

EMAGE mouse embryo spatial gene expression database: 2010 update

Lorna Richardson, Shanmugasundaram Venkataraman, Peter Stevenson, Yiya Yang, Nicholas Burton, Jianguo Rao, Malcolm Fisher, Richard A. Baldock, Duncan R. Davidson and Jeffrey H. Christiansen*

MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, Western General Hospital, Edinburgh EH4 2XU, UK

Received August 27, 2009; Accepted August 31, 2009

ABSTRACT

EMAGE (<http://www.emouseatlas.org/emage>) is a freely available online database of *in situ* gene expression patterns in the developing mouse embryo. Gene expression domains from raw images are extracted and integrated spatially into a set of standard 3D virtual mouse embryos at different stages of development, which allows data interrogation by spatial methods. An anatomy ontology is also used to describe sites of expression, which allows data to be queried using text-based methods. Here, we describe recent enhancements to EMAGE including: the release of a completely re-designed website, which offers integration of many different search functions in HTML web pages, improved user feedback and the ability to find similar expression patterns at the click of a button; back-end refactoring from an object oriented to relational architecture, allowing associated SQL access; and the provision of further access by standard formatted URLs and a Java API. We have also increased data coverage by sourcing from a greater selection of journals and developed automated methods for spatial data annotation that are being applied to spatially incorporate the genome-wide (~19 000 gene) 'EURExpress' dataset into EMAGE.

INTRODUCTION

The development of multicellular organisms from single cells into complex functioning entities, is governed by a multitude of gene expression events that are strictly spatio-temporally controlled. Quantitative assays of these events use techniques such as microarrays, serial analysis of gene

expression (SAGE) and 'deep sequencing'; however, the finest spatial granularity available through these methods is the size of the original sample, and an appreciation of the intricacies of level and distribution of the global expression profile across a complex biological tissue is usually not attained.

Complementary to these techniques are *in situ* expression profiling methods [e.g. *in situ* hybridization (ISH), immunohistochemistry (IHC) and the use of targeted knock-in or gene trapped *in situ* reporters (ISR)] that provide an excellent appreciation of the spatial intricacies of a gene expression pattern across a structurally complex sample.

Whilst *in situ* expression techniques are used routinely to assess gene expression, it remains a major challenge to integrate the data-dense information contained in the output images into formats that are amenable for data storage and subsequent computational search and analysis. Generally, the method chosen by *in situ* expression databases has been for a human annotator to use a controlled anatomy vocabulary to describe the pattern of gene expression that is seen in the data images (1–3). This method, whilst being useful for simple indexing, cannot easily be used to describe spatial intricacies of gene expression patterns. To complement the text annotation method, we have pioneered the use of a spatial annotation approach, where digital representations of different gene expression patterns are integrated into a common spatial framework (4). This forms the basis of the EMAGE database (5). Importantly, we have also developed tools to query and mine these data based on spatial relationships between any number of gene expression patterns in the EMAGE database (5).

Data in EMAGE are sourced from the community (primarily via the literature and mid- to large-scale screening projects) and is curated by full-time editorial staff to ensure data accuracy and consistency. New data are added regularly. Here, we focus on recent progress in

*To whom correspondence should be addressed. Tel: +44 131 332 2471; Fax +44 131 467 8456; Email: Jeff.Christiansen@hgu.mrc.ac.uk

terms of additional content, improved data accessibility to users, new querying capabilities and changes to the database architecture.

DATA CONTENT

Manually mapped data

The manual spatial mapping method employed by EMAGE has been described earlier (6). This method requires a human annotator to identify anatomically similar points between each data image and the EMAGE reference embryos, as well as to determine the above-noise signal in each data image. This case-by-case approach is particularly suitable for annotating data from the literature, as these images are photographed in many different laboratories in non-standardized ways, varying widely in embryo posture/view and colour of signal.

Over the past 2 years, we have focused our manual annotation attentions on increasing gene coverage for whole-mount stained samples at mid-gestation stages [9.5–11.5 *dpc* (days *post coitum*)]. The primary source of manually annotated data in EMAGE continues to be the literature and locating suitable data has been achieved by searching the GXD Gene Expression Literature Index (1). Over the past 2 years, we have worked with the publishers of 37 additional journals to grant us copyright agreements that allow reproduction of images and incorporation of data into EMAGE. Currently, data images originating from 41 journals (see http://www.emouseatlas.org/emage/about/copyright_journal_list.html) are incorporated into EMAGE.

In addition, one large ISH dataset has been directly submitted to EMAGE (7), and this data, which covers approximately 400 entries, primarily for whole-mount specimens aged between 7.0 and 9.0 *dpc*, has been manually annotated and incorporated into EMAGE.

As of August 2009, there were 5772 entries with a spatial annotation generated by manual methods for 2532 genes in EMAGE (~1000 entries/~600 genes added each year—see Supplementary Figure S1 for a graph showing manual annotation rate and Supplementary Figure S2 for a breakdown of mid-2009 content by stage of development). As with all data in EMAGE, these data also include an accompanying standardized text-based description of the sites of expression as well as a description of the experimental details that adheres to the minimum information standard for ISH and IHC experiments (MISFISHIE) (8).

Automated spatially annotated data

In addition to the manually annotated entries discussed above, EMAGE now includes data that has been spatially mapped using an automated method. Automation requires data that has been consistently produced, such as the genome-scale EURExpress dataset (<http://www.eurexpress.org>). The EURExpress project has generated ISH data for ~19 000 mouse genes on sections derived from ~5000 14.5 *dpc* embryos, according to a standard protocol (~25 equally spaced sagittal sections per gene). In order to spatially annotate this very large dataset,

we have devised fully automated methods to discern between (i) tissue and microscope slide and (ii) unstained tissue and signal (and apparent strength of the signal) in these images. We have developed methods to align the tissue sections from each data embryo, based on edge recognition and the shape of the adjacent sections (see Supplementary Figure S3 for an outline of the method, J. Rao *et al.*, manuscript in preparation). Ultimately, we plan to carry out full 3D mapping of the EURExpress data into EMAGE by spatially registering data from each assayed section into the corresponding 3D co-ordinates of an age-matched 3D reference model embryo. In a first-pass approach, we have used a 2D projection through the aligned sections for each embryo to create an image similar to a conventional image of a whole-mount preparation. This ‘pseudo-whole-mount’ image is then mapped to a corresponding 2D ‘whole-mount’ view of the reference embryo model using a semi-automated version of our usual methods. The automated mapping method is constrained by a small number of manually placed tie points. For each gene, the apparent strength of signal from each constituent section is determined automatically on the basis of optical density and colour, then the values for all contributing sections are averaged across the alignment. This expression signal is automatically transferred to the correct spatial co-ordinates of the 2D reference object using the data embryo to reference-model mapping. The mapped expression domains can be used for spatial interrogation using standard EMAGE methods. Data of this type started to be incorporated into EMAGE in August 2009 (see EMAGE:6089 <http://www.emouseatlas.org/emagewebapp/router?id=6089> for an example). Entry of the full ~19 000 gene set is scheduled for completion by December 2009 (Supplementary Figure S4).

Other data

We continue to incorporate 3D image data [produced using Optical Projection Tomography—OPT—(9)] from the National Institute of Dental and Craniofacial Research funded (NIDCR P50 DE016215-01) craniofacial screening consortium at four stages of development (9.5, 10.5, 11.5 and 12.5 *dpc*). The sites of expression in these data are described using an anatomy text annotation. As of August 2009, data for 70 genes (from 250 embryo samples) have been incorporated into the public database, and a further ~1000 entries for 197 genes are in the curation pipeline. These data are intrinsically 3D and we are actively developing semi-automated techniques to map this data in 3D. Preliminary studies using existing tools (Amira—<http://www.amiravis.com>) that can be applied to small sub-regions of the embryo have already proved successful for data mapping (10). Full 3D mapping is technically more demanding because of embryo variation particularly in terms of posture, which rules out existing techniques that can be applied interactively. We are currently testing a novel method based on the constrained distance transform of Hill and Baldock (<http://www.emouseatlas.org/Papers/Hill.pdf>).

NEW WEBSITE WITH INTEGRATED GRAPHICAL USER INTERFACES

We have recently deployed a completely re-designed EMAGE website and associated Graphical User Interfaces, accessed via the new web address <http://www.emouseatlas.org/emage> (Figure 1). The new design is in response to feedback from our users and, most importantly, allows integration of multiple search methods in one simple HTML interface, where several disparate applications were previously required (5,6). All results are now returned in a standard tabular format. Usage assessments were carried out by target end-users (from novice to expert) and their suggestions incorporated in the design. The recommended browser for the EMAGE web interfaces is Firefox 3.0, which is tested thoroughly on MacOS 10.4+, Windows XP, Windows VISTA and SUSE Linux platforms. We also test using Internet Explorer 7 (for Windows XP and VISTA) and Safari 3.x (for Mac OS 10.4+).

All content of EMAGE (except where otherwise noted—see below), is now licensed under a Creative Commons Attribution 3.0 License (<http://creativecommons.org/licenses/by/3.0/>). This constitutes true Open Access and anyone can freely copy, distribute and adapt the website/

database content without permission, as long as the source is attributed as EMAGE. The notable exceptions to this rule are data images that have been sourced from the literature. These are usually licensed under specific copyright agreements between EMAGE and each journal publisher and this is clearly indicated in the notes associated with the images. If EMAGE users wish to reproduce these images, they need to contact each copyright holder directly.

Currently, three main options are available to search EMAGE: by Embryo Space, Gene and Anatomical Structure Name. All searches are now performed by either selecting the appropriate icon on the EMAGE homepage, or by selecting the appropriate item from the Search menu. Additionally, a simple ‘Quick Search’ function can be found at the top right-hand corner of any EMAGE page, which allows searching of the database by several data aspects (original source of data, ID, gene symbol, detection reagent, stage, assay type, specimen type and mutant allele).

Search by embryo space

Search by a user-defined arbitrary region. We have developed a new HTML interface to retrieve examples of gene expression patterns with localized spatial

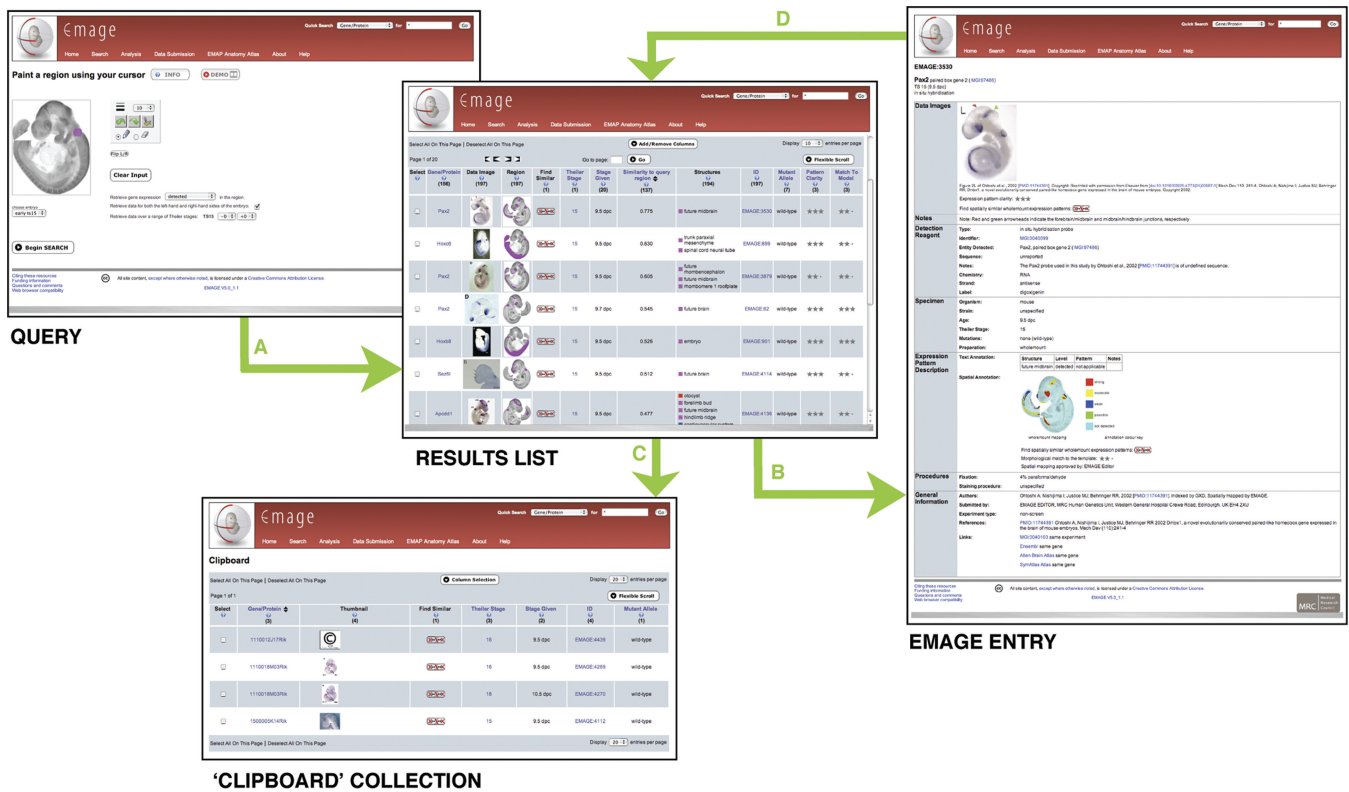


Figure 1. Examples and use of the re-designed EMAGE graphical user interfaces. Four component pages are shown. The **QUERY** page (top left) is used to define a search item. A spatial query is shown here, but gene and anatomical structure queries are also supported. Following a search (process A), a common style **RESULTS LIST** (central panel) is returned from all queries, which is formatted in rows (one per EMAGE entry) and columns (for various data aspects of the EMAGE entries). The table is interactive and the user can choose which columns to display and item ordering. Following the EMAGE:ID hyperlink in a results table (process B), opens an **EMAGE ENTRY** page (shown on the right), which contains full experimental details. A user can also add items from the **RESULTS LIST** to a ‘**CLIPBOARD**’ COLLECTION (lower left) for local storage (process C). The spatial annotation in any **EMAGE ENTRY** can also be used as a search item by using the ‘Find Spatially Similar Patterns’ icon [x-X-x] (process D).

similarity to a user-defined query region. The user first selects the appropriate embryo age and a data type (either whole-mount or 3D/section data) using drop-down lists. If the whole-mount option is selected, the user is presented with an image of the embryo and a palette of painting tools, paints a query region directly onto the embryo image and then selects the search type, e.g. either find genes that are detected in, or not detected in the query region, and finally clicks on the 'Begin Search' icon. The drawing function is provided by a set of drawing tools based on the 'Canvas Painter' javascript library developed by Robayna (<http://caimansys.com/painter/index.html>). This enhanced set of javascripts provides image loading and display, graphical interaction (i.e. drawing), undo/redo, mapping to Microsoft VML (for Internet Explorer compatibility) and EMAGE query. The query is defined by encoding the drawn region as a series of draw operations that is then passed to the Java servlet to be converted to a spatial domain that can be used to query the database. The query domain is presented as a thumbnail image at the top of the standard tabulated results page (described below). Results are ranked according to the spatial similarity between the query region and the gene expression patterns. This search utilizes the Local Spatial Similarity Search Tool (LOSSST) algorithm described previously (5). See Supplementary Movie S1 for an example.

Note that if the 3D/section data search option is selected, the user is currently directed to conduct the search using the EMAGE Java Client (6). We are developing the infrastructure to transfer this functionality to a HTML browser (see Future Directions section).

Search by similarity to an existing EMAGE entry. Recently, we implemented a new form of spatial search: Find Similar Patterns. This function allows the user to query EMAGE to retrieve spatially similar expression patterns to any pattern already housed in EMAGE by simply clicking on the icon marked [x-X-x], which can be found in both an EMAGE entry and in any results list in the 'Find Similar' column. Both WM or 3D/section searches are supported (see Figure 2 for a 3D section example). The search utilizes the LOSSST algorithm (5) and the query domain is a union of the sub-domains from the EMAGE entry that is being used as a query. The results are returned in standard format, ranked by similarity to the query pattern.

Search by anatomical structure

The new interface to search by anatomical terms allows the users to conduct this query across all stages of development. The user types free text into a search query field and this is compared with the EMAP 'Abstract Mouse' Ontology [i.e. a structured list of all anatomical structures across all stages of mouse embryo development (11)]. Matching text strings are displayed in a drop-down list (implemented using javascript) and allows direct linking with the database. As more text is typed, the list narrows. The user selects one or more items from the list to use as search terms (multiple items can be selected using either

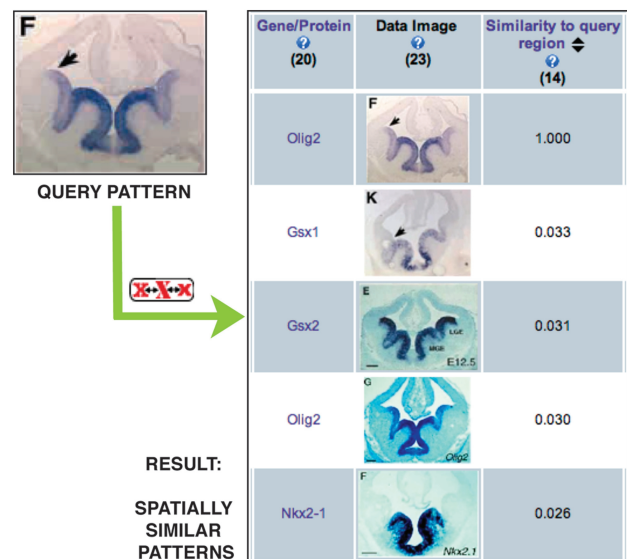


Figure 2. Example of a 'Find Similar Pattern' 3D search. The top left image shows an example of *Olig2* expression in the ventricular layer of both the medial (MGE) and lateral ganglionic eminence (LGE) of the telencephalon at TS20 (EMAGE:768). When the 'Find Similar Pattern' function is used [x-X-x], other examples of spatially similar expression patterns are retrieved from EMAGE—the top 5 matches are shown. Note the first four are also expressed in the ventricular layer of the MGE and LGE, and the 5th entry is progressively less similar and is only expressed in MGE ventricular layer. A high resolution search such as this would not be possible using a text-based approach as the accompanying text annotations for these entries is at a considerably lower resolution (all are annotated to the broad brain region 'telencephalon'). Images reproduced with permission from Development, see EMAGE:452, 768, 951, 1004 and 3334 for full citation details.

the shift or option key in standard fashion). See Supplementary Movie S2 for an example. The underlying search mechanism is based on retrieving EMAGE entries with text annotations to EMAP Anatomy Ontology (11) terms that text-string match the query term, as well as EMAGE entries that are either annotated to a part (for 'detected' queries) or a structural parent (for 'not detected' queries) of such terms. This approach allows easier access for users who are not familiar with the structure or contents of the EMAP Anatomy Ontology.

Search by gene/protein

The new EMAGE gene search allows multiple terms to be used (entered as a comma separated list). See Supplementary Movie S3 for an example. A much greater variety of mouse (and some human) gene/protein identifiers are now supported including: gene/protein symbols; names; synonyms (i.e. withdrawn or unofficial names/symbols); gene IDs [MGI (12), Ensembl (13), VEGA (14), Entrez (15), UniGene (16)] and sequence IDs [GenBank (16), RefSeq (17), SwissProt (18)]. Our nomenclature index is compiled regularly from the MRK_Dump2.rpt and MGI_MouseHumanSequence.rpt reports produced by the acknowledged authority in mouse gene nomenclature, Mouse Genome Informatics (12).

Common results format

We have standardized our results format to provide a consistent layout regardless of the type of query. The results are returned in a table, where each row corresponds to an EMAGE entry and each column refers to a different data field (e.g. ID, thumbnail image, annotations, etc.). The user can control what information is displayed in the table by adding/removing columns or sorting the rows in alpha-numerical order of the contents of relevant columns. See Supplementary Movie S4 for an example showing how to perform these tasks. Items in a results list can also be selected and saved to a 'clipboard' collection for future reference (this functionality uses browser cookies). The use of the clipboard collection function will be extended in the future by developing a secure user login system, which will allow storage of multiple collections on the server side and Boolean comparisons between multiple collections.

Improved layout of EMAGE entries

The layout of the individual EMAGE entry pages has also been re-arranged in accordance with the specification for data structure described by the MISFISHIE information standard (8). The data in each EMAGE entry are now grouped under the broad MISFISHIE data aspects: Data Images; Detection Reagent; Specimen; Expression Pattern Description; Procedures and General Information (Figure 1).

DATABASE ARCHITECTURE

Over the past 2 years, the EMAGE database has been re-engineered from an object oriented to a relational architecture. The new schema has been designed to accommodate the core ISH, IHC or ISR data types in a format that is consistent with the MISFISHIE data standard (8), yet has also been designed such that the database can be easily extended to other organisms or biological data types (e.g. to describe patterns of biological processes such as cell death, division or lineage that may be detected using different experimental procedures).

The EMAGE logic database consists of two parts: the text part of the EMAGE mouse spatial gene expression database (implemented in DB2—<http://www.ibm.com/db2>) and an image add-on component (including both the original data images and mapped expression domains). The schema of the DB2 part is available online (<http://www.emouseatlas.org/emage/search/sql.html>).

In order to provide multiple types of access, most EMAGE clients communicate with EMAGE application servers that access the DB2 database and process image manipulations. The EMAGE application server software is written in Java. It accesses the DB2 database via the Hibernate object/relational mapping (ORM) solution (<http://www.hibernate.org>) and communicates with a CORBA (<http://www.corba.org>) image server which is written in C++ and uses our bespoke Woolz C library for image processing.

DATA ACCESS

In addition to our previously described EMAGE DAS (6) and Web Services (5) methods, we have implemented three new protocols to allow programmers direct access to EMAGE data.

Invoking an EMAGE query using a URL

We have implemented a general method for gene/anatomy queries of EMAGE that can be invoked using a URL. The URL includes name and value pairs in this general form: http://www.emouseatlas.org/emagewebapp/pages/emage_general_query_result.jsf?name1=value1,value2&name2=true. For example, http://www.emouseatlas.org/emagewebapp/pages/emage_general_query_result.jsf?genes=Fgf8&stages=14,15&strengths=detected will result in the gene expression data of the gene Fgf8 expressed at either Theiler Stages 14 or 15. Further instructions can be found at <http://www.emouseatlas.org/emage/search/url.html>.

SQL access

The text part of the EMAGE database can also be accessed using direct SQL queries from anywhere. This is primarily for the benefit of the biologist who has basic SQL skills and only wishes to access the text part of the EMAGE data. Full documentation can be found at <http://www.emouseatlas.org/emage/search/sql.html>.

Java API

For full EMAGE capability (both text and image), EMAGE Java API is provided and can be accessed through the EMAGE RMI application Server (RMI invocation <http://java.sun.com/j2se/1.4.2/docs/api/java/rmi/package-summary.html>). The documentation about how to use the EMAGE Java API can be found at <http://www.emouseatlas.org/emage/help/emageAPI/index.html>

FUTURE DIRECTIONS

EMAGE will continue to include *in situ* gene expression data from a variety of sources, including the literature and large-scale screens, aiming towards genome-wide coverage at multiple stages of development. We will continue to develop and enhance methods to search, analyse and visualize such data in response to the requirements of the research community.

Extended search functions

Currently under development are additional methods for users to search EMAGE for expression of multiple genes at once. These could include genes that are grouped based on function [e.g. by Gene Ontology (19) functional classifications, or pathway groupings [e.g. KEGG (20)] or those found in particular chromosomal locations. The grouping could also be user-defined (e.g. corresponding to the contents of a user's 'clipboard' collection). When used for spatial searches, it will be crucial that appropriate user-feedback is given to assess regions of the embryo where expression of multiple genes is detected. We have

previously shown that a ‘heat-map’ representation is an effective vehicle for delivering this information, and we have already implemented this mode for user feedback in assessing the results of hierarchically clustered, spatially similar gene expression patterns in EMAGE (5). Currently, all heat-maps displayed in EMAGE are pre-generated and the ability to produce such representations in real time for any set of data is required. We are exploring a multiple image overlay functionality offered by an IIP server (see below) to achieve this goal. Interfaces to allow one to extract relevant spatial domains from heat-map images, for example, the maximum region of common expression, then use the selected regions for a query are also required—the underlying algorithms already exist for such calculations (R.A. Baldock, unpublished data)

Additional planned extensions to the EMAGE search interfaces include a secure log-in function to allow multiple search results to be saved, personalized display and filtering options and results download as well as comparison of multiple results lists (i.e. clipboard collections) by multiple Boolean operations. As well as extending the query capabilities of our interfaces we will also implement a BioMart (21) interface to provide interoperable and integrated access across many genomic, proteomic, interaction and phenotype databases for power-bioinformatics users and complex queries. The BioMart interface will also provide standardized web-service and DAS interfaces to complement those already in place.

Improved delivery of images

We are developing a method that allows very large images to be delivered via the web in a timely manner, including images of sections cut from virtual 3D objects in any plane. A prototype version can be accessed by clicking on the ‘Prototype IIP section chooser’ button on the following page: http://www.emouseatlas.org/emage/temp/3d_search.html. The method uses the Internet Imaging Protocol (IIP—<http://iipimage.sourceforge.net/>), which offers a streaming client-server system for the web-based viewing and zooming of high-resolution images via a tile-based approach (a method similar to that used by Google™ Earth). This will provide a user-friendly HTML interface to scroll through the 3D models, find specific sections in which to define a query region and thereby carry out a 3D spatial search as well as to interactively view overlays of multiple gene-expression patterns.

USER HELP

There is dedicated User Support for EMAGE. Please write to ma-edit@hgu.mrc.ac.uk. We also have a User Group where we announce new releases and other relevant information. To subscribe, visit <http://www.jiscmail.ac.uk/lists/MA-EMAGE.html>. EMAGE regularly holds trade display stands at international developmental biology and mouse genetics conferences where hands-on tutorials are available.

CITING EMAGE

To reference EMAGE, please cite this article. For specific data entries, please list the EMAGE:ID and also mention that the data was retrieved from EMAGE, MRC Human Genetics Unit, Edinburgh, UK (<http://www.emouseatlas.org/emage>).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We would like to thank Dr Martin Ringwald and our colleagues at the GXD (MGI, Jackson Laboratory, Maine, USA) for helpful input; The Company of Biologists, Elsevier, Wiley, Cell Press, Nature Publishing Group, the American Society for Microbiology, the Society for Neuroscience and the National Academy of Sciences USA, for copyright agreements allowing the use of data in EMAGE.

FUNDING

Medical Research Council (EMAGE). Funding to pay the Open Access publication charge: Medical Research Council.

REFERENCES

- Smith,C.M., Finger,J.H., Hayamizu,T.F., McCright,I.J., Eppig,J.T., Kadin,J.A., Richardson,J.E. and Ringwald,M. (2007) The mouse Gene Expression Database (GXD): 2007 update. *Nucleic Acids Res.*, **35**, D618–D623.
- Sprague,J., Bayraktaroglu,L., Bradford,Y., Conlin,T., Dunn,N., Fashena,D., Frazer,K., Haendel,M., Howe,D.G., Knight,J. *et al.* (2008) The Zebrafish information network: the zebrafish model organism database provides expanded support for genotypes and phenotypes. *Nucleic Acids Res.*, **36**, D768–D772.
- Visel,A., Thaller,C. and Eichele,G. (2004) GenePaint.org: an atlas of gene expression patterns in the mouse embryo. *Nucleic Acids Res.*, **32**, D552–D556.
- Davidson,D. and Baldock,R. (2001) Bioinformatics beyond sequence: mapping gene function in the embryo. *Nat.Rev.Genet.*, **2**, 409–417.
- Venkataraman,S., Stevenson,P., Yang,Y., Richardson,L., Burton,N., Perry,T.P., Smith,P., Baldock,R.A., Davidson,D.R. and Christiansen,J.H. (2008) EMAGE—Edinburgh Mouse Atlas of Gene Expression: 2008 update. *Nucleic Acids Res.*, **36**, D860–D865.
- Christiansen,J.H., Yang,Y., Venkataraman,S., Richardson,L., Stevenson,P., Burton,N., Baldock,R.A. and Davidson,D.R. (2006) EMAGE: a spatial database of gene expression patterns during mouse embryo development. *Nucleic Acids Res.*, **34**, D637–D641.
- Tamplin,O.J., Kinzel,D., Cox,B.J., Bell,C.E., Rossant,J. and Lickert,H. (2008) Microarray analysis of Foxa2 mutant mouse embryos reveals novel gene expression and inductive roles for the gastrula organizer and its derivatives. *BMC Genomics*, **9**, 511.
- Deutsch,E.W., Ball,C.A., Berman,J.J., Bova,G.S., Brazma,A., Bumgarner,R.E., Campbell,D., Causton,H.C., Christiansen,J.H., Daian,F. *et al.* (2008) Minimum information specification for in situ hybridization and immunohistochemistry experiments (MISFISHIE). *Nat.Biotechnol.*, **26**, 305–312.
- Sharpe,J., Ahlgren,U., Perry,P., Hill,B., Ross,A., Hecksher-Sorensen,J., Baldock,R. and Davidson,D. (2002) Optical projection tomography as a tool for 3D microscopy and gene expression studies. *Science*, **296**, 541–545.

10. Fisher, M.E., Clelland, A.K., Bain, A., Baldock, R.A., Murphy, P., Downie, H., Tickle, C., Davidson, D.R. and Buckland, R.A. (2008) Integrating technologies for comparing 3D gene expression domains in the developing chick limb. *Dev. Biol.*, **317**, 13–23.
11. Bard, J.B.L., Kaufman, M.H., Dubreuil, C., Brune, R.M., Burger, A., Baldock, R.A. and Davidson, D.R. (1998) An internet-accessible database of mouse developmental anatomy based on a systematic nomenclature. *Mech Dev*, **74**, 111–120.
12. Bult, C.J., Eppig, J.T., Kadin, J.A., Richardson, J.E., Blake, J.A. and Group, M.G.D. (2008) The Mouse Genome Database (MGD): mouse biology and model systems. *Nucleic Acids Res.*, **36**, D724–D728.
13. Hubbard, T.J.P., Aken, B.L., Ayling, S., Ballester, B., Beal, K., Bragin, E., Brent, S., Chen, Y., Clapham, P., Clarke, L. *et al.* (2009) Ensembl 2009. *Nucleic Acids Res.*, **37**, D690–D697.
14. Wilming, L.G., Gilbert, J.G.R., Howe, K., Trevanion, S., Hubbard, T. and Harrow, J.L. (2008) The vertebrate genome annotation (Vega) database. *Nucleic Acids Res.*, **36**, D753–D760.
15. Maglott, D., Ostell, J., Pruitt, K.D. and Tatusova, T. (2007) Entrez gene: gene-centered information at NCBI. *Nucleic Acids Res.*, **35**, D26–D31.
16. Sayers, E.W., Barrett, T., Benson, D.A., Bryant, S.H., Canese, K., Chetvernin, V., Church, D.M., DiCuccio, M., Edgar, R., Federhen, S. *et al.* (2009) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **37**, D5–D15.
17. Pruitt, K.D., Tatusova, T. and Maglott, D.R. (2007) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.*, **35**, D61–D65.
18. Bairoch, A., Consortium, U., Bougueleret, L., Altairac, S., Amendolia, V., Auchincloss, A., Argoud-Puy, G., Axelsen, K., Baratin, D., Blatter, M.C. *et al.* (2009) The Universal Protein Resource (UniProt) 2009. *Nucleic Acids Res.*, **37**, D169–D174.
19. Harris, M.A., Deegan, J.I., Lomax, J., Ashburner, M., Tweedie, S., Carbon, S., Lewis, S., Mungall, C., Day-Richter, J., Eilbeck, K. *et al.* (2008) The gene ontology project in 2008. *Nucleic Acids Res.*, **36**, D440–D444.
20. Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., Katayama, T., Kawashima, S., Okuda, S., Tokimatsu, T. *et al.* (2008) KEGG for linking genomes to life and the environment. *Nucleic Acids Res.*, **36**, D480–D484.
21. Smedley, D., Haider, S., Ballester, B., Holland, R., London, D., Thorisson, G. and Kasprzyk, A. (2009) BioMart—biological queries made easy. *BMC Genomics*, **10**, 22.