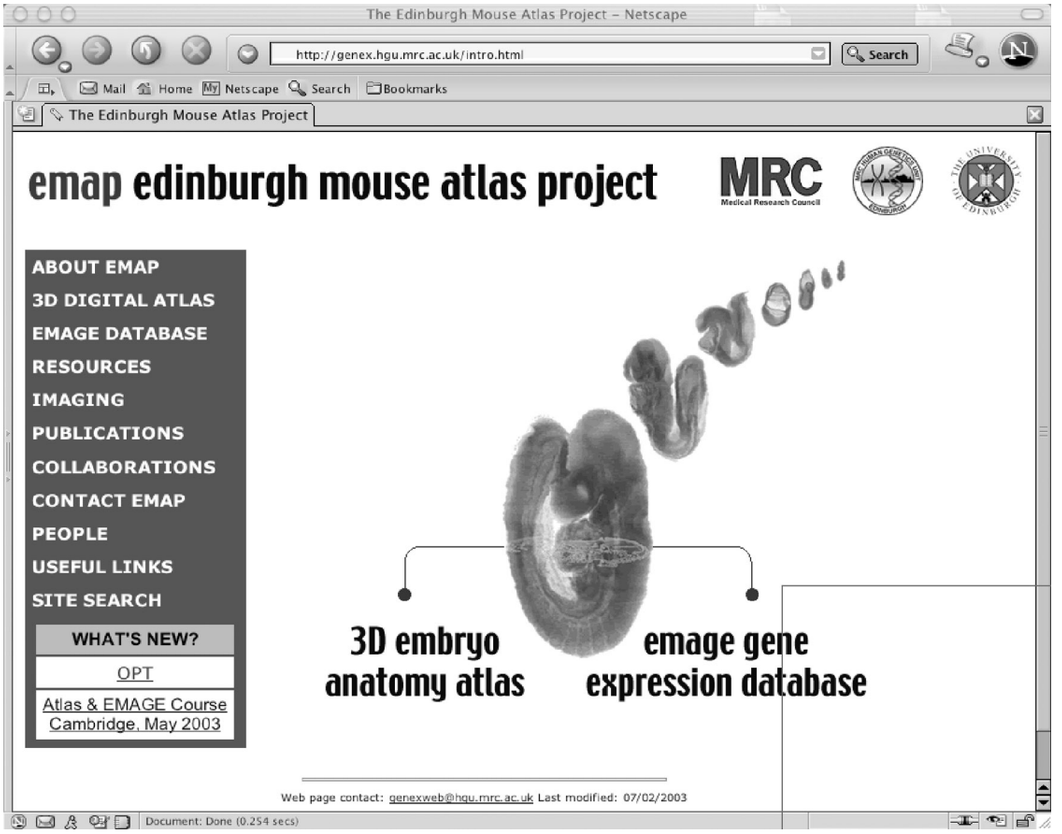

Fig. 3. Worldwide web interface to the EMAP and EMAGE internet resources. The site also includes links for other resources, embryo developmental information, software, and image data.

The atlas is implemented as an object-oriented database, written in C++ and uses ObjectStore (www.objectstore.net) as the database management system (DBMS).

The anatomical nomenclature and many of the graphical features of the Atlas can be accessed online using Java applications from the EMAP website (http://genex.hgu.mrc.ac.uk/) (Fig. 3).

The complete voxel models, anatomical hierarchies, and domains are accessed from CD-ROMs produced by the Medical Research Council (MRC) and can be ordered from the same website.

EMAGE Spatially Mapped Gene Expression Database

The EMAP atlas was developed for the purpose of housing a gene-expression database for mouse development. The anatomy ontology has already been used within a number of other databases, including the Jackson Laboratory Mouse Genome Informatics gene-expression database (MGI/GXD), which indexes expression patterns using text descriptions. The aim of EMAGE however is to contain spatially mapped gene-expression data. The basic idea is simple. Each voxel can be thought of as a region of space at a certain loca-

tion in time. Data associated with that location can be added to that voxel and later queried in terms of location as well as other properties of the data. Each voxel can therefore be considered a "pigeon hole" into which information is posted. This type of information could be represented in a relational database but the mapping and spatial processing would be very inefficient. More appropriate is to gather a voxel or group of voxels (a domain) into a single image structure associated with the body of other data. A query is then a combination of standard database-style processing with image processing operations to access the spatial data. This combination and data organization makes the use of an object-oriented design most appropriate. *In situ* gene-expression data is imaged in a variety of ways. Initial studies typically involve whole-mount processing and image capture using a low-power dissection microscope. This data is a 2D projection of the 3D embryo and can therefore only act as a guide to the true expression distribution. If more detailed analysis and accurate spatial location is required, then the next step is to physically section the embryo and stain individual sections for high-power microscopy. This data is intrinsically three-dimensional but with the problem that the mapping of the section into 3D space needs to be recovered. Other techniques can provide more direct 3D information for example confocal microscopy (CLSM), magnetic resonance microscopy (MRM) and the recently invented OPT.

Bringing the *in situ* data into registration with the embryo models requires that the non-linear warp transformation from experimental image to the model be established. Tools for this process are discussed in the following. In addition to the image data for each submission, there is a set of additional central information to fully define the experiment.

Resources and Tools

EMAP and EMAGE are readily accessible over the Internet (Fig. 3). In this section we present some of the resources and tools that are available with an overview of the capabilities for each tool. In most cases, the software for the tool is written in Java and is run as an application on the user's machine. So that the software is updated automatically, all Java software is delivered and managed using Java Webstart, an application program freely available from Sun Microsystems for all common operating systems including Sun Unix, Linux, Mac OS X, and Microsoft Windows. The benefit of Webstart is that Web Browser dependencies are eliminated and in fact the application, although originally started by "clicking" on a URL link on a webpage, can be started from Webstart itself or from a desktop icon. Webstart maintains a local cache of the software, which it compares with the remote copy (if online), and, if necessary, updates the local copy. If the software is stable, then this significantly reduces the start-up time because very little data needs to be transferred. To use the interfaces, it is necessary to have Webstart installed once (instructions for this are provided on the EMAP website).

WWW and CD-ROM Atlas

Internet access provides a graphical view and browsing interface to the atlas models and databases. It is possible to browse the atlas models and ontologies by using tools provided on the website pages. For more sophisticated access to the models, for example interactive resectioning of the histology models at arbitrary planes or 3D visualization of the anatomical domains, there are a number of CD-ROMs available. These provide the full 3D data and allow mapping of gene-expression data for submission to the database.

The EMAP website has two primary access pages, the first is to the EMAP atlas itself, where the user is presented with a graphical index of the reconstructions, anatomy, and anatomical domains. The other is to the gene-expression database. The EMAGE interface has been developed in Java and is downloadable from the EMAGE webpage. This download provides the query and submission interface, which allows the user to build a submission for the database, manage local data, and review and compare data from the database. For mapping and submission of data there is a specific "Mapping and Analysis" CD. Specific interfaces available from these pages are discussed in the following.

Databases

The anatomy ontology databases, both mouse and human, and the gene-expression database EMAGE can be accessed online. In each case, the user interfaces have been developed using Java and are downloaded automatically when the link on the webpage is selected. Subsequently, the applications can be started directly from a desktop icon, local file, or by reselecting the link. In the case of the EMAGE query and submission interface, the first-time user is requested to register their name and email address. This is to determine the number and range of users of the database.

The mouse anatomy database is also provided with a series of webpages of supplementary information on mouse development, including a table of the primary staging criteria with links to stage-specific pages with more detail, diagrams, and images of representative embryos.

Atlas Browsing Tools

The EMAP atlas is available in a number of ways. The simplest is directly over the Internet using easily accessible viewers. The most comprehensive view of the data is provided by tools that use the EMAP CD-ROMs to provide interactive viewing of the grey-level models with the anatomy overlaid. These tools also provide opportunity for mapping and defining new spatial information.

Anatomy Database Browser

This interface provides query and browsing of the Mouse Atlas anatomy ontology (www.ana.ed.ac.uk/anatomy/database/humat).

The interface is implemented in Java and is downloaded and run using Java Webstart. Figure 4 shows a view of the interface.

The primary tool is a simple tree-browser. For this, the user selects the developmental (Theiler) stage required, which is then displayed in collapsed form. The user can then expand the tree as required or select expansion of the whole tree to a certain depth. The tree components are active and right-button select (<Control> select on Mac) will result in a menu. The capability at present is to select "Details" which will result in a dialog with more detail of the selected component: full name of the component (as a path in the tree), range of Theiler stages for this component, lineage links. Synonyms, and status flags. In addition to the stage browser, the interface provides a query facility to search on name and anatomy component identity number across all stages or a selected stage range. This also provides links to the component-details box.

Section and Anatomy Viewer

The section browser is a Java applet that is accessible for all architectures. It provides a limited series of pre-calculated section views with mapped anatomy and the anatomy ontology. The views can be selected from transverse, frontal, and sagittal planes through the embryo reconstruction. For each selection, the visible

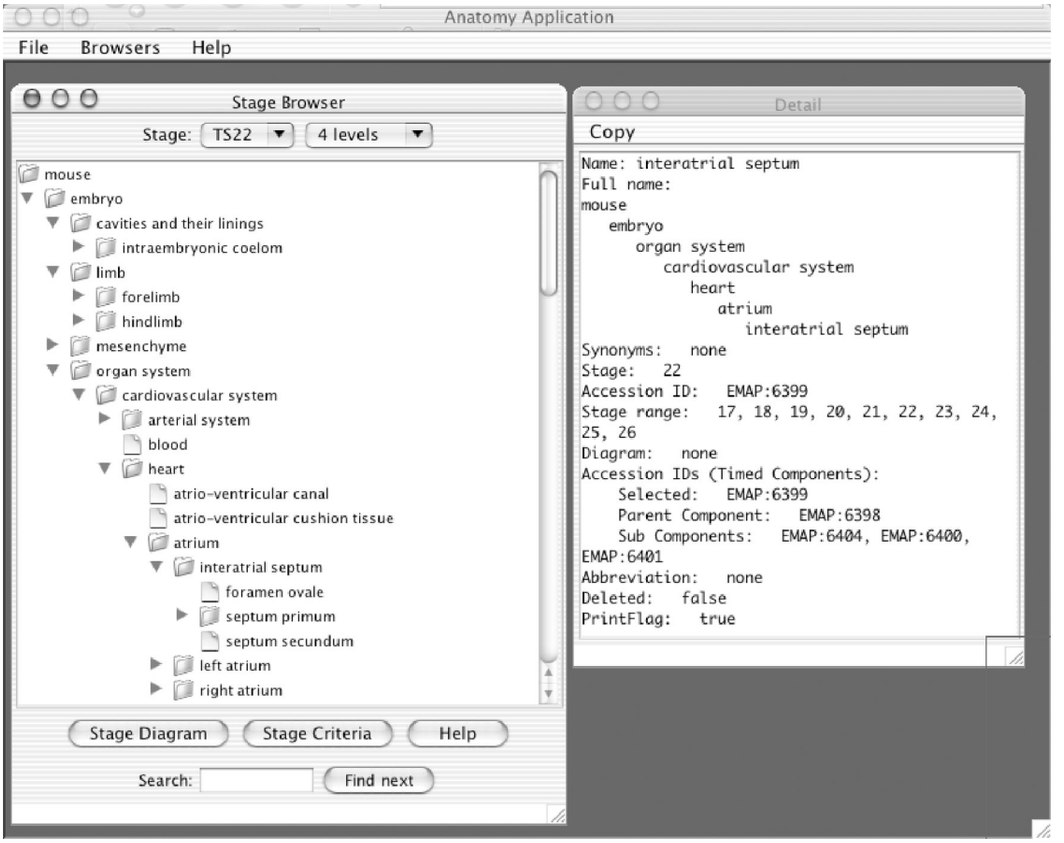

Fig. 4. The anatomy database browser. The left window is an interactive tree view of the anatomy ontology, which is a stage specific part-of hierarchy of the anatomical terms at each developmental stage. The right-hand window provides additional detail of a selected component and its UIDs and relationships.

defined anatomy can be viewed overlaid on the sections by moving the cursor or selecting a point on the section. Each coordinate in the section view is linked to the enclosing tissue domain and thence to the tissue anatomy component. The view provides a simple and effective way to browse the grey-level voxel model of the embryo and enables the user to explore the defined anatomy and anatomy ontology in a natural way.

The browser starts automatically by accessing a link on http:/genex.hgu.mrc.ac.uk/Atlas/and is shown in Fig. 5.

This page also provides links to other resources available for the selected Theiler stage. In addition to providing a graphical view of the atlas, the section browser can be used to query the MGI/GXD gene-expression database at the Jackson Laboratory. This is a very large database, which includes gene-expression data in the mouse indexed using the anatomical ontology from EMAP. Key to the interoperability are the unique identifiers (UID) or accession numbers that each component acquires in the EMAP database. These UIDs allow query from the Section Browser

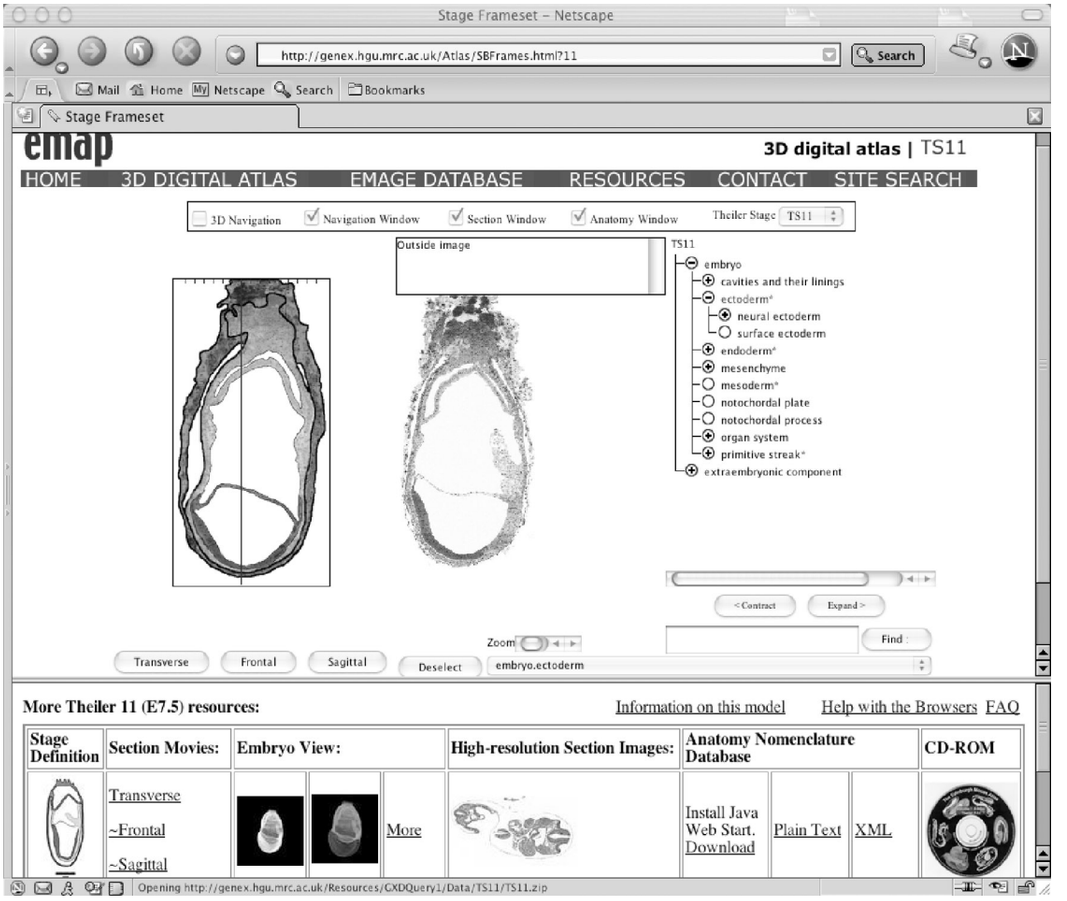

Fig. 5. Section Browser web interface plus additional resources links. The main interface in the top frame is interactive and will highlight tissues as the cursor is passed over the corresponding pixel locations. Tissues can be selected in space or from the ontology trees and used to query the GXD database.

directly on the GXD and thereby provides a graphical query interface.

Browsing the Original Section Images Used to Make the Models

To create the voxel models for each stage of development, serial histological sections of each embryo were digitised and reassembled (stacked) to form the voxel model. The models are then used for resectioning and mapping so the reassembly involves alignment and non-linear warping of the section images onto the 3D spatial coordinate frame. In addition, the original section views are sub-sampled so that the resolution of the stack is approximately isotropic. This makes the resectioning simpler and the digital sections more realistic. The cost is a loss of resolution in the original planes. To allow review and more detailed analysis of the full image data these "original-sections" are provided within a Java user interface. Select the "High resolution section-image" option on the

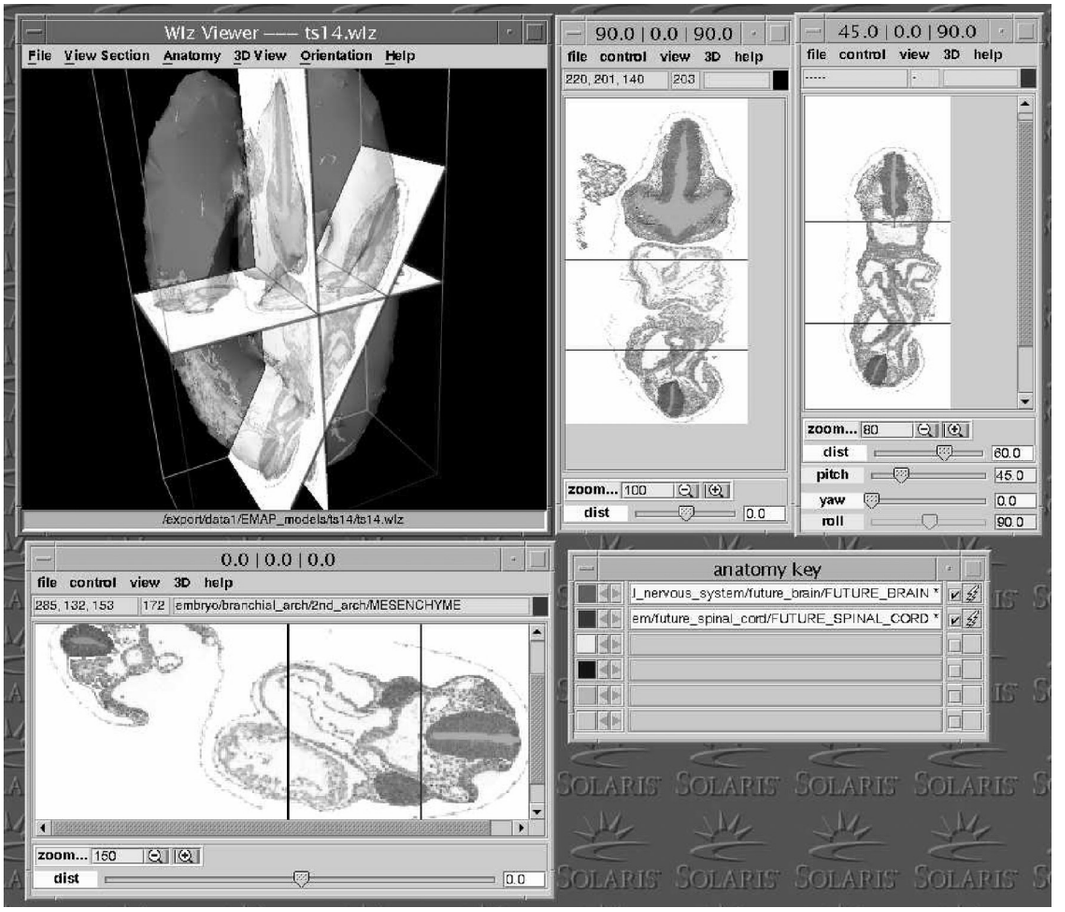

Fig. 6. Screen shot of the Java Atlas viewer. The top-left window is the primary startup window for the application and provides the 3D view and feedback. Shown as a transparent surface is a view of the whole embryo to aid navigation. Any number of section views can be requested, each of which has a user definable orientation, highlight color and 3D display options. In this case, all three have been selected for texture map view which displays the grey-level image in 3D space. In addition the intersecting lines between sections are shown in their respective colors. The top-right section view also shows some of the orientation controls that are available for manipulating the view interactively. The bottom-right window shows the currently selected anatomy components, shown in color on each view and in the 3D window. (Color figure available online.)

page displayed in Fig. 5 allows use of this interface. The window on the left provides a section selector that can be controlled using the up/down arrows to locate any required section. The section number is identical to the plane or z-coordinate provided by other viewers.

Java Atlas Viewer

The Section Browser provides constrained views of the atlas models. For Unix systems, other programs (e.g., MAPaint) provide for arbitrary sectioning and fast interactive browsing. To overcome UNIX dependencies, we have

recently released the Java Atlas Browser, which is a Java application, accessed using Webstart that will provide this functionality for all architectures, including Microsoft Windows based systems. The browser is shown in Fig. 6.

There are two basic windows. The first is the section view, which provides a cut section through the embryo model at any orientation and position. Any number of section views can be displayed simultaneously. The other primary window is the 3D feedback view, which displays a surface model of the embryo, the bounding-box of the data and the location of each of the section views. Other windows provide controls for viewing the anatomy. This browser can be freely downloaded from the genex website (http://genex.hgu.mrc.ac.uk/).

Gene-Expression Data Submission

The basic principle of the EMAGE database is that spatially organized data, for example, the domain of expression of a particular gene at a given stage of development, is spatially mapped by image registration and warping onto the standard framework. The mapped pattern is stored along with the original images and supplementary ancillary data. Interfaces for these aspects are discussed in the following. The philosophy of the database is to accept any submission that satisfies quality criteria in terms of the submitted data and the original image. No check is made of consistency with existing submission, i.e., we expect to include apparently inconsistent results, which will be resolved by further and more detailed experiments.

Mapping Data for Inclusion in EMAGE Using MAPaint

The key capability of MAPaint is the option of defining delineated regions using any section view through the embryo. Thus MAPaint can be used to map gene expression domains or any other spatially organized data from actual histological sections into the equivalent virtual section through the 3D Atlas model.

The first step in this process is to select the appropriate Theiler-stage model using the staging criteria that can be browsed on the EMAP website and, where more than one model exists for a particular stage, choosing the model that most closely resembles the conformation of the original specimen from which the histological section or wholemount photograph comes. It is worth noting here that it will be an advantage to photograph embryos before histology processing.

The second step is to select the digital section plane that best matches the plane of the actual section containing the data to be mapped. Navigation through the model is interactive (provided the voxel model can fit into the computer's main memory). To assist in finding the desired plane, the controls allow rapid location of specific morphological features or existing delineated domains in the model; it is then possible to fix first one, then two points in the visible plane so that all subsequent changes in section angle will keep these points in the plane. Fixing one point leaves two rotational degrees of freedom; fixing two leaves a single rotational degree of freedom. With a little practice, the optimal match can be quickly established in this way. The parameters defining a given section can be saved and restored at any time.

Once the plane of section is selected, there are three options for mapping the data, say a gene expression pattern, into the 3D space at this plane. The mapping process can be painting, compilation of existing defined regions followed by manual editing, and finally these in combination with mapping directly from a source image.

Painting or region delineation is achieved using a number of tools; some of which are purely manual, e.g., simple polyline enclosure, paint-ball, and thresholding. Others include

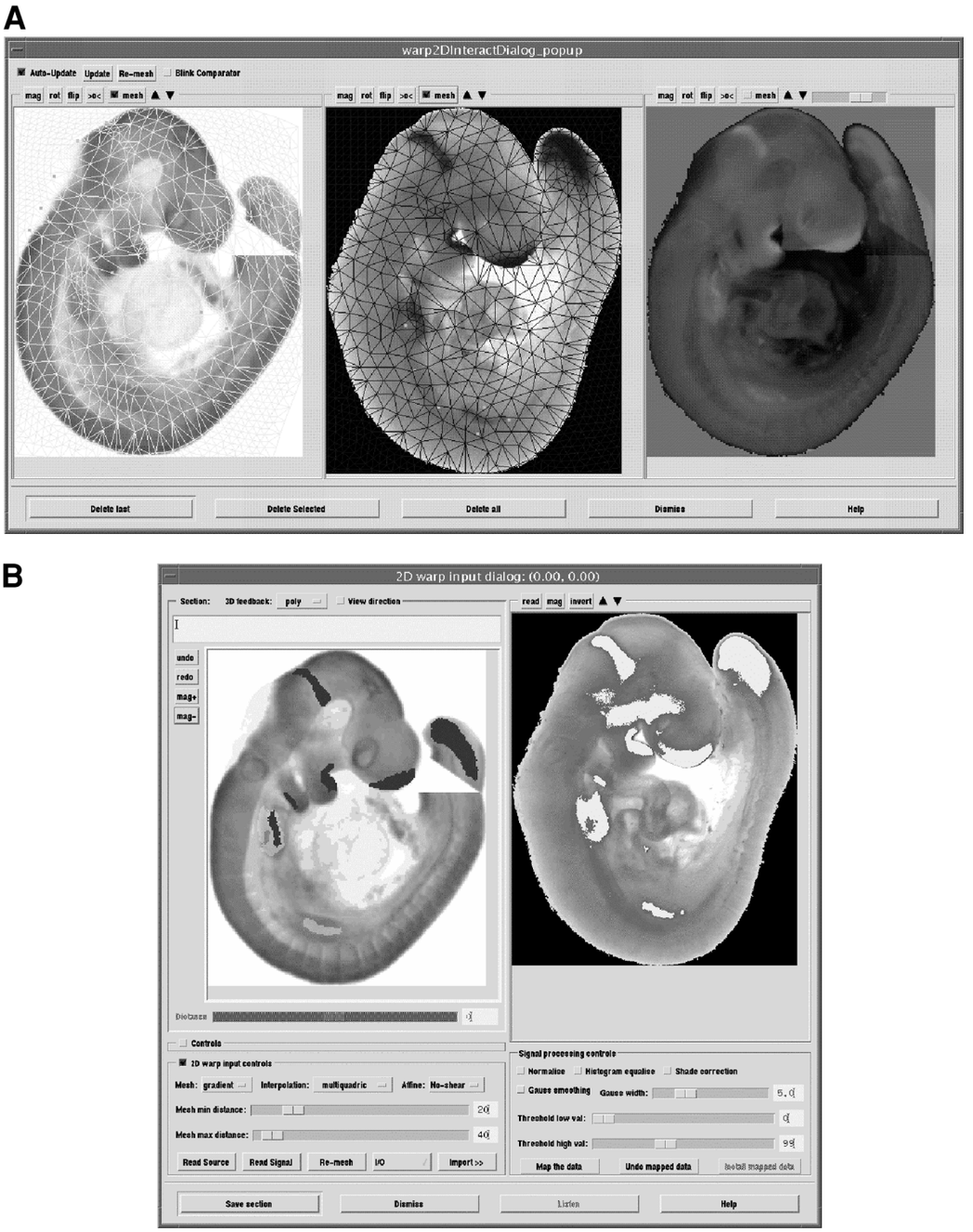

Fig. 7. **(A)** Mapping 2D data onto the standard model. A non-linear warp is defined by a series of tie-points put in by the user. The process takes a few minutes. **(B)** mapping the signal by simple thresholding of the stained embryo image and spatial mapping.

automation for example edge tracking, propagation, and hierarchical dominance. The program is designed to allow simultaneous definition of up to 32 domains. In practice, operators usually restrict painting activity to about 6 domains at a time. As each domain is delineated it immediately becomes visible in any view and can be displayed as a 3D structure in the navigator window (*see* Fig. 7).

Any number of domains can be defined because domains are saved as independent 3D image structures in separate files, for example at Theiler stage 26 there are about 450 anatomical tissues delineated. Domains defined during the same session are assumed to be mutually exclusive: this is a constraint imposed by the system. So that common boundaries only need to be defined once, there is a controllable dominance hierarchy so that the user can control and modify which domain will remain when two domains intersect. Overlapping domains can, of course, be defined and displayed in separate sessions.

Domains that have been defined previously can be edited using a number of command-line operations to create new domains that are, for example, the union, intersection and difference of existing ones. Existing domains, for example, those representing named anatomical structures and previously delineated gene expression patterns can be used as starting points in this process. The resulting new domains can then be manually edited using the MAPaint program as described previously.

Mapping directly from the source image requires the definition of the potentially complex spatial transform from the actual histological section to the standard model. (Note that the same mapping process as is applied to sections can be applied to 2D images of stained embryos in whole-mount preparations.) This process is termed warp input and the current system requires manual interaction to define a series of tie-points. Each tie-point pair establishes a correspondence between the source image and the target view in the model. As the tie-points are defined the source image is warped using thin-plate spline (Baldock and Hill, 1999) or multiquadric interpolation and the user is presented with a composite image that shows the current state of the transformation. In many cases, the spatial transformation can be very accurate with relatively few points and only takes a few minutes of operator time. For larger scale studies, automated mechanisms for this matching are being developed. This sequence of mapping is shown in Fig. 7.

Once the spatial transformation has been established, the signal image can be imported and the signal extracted by simple smoothing and thresholding. This is best undertaken with the original section or whole-mount available under the microscope so that the threshold values can be checked against high-resolution observation of the data. Once the signal has been extracted, the mapped data can be edited to remove artefacts and correct for minor misalignments arising from errors in the non-linear mapping. Using this mechanism, data from single, or a series of, sections can be mapped onto the model and submitted to the database. *See* Fig. 8 for an example of mapped data using the standard protocol for required and optional signal domains.

In summary, mapping directly from the source image involves the following steps:
Select the appropriate Theiler stage model.

- Locate the corresponding section in the model.
- Select the 2D warp input interface.
- Read in the source image and define tie-point correspondences.
- Read in the signal image and define the signal thresholds.
- Map the data and save.

EMAGE Submission Interface

The key property of the EMAGE database is spatial mapping of the gene-expression pat-

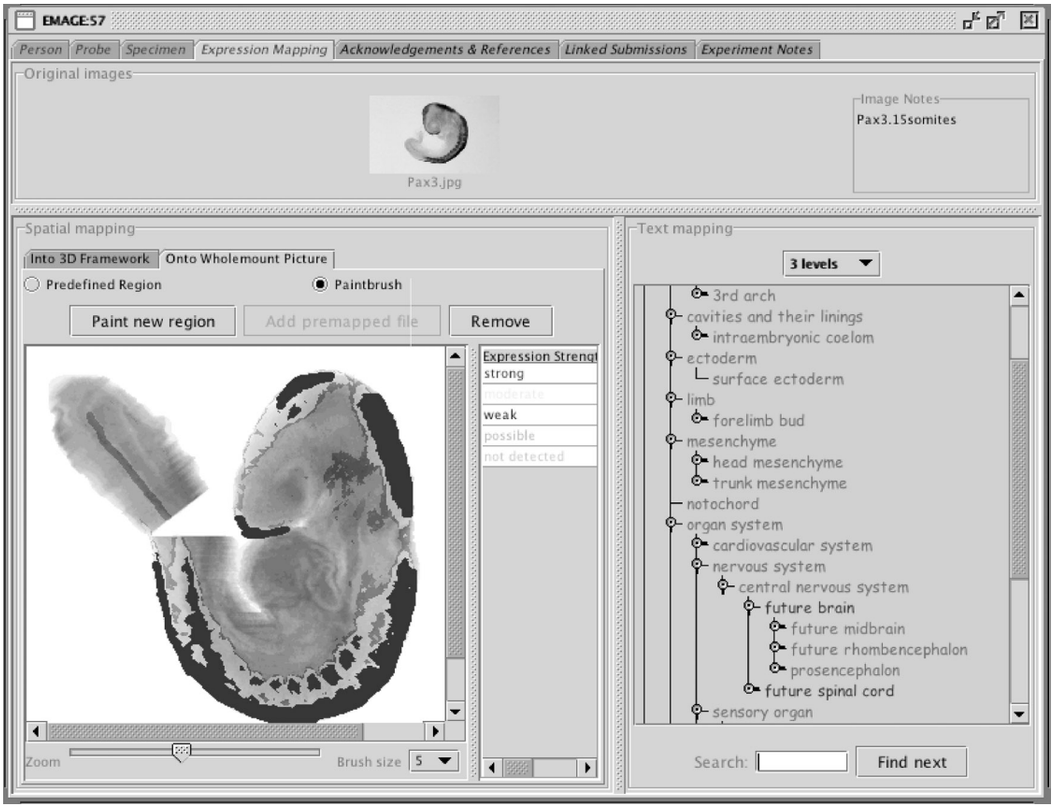

Fig. 8. Submission interface for the EMAGE gene-expression database. This is the "Expression mapping" tab showing spatially mapped and anatomically annotated data. The thumbnail image at the top is the original image data that is a required part of every submission.

terns which makes possible the spatial search and comparison. To make the database useful, it is necessary to include a significant amount of supplementary or ancillary data. This includes source embryo details, probe details, probed gene symbol, author and laboratory details, reference to the publication describing the data (if available), and original images of the unmapped data. This information is managed in the database as part of a single submission and can be returned complete, or in part, as the result of a query on the database. To manage and support the building of a database submission, we have developed a submission interface. This presents the user with a series of forms for attaching essential and optional data, for adding image data, and for defining the anatomical annotation. It also manages the submissions in a local database so an individual user or group can build up their own personal database. This can include oft-repeated data such as personal information, addresses, strain information, probe information, etc. All of this information can then be entered by simple selection from a list.

Figure 8 shows a screen shot of the interface that is available by direct download from the website (http://genex.hgu.mrc.ac.uk/

Emage/) or by using Webstart. When the interface starts up, a query window is displayed. To build a submission, the user is first prompted to create a local database. A new submission to this local database can then be made by selecting 'new-submission' on the File menu. This initiates an empty submission form in which the essential fields indicated by red titles. The interface (*see* Fig. 8) provides a set of pages (tabs) to be filled. Instructions for the entire process can be accessed using the help menu of the interface or downloaded from the website as notes for a training course. The first essential information is the name and contact details of the author and source laboratory. Essential information on the second page defines the probe and method used to visualize the gene-expression pattern, the third page asks for detail of the embryo, strain, stage, etc. The fourth page is for the gene-expression results in the form of a spatial mapping or anatomical annotation. A fifth page includes space for details of any publication or linked submissions. On each of these pages, there are additional optional fields and space for additional free text to provide extra detail of the experimental procedure, data interpretation or analysis of particular patterns. Finally there is a page for lab-specific detail that is for the local use, and not for transmission to the central database.

Each submission record corresponds to a single gene (probe) on a single developmental stage captured during a single experiment. Typically this will be a single embryo but that is not necessary. If the same embryo is used for multiple gene-expression assays at the same time, e.g., by double (or more) labeling, or by using sections in sequence for each gene then the data-mapping and comparison is likely to be more reliable than for data from two completely independent studies. For this situation, different submissions can be linked and the nature of the link recorded as "part of the same experiment."

Gene expression can be spatially mapped to any of the Atlas models. Expression patterns can be mapped not only to the 3D reconstruction for each stage, but also to a number of other standard views, typically lateral views of the whole embryo. The standard procedure requires the definition of a number of domains (which can be "empty") that specify the regions of expression under the headings "strong," "medium," "weak," "possible," and "not-detected." Finally it is required to make sure that the "not-detected" excludes any areas that are not examined, for example if the section is incomplete or there are areas that are occluded or hidden. This is because it is important to be able to distinguish between parts of the embryo that have been examined during a study but for which no expression could be detected and parts that have not been examined at all. Note that we use the designation "not-detected" rather than "not-expressed" because there may always be expression at a level too low for any given experiment to detect.

Gene-Expression Query

Gene-expression patterns are mapped onto the standard framework, either directly onto the 3D digital models or indirectly by anatomical annotation. This mapping enables query of the data in purely spatial terms in addition to search on the textual and numerical data. In addition, the EMAP Section Browser interface can be used to query the MGI/GXD gene expression database, which uses the standard anatomical terms to annotate expression patterns described in the literature.

The EMAGE query interface is shown in Fig. 9. This interface currently allows a small set of key queries. Spatial queries are invoked by defining a spatial region or query domain, which is then compared with the domains of gene-expression (or not-detected) held in the database. In exactly the same way that the

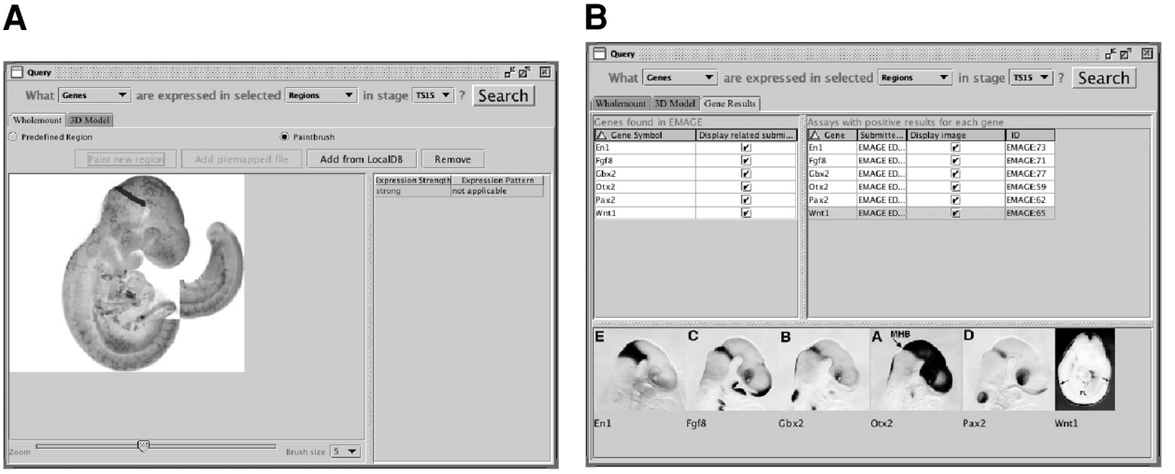

Fig. 9. **(A)** The EMAGE query interface showing a spatial query on Theiler stage 15. **(B)** Results are returned as shown. Double-click on any of the thumbnail image or row in the results table will download the respective submission which can be viewed in detail and stored in a local database.

mapped region is defined directly on the digital models or indirectly via the anatomy, the query domain is also defined directly or indirectly. All submissions in the database are subjected to the query, independently of how the data was originally mapped.

Defining a spatial query is done by selecting the required stage, selecting the required model, and then using one of three methods to define the query domain. The first is to select one or more pre-defined regions, the second is to paint directly onto the digital model, and the third is to read in a pre-defined region comprising a domain delineated by MAPaint or a domain of expression of another gene. In fact, any combination of these options is possible, allowing arbitrarily complex combinations of existing domains followed by manual editing.

Once defined the user selects "Search" to return "hits" on the database in the form of two simple lists. The first is the list of genes that are found to be expressed and the second is the full list of submissions that satisfy the query. The latter will, in general, be longer than the gene list since there will be more than one submission per gene.

Conclusion

In this paper we have presented some of the basic tools developed for the Edinburgh Mouse Atlas (EMAP) and an associated gene-expression database (EMAGE). The essence of the framework is to allow the comparison and interpretation of data within the same spatial coordinate system. This includes direct spatial querying. The tools presented here allow browsing of the atlas and anatomical ontology, mapping and analysis of the spatial data and query of the database. The tools are primarily developed in Java and delivered using Java Webstart which allows for automatic update and ensures that the interfaces are largely architecture independent. Specifically all Java tools will work under Unix (Solaris, IRIX), Linux, Mac OS X, and Microsoft Windows. MAPaint is available on all except Windows. All soft-

ware is freely available for academic, non-profit use, including source code. The anatomy ontologies are available in a number of forms, including XML in OBO (www.geneontology.org/doc/gobo.html), for importation to other databases provided the component identity numbers are preserved. The atlas is provided online or on CD-ROM and the gene-expression database is online.

EMAGE is the first spatially mapped and searchable gene-expression database for any organism, but it is not the last. There are many projects underway to provide similar functionality across other model organisms, embryo and adult. A key challenge is to make sure that these databases are interoperable. This is being addressed at the ontology level by the Cross Species Anatomy Network (XSPAN: http://www.xspan.org/). In addition, the Standards and Ontologies for Functional Genomics (SOFG: http://genex.hgu.mrc.ac.uk/sofg) group are investigating the integration of mouse and human terms. Within EMAP we are investigating direct spatial mapping between mouse and human. These efforts will extend to include temporal as well as cross-species mapping.

Acknowledgments

The development of the EMAP atlas and EMAGE database has been undertaken as a core-funded MRC program at the MRC Human Genetics Unit in collaboration with Matthew Kaufman and Jonathan Bard at the University of Edinburgh. We acknowledge the support of the BBSRC and MRC to the University of Edinburgh and EU FP5 (EURExpress - QLG2-1999-00793) and NIH (NIH HD39928-02) funding to the group in the MRC.

References

Baldock, R. A., Bard, J., Kaufman, M. H., and Davidson D. (1992) A real mouse for your computer. BioEssays 14, 501–502.

Baldock, R. A. and Hill, W. (1999) Image Warping and Spatial Data Mapping, Image Processing and Analysis: A Practical Approach. (Baldock, R. A. and Graham, J., eds.) Oxford University Press, Oxford, UK, pp. 261–288

Bard, B. L. B., Kaufman, M., Dubreuil, C., et al. (1998) An internet-accessible database of mouse developmental anatomy based on a systematic nomenclature. Mechanisms Dev. 74, 111–120.

Brune, R., Bard, J. B. L., Dubreuil, C., et al. (1999) A three-dimensional model of the mouse at embryonic day 9. Dev. Biol. 216, 457–468.

Davidson, D., Bard, J., Brune, R., et al. (1997) The mouse atlas and graphical gene-expression database. Seminars Cell Dev. Biol. 8, 509–517.

Downes, K. M. and Davies, T. (1993) Staging of gastrulating mouse embryos by morphological landmarks in the dissecting microscope. Development 118, 1255–1266.

Kaufman, M. H. (1992) The Atlas of Mouse Development, Academic Press, London, UK.

MacKenzie-Graham, A., Jones, E. S., Shattuck, D. W., Dinov, I., Bota, M., and Toga, A. W. (2003) The informatics of a C57BL/6 mouse brain atlas. Neuroinformatics 1, 397–410.

Martone, M. E., Zhang, S., Gupta, A., et al. (2003) The cell-centered database: a database for multiscale structural and protein localization data from light and electron microscopy. Neuroinformatics 1, 379–396.

Rosen, G. D., La Porte, N. T., Diechtiareff, B., et al. (2003) Informatics center for mouse genomics: the dissection of complex traits of the nervous system. Neuroinformatics 1, 327–342.

Sharpe, J., Ahlgren, U., Perry, P., et al. (2002) Optical projection tomography as a tool for 3D microscopy and gene expression studies. Science 296, 541–545.

Theiler, K. (1989) The house mouse: Atlas of embryonic development. Springer Verlag, New York.