

The mouse Gene Expression Database (GXD): 2007 update

Constance M. Smith, Jacqueline H. Finger, Terry F. Hayamizu, Ingeborg J. McCright,
Janan T. Eppig, James A. Kadin, Joel E. Richardson and Martin Ringwald*

The Jackson Laboratory, 600 Main Street, Bar Harbor, ME 04609, USA

Received September 14, 2006; Revised October 26, 2006; Accepted October 27, 2006

ABSTRACT

The Gene Expression Database (GXD) provides the scientific community with an extensive and easily searchable database of gene expression information about the mouse. Its primary emphasis is on developmental studies. By integrating different types of expression data, GXD aims to provide comprehensive information about expression patterns of transcripts and proteins in wild-type and mutant mice. Integration with the other Mouse Genome Informatics (MGI) databases places the gene expression information in the context of genetic, sequence, functional and phenotypic information, enabling valuable insights into the molecular biology that underlies developmental and disease processes. In recent years the utility of GXD has been greatly enhanced by a large increase in data content, obtained from the literature and provided by researchers doing large-scale *in situ* and cDNA screens. In addition, we have continued to refine our query and display features to make it easier for users to interrogate the data. GXD is available through the MGI web site at <http://www.informatics.jax.org/> or directly at http://www.informatics.jax.org/menus/expression_menu.shtml.

INTRODUCTION

The laboratory mouse serves as a premier animal model in studying the complex molecular networks that underlie the processes of human development, differentiation and disease. To gain insights into these networks, it is essential to know where, when and in what amounts transcripts and proteins are expressed, and how their expression varies in different mouse strains and mutants. The Gene Expression Database (GXD) addresses this objective in a uniquely comprehensive way. GXD is the only resource that acquires mouse expression data from the literature in a systematic manner, as well as acquiring data directly from conventional and large-scale

providers via electronic data submission and bulk data downloads. GXD integrates various types of mRNA and protein expression information, collects data from all tissue and developmental stages and includes data from many different mouse strains and mutants. Annotations in GXD make extensive use of controlled vocabularies and ontologies to provide the standardization of data that enables complex queries. In addition, GXD is fully integrated with the other databases of the Mouse Genome Informatics (MGI) resource, including the Mouse Genome Database (MGD) (1,2) and the MGI part of the Gene Ontology Project (GO) (3). MGI also maintains comprehensive links to external resources such as sequence databases, Entrez Gene, UniProt, InterPro, Online Mendelian Inheritance in Man (OMIM), PubMed and other mammalian databases (4–15). This robust integration puts the expression data annotated in GXD into a much larger biological and analytical context. Thus, users are able to query using extensive genetic, sequence, functional, expression and phenotypic information.

Other public and laboratory databases have been developed in recent years to store mouse expression data (16–26). They store data from one or two specific assay types and/or focus on specific tissues/developmental stages; they are often dedicated to specific data generation projects. These databases are complementary to the GXD effort. Due to its broad scope, its thorough approach and its data integration and querying capabilities, though, GXD provides a unique resource to the biomedical research community. New data are entered and made publicly available on a daily basis. GXD and its query interfaces have been described previously (27–30). Here we focus on recent progress in terms of data acquisition and querying capabilities.

DATA CONTENT

The Gene Expression Literature Index

GXD curators survey journals to find all published papers that describe endogenous gene expression and knock-in reporter studies done in the embryonic mouse. In a first annotation step, the curators record the genes and ages analyzed and the expression assay types used in these publications. GXD

*To whom correspondence should be addressed. Tel: +1 207 288 6436; Fax: +1 207 288 6132; Email: ringwald@informatics.jax.org

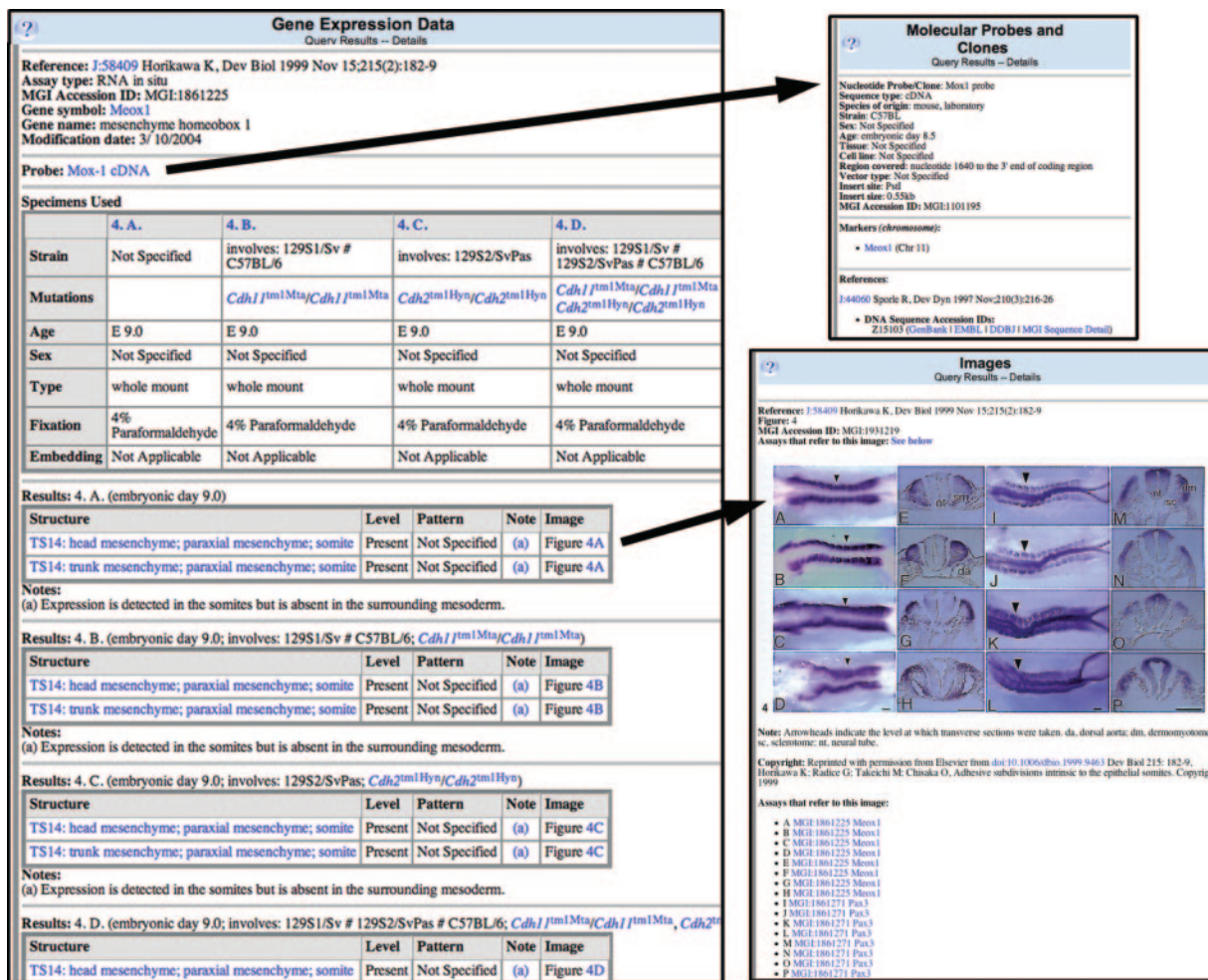


Figure 1. GXD contains detailed expression annotations. This record for an RNA *in situ* hybridization assay (left) illustrates the details included in annotations of expression results. It reports the reference the data were derived from, the assay type and the gene analyzed. Details regarding the nucleotide probe (or antibody) used in the assay are entered into the database and can be accessed via a link on the expression detail page; the detail page for the nucleotide probe used in this assay is pictured (upper right). In the Specimens Used section, the genetic background, age, sex and preparation method of the specimens are described; links are provided to pages that detail the characteristics of mutants used. In the Results section, we report the Theiler stage and tissue examined, as well as the level and pattern of expression as described by the author. If it is a blot assay, the number and sizes of detected bands are reported. We have permission from several publishers, as well as from researchers who have contributed data to us directly, to provide images of the original expression data (lower right) to help users analyze the results.

combines these data with information obtained from PubMed and makes them available for searching via the Gene Expression Literature Index. Therefore, users can query for specific types of expression information in combination with bibliographic information (author, journal, year) or specific words in the title or abstract of publications. The Literature Index is comprehensive and up-to-date; it contains all pertinent journal articles from 1993 to the present and articles from major developmental journals from 1990 to the present. Currently, the index contains >56 500 entries covering nearly 12 300 references analyzing nearly 8700 genes. Thus, it provides a powerful tool to quickly locate expression information in the literature.

Gene expression data

GXD currently collects detailed expression data from the following assay types: RNA *in situ* hybridization,

immunohistochemistry, *in situ* reporter (knock in), northern blot, RT-PCR, western blot, RNase protection and nuclease S1 protection studies. Work is underway to incorporate microarray data as well. As illustrated in Figure 1, expression records in GXD are detailed. Each entry contains a description of the assay type and the molecular probe used in the assay, the genetic origin of the sample and the experimental conditions used. The time and tissue of expression, the authors' description of pattern and strength of expression, the number and sizes of detected bands and sequence information are also recorded. Expression patterns are described using an extensive dictionary of standardized anatomical terms that lists the anatomical structures for each developmental stage in a hierarchical fashion, thus enabling the recording of expression results from assays with different spatial resolution in a consistent manner. The embryonic part of the anatomical dictionary was developed by our collaborators from the Edinburgh Mouse Atlas and Gene Expression

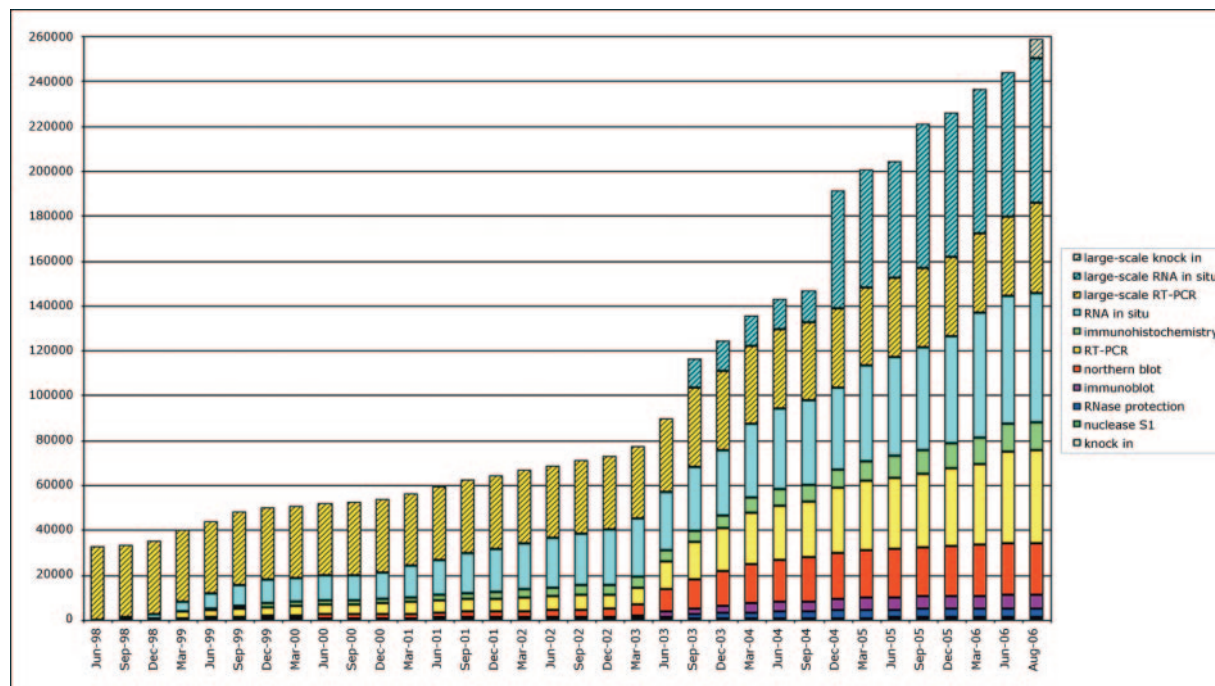


Figure 2. Quantity and type of detailed expression results in GXD. Since first released in June 1998, the amount of data in GXD has steadily increased to >260 000 results today. More than half of these results are from RNA *in situ* hybridization, immunohistochemistry and *in situ* reporter (knock in) assays; these studies yield complex descriptions of expression patterns. The remaining results were obtained from blot assays.

Database (EMAGE) project (31); the adult part was developed by the GXD project (32). As well as enabling complex querying capabilities, these detailed annotations make it easier to interpret and compare expression data.

GXD's data content has increased significantly in recent years (Figure 2). Currently, GXD contains data from >24 600 assays that provide >260 000 detailed expression results for nearly 7700 genes, including expression data from almost 1000 different mouse mutants. Two-thirds of these data are linked to images of the primary expression data; GXD currently contains >43 000 images of expression data. This rapid growth in data content was made possible by the daily annotation of expression data from the literature and through the incorporation of large sets of expression data from large-scale RNA *in situ* hybridization and RT-PCR screens. Recently acquired large data sets include: RNA *in situ* and RT-PCR studies of mouse genes homologous to human chromosome 21 genes (33,34); RNA *in situ* studies that analyzed expression patterns of >1300 transcription factors in the developing mouse brain (35); RNA *in situ* studies examining the expression of >1000 genes during retinal development (36); and RNA *in situ* and RT-PCR studies of >300 RNA-binding protein encoding genes in the developing mouse brain (37). In these instances curators worked with the researchers to fully integrate into GXD the large amounts of supplemental material that accompanied these publications. Curators worked with the laboratories to bring the data into standardized formats and to resolve issues pertinent to nomenclature and referential integrity; the data was reviewed both computationally and manually. The integration of these data into GXD greatly expands the research community's ability to query these data, increasing their utility.

GXD also stores expression data on the tissue or cell-line source of mouse cDNA clones, data mainly acquired via large data downloads. We have made strong progress in this area. In collaboration with the other members of the FANTOM consortium, including our colleagues from the MGD, we incorporated all cDNA data derived from the FANTOM project into GXD (38–40). We have also incorporated all other publicly available mouse cDNA and EST data, including cDNA data from the I.M.A.G.E consortium, the sets from the National Institute of Aging (NIA) (41,42) and the Mammalian Gene Collection (MGC) (43,44). We have loaded each clone's IDs and sequence information, including ESTs and any longer sequences that are known. Clone records have been associated with genes via manual curation and computational analysis of sequence associations. Likewise, using coordinated automatic and manual processes, we have mapped the source information for each cDNA library (such as information about strain, tissue, cell line and sex) to our controlled vocabularies. Thus, all the cDNA source data are now recorded in standardized form. This annotation and integration of all these data allows users to do comprehensive, expression-related queries based on cDNA source information. Currently GXD contains data for >1.7 million clones for nearly 28 000 genetic markers.

The Gene Expression Notebook

GXD's usefulness is directly proportional to the amount of data it contains, but extracting data from the literature is a time-intensive process. Therefore, to facilitate direct submissions of expression data from the research community, we have developed the Gene Expression Notebook (GEN).

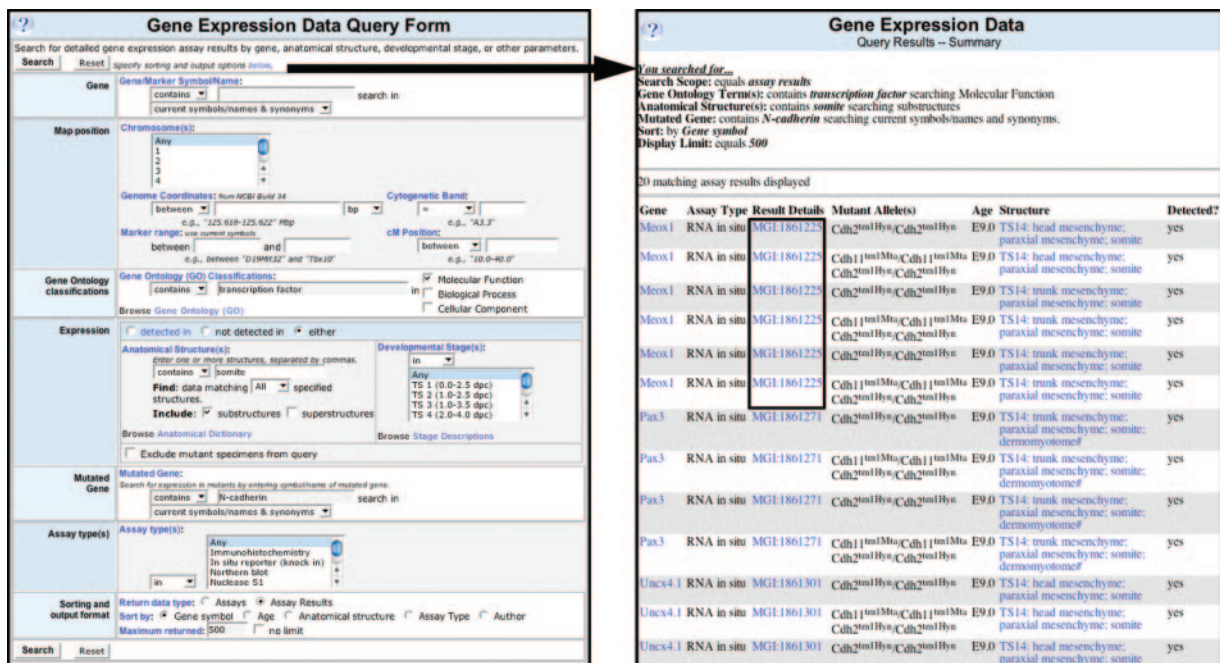


Figure 3. Querying for detailed expression data. The standard Gene Expression Data query form (left) allows users to query expression data using the parameters that are of most interest to them. Searchable fields include gene symbol/name (and synonyms); map position, anatomical structure, developmental stage and assay type. The Gene Ontology (GO) Classifications field allows users to search for genes that have specific molecular functions, are involved in particular biological processes or associated with specific subcellular components. Anatomical structures are named according to a controlled vocabulary system. Due to the hierarchical structure of this vocabulary, spatial queries can include anatomical substructures or superstructures. The Mutated Gene field allows querying for expression data in specific mutants. The query formulated in the figure uses three search parameters to ask for transcription factors expressed in the somites of N-cadherin (Cdh2) mutants. The results summary (right) lists the symbol of the gene examined, the assay type used, result details (given as an MGI accession ID linked to an assay record), allele pairs describing the mutant genotype of the specimen, age of the specimen, anatomical structure examined and whether or not expression was detected. Clicking on the boxed accession IDs will yield the detailed expression record pictured in Figure 1.

GEN is an Excel-based application designed to function as an electronic notebook to store and organize expression data and images in the laboratory. Researchers can then, with minimal additional effort, select data stored in the GEN and submit it to GXD. Data submissions are reviewed by GXD curators and receive accession numbers that can be cited in publications. GEN has been described in detail previously (45). Since then we have developed new versions of the GEN to be compatible with recent versions of Excel: 2000, 2003 and X. GEN is freely available for download at: <http://www.informatics.jax.org/mgihome/GXD/GEN/>.

ACCESSING GXD'S EXPRESSION DATA

Users primarily access GXD using web-based query forms. A complete listing of our query forms is provided in the supplementary material. The standard (Figure 3) and expanded Gene Expression Data query forms provide access to the detailed expression data. Both forms provide users with a wide variety of query fields to allow them to extract the data from the database that is of the most interest to them. The expanded GXD query form is designed to allow users to query for genes expressed in some anatomical structures and/or developmental stages but not others. The functionality of the map position section on these forms (as well as on the cDNA clone query form) was expanded to allow searching by genome coordinates and by a range of markers. When

combined with the previously existing ability to search by centiMorgan position and cytogenetic band, users have a powerful tool to limit their expression searches to specified chromosomal regions, useful when doing positional cloning studies to hunt for candidate genes. All queries in GXD, including those from the standard and expanded query forms, return data summaries; an example is pictured in Figure 3. Links in the Results Detail column of these summary pages provide access to our detailed data entries; an example of a detailed entry is provided in Figure 1.

GXD is implemented in the Sybase relational database management system. For those users who wish to perform custom queries or analyses not possible through the web-based query forms, direct SQL access or custom database reports can be requested from User Support (contact details supplied below).

MOUSE GENE EXPRESSION INFORMATION RESOURCE

GXD has a longstanding collaboration with the EMAGE project (16). EMAGE provides for 3D graphical storage of mouse developmental *in situ* expression patterns (from wild-type mice). GXD makes all pertinent text-annotated *in situ* data, including primary image data, available to EMAGE so that it can be spatially mapped. EMAGE already provides links from its Anatomical Section Browser

(<http://genex.hgu.mrc.ac.uk/Resources/GXDQuery1/Section-Browser.html>) to the appropriate GXD data summaries. GXD is in the process of adding links to EMAGE. The ultimate goal of the collaboration between our two databases is the creation of the Mouse Gene Expression Information Resource (MGEIR) that will fully combine standardized text-based and graphical means for storing and querying expression data (46).

USER SUPPORT

GXD provides support to its users through detailed on-line documentation and a dedicated User Support staff. The on-line documentation can be accessed by clicking on the question mark found in the upper left hand corner of most web pages. Our User Support personnel can be contacted via email at mgi-help@informatics.jax.org or by clicking User Support on our web pages. They can also be reached by phone at 1-207-288-6445.

CITING GXD

The following citation format is suggested when referring to data from GXD: These data were retrieved from the Gene Expression Database (GXD), Mouse Genome Informatics (MGI), The Jackson Laboratory, Bar Harbor, Maine, USA (URL: <http://www.informatics.jax.org>). [Type in date (month, year) when you retrieved the data cited.] To reference the database itself, please cite this article.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We would like to thank our colleagues from the other MGI projects for their contributions to the GXD project and to the larger, integrated MGI resource. We thank Drs. Carol Bult and Gregory Cox for critically reading the manuscript. GXD is available to the public for free due to funding from NIH grant HD033745. Funding to pay the Open Access publication charges for this article was provided by NIH grant HD033745.

Conflict of interest statement. None declared.

REFERENCES

- Blake, J.A., Eppig, J.T., Bult, C.J., Kadin, J.A., Richardson, J.E. and the Mouse Genome Database Group (2006) The Mouse Genome Database (MGD): updates and enhancements. *Nucleic Acids Res.*, **34**, D562–D567.
- Eppig, J.T., Blake, J.A., Davisson, M.T., Kadin, J.A., Richardson, J.E. and the Mouse Genome Database staff (2000) The Mouse Genome Database: a resource for today and tomorrow. *Lab Anim. (NY)*, **29**, 39–43.
- Gene Ontology Consortium (2006) The Gene Ontology (GO) project in 2006. *Nucleic Acids Res.*, **34**, D322–D326.
- Wheeler, D.L., Barrett, T., Benson, D.A., Bryant, S.H., Canese, K., Chetverin, V., Church, D.M., DiCuccio, M., Edgar, R., Federhen, S. *et al.* (2006) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **34**, D173–D180.
- Benson, D.A., Karsch-Mizrachi, I., Lipman, D.J., Ostell, J. and Wheeler, D.L. (2006) GenBank. *Nucleic Acids Res.*, **34**, D16–D20.
- Birney, E., Andrews, D., Caccamo, M., Chen, Y., Clarke, L., Coates, G., Cox, T., Cunningham, F., Curwen, V., Cutts, T. *et al.* (2006) Ensembl 2006. *Nucleic Acids Res.*, **34**, D556–D561.
- Lee, Y., Tsai, J., Sunkara, S., Karamycheva, S., Pertea, G., Sultana, R., Antonescu, V., Chan, A., Cheung, F. and Quackenbush, J. (2005) The TIGR gene indices: clustering and assembling EST and known genes and integration with eukaryotic genomes. *Nucleic Acids Res.*, **33**, D71–D74.
- Maglott, D., Ostell, J., Pruitt, K.D. and Tatusova, T. (2005) Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res.*, **33**, D54–D58.
- Wu, C.H., Apweiler, R., Bairoch, A., Natale, D.A., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R. *et al.* (2006) The Universal Protein Resource (UniProt): an expanding universe of protein information. *Nucleic Acids Res.*, **34**, D187–D191.
- Mulder, N.J., Apweiler, R., Attwood, T.K., Bairoch, A., Bateman, A., Binns, D., Bradley, P., Bork, P., Bucher, P., Cerutti, L. *et al.* (2005) InterPro, progress and status in 2005. *Nucleic Acids Res.*, **33**, D201–D205.
- Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M., Estreicher, A., Gasteiger, E., Martin, M.J., Michoud, K., O'Donovan, C., Phan, I. *et al.* (2003) The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.*, **33**, 365–370.
- Hamosh, A., Scott, A.F., Amberger, J.S., Bocchini, C.A. and McKusick, V.A. (2005) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res.*, **33**, D514–D517.
- de la Cruz, N., Bromberg, S., Pasko, D., Shimoyama, M., Twigger, S., Chen, J., Chen, C.F., Fan, C., Foote, C., Gopinath, G.R. *et al.* (2005) The Rat Genome Database (RGD): developments towards a phenome database. *Nucleic Acids Res.*, **33**, D485–D491.
- Hu, J., Mungall, C., Law, A., Papworth, R., Nelson, J.P., Brown, A., Simpson, I., Leckie, S., Burt, D.W., Hillyard, A.L. and Archibald, A.L. (2001) The ARKdb: genome databases for farmed and other animals. *Nucleic Acids Res.*, **29**, 106–110.
- Gas, S., Eggen, A., Samson, F., Christophe, C., Mungall, C., Bessières, P. and Levéziel, H. (1996) The BOVMAP database. XXVth International Conference on Animal Genetics, July 21–25, 1996, Tours, France. *Anim. Genet.*, **27** (Suppl. 2), 56.
- 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.
- 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.
- Heintz, N. (2004) Gene expression nervous system atlas (GENSTAT). *Nature Neurosci.*, **7**, 483.
- Magdaleno, S., Jensen, P., Brumwell, C.L., Seal, A., Lehman, K., Asbury, A., Cheung, T., Cornelius, T., Batten, D.M. *et al.* (2006) BGEM: an *in situ* hybridization database of gene expression in the embryonic and adult mouse nervous system. *PLoS Biol.*, **4**, E86.
- Zwingman, T.A., Thompson, C.L., Mortrud, M.T., Dong, H.W., Jones, A.R., Hawrylycz, M. and Lein, E.S. (2005) High-throughput *in situ* hybridization in adult mouse reveals novel cell-specific markers and new cerebellar compartments. *Neurosci. Abst.*, **31**, 145.13.
- Barrett, T., Suzek, T.O., Troup, D.B., Wilhite, S.E., Wing-Chi, N., Ledoux, P., Rudnev, D., Lash, A.E., Fujibuchi, W. and Edgar, R. (2005) NCBI GEO: mining millions of expression profiles—database and tools. *Nucleic Acids Res.*, **33**, D562–D566.
- Parkinson, H., Sarkans, U., Shojatalab, M., Abeygunawardena, N., Contrino, S., Coulson, R., Farne, A., Lara, G.G., Holloway, E., Kapushesky, M. *et al.* (2005) ArrayExpress—a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res.*, **33**, D553–D555.
- Su, A.I., Wiltshire, T., Batalov, S., Lapp, H., Ching, K.A., Block, D., Zhang, J., Soden, R., Hayakawa, M., Kreiman, G. *et al.* (2004) A gene atlas of the mouse and human protein-encoding transcriptomes. *Proc. Natl Acad. Sci. USA*, **101**, 6062–6071.
- Ball, C.A., Awad, I.A., Demeter, J., Gollub, J., Hebert, J.M., Hernandez-Boussard, T., Jin, H., Matese, J.C., Nizberg, M., Wymore, F. *et al.* (2005) The Stanford Microarray Database accommodates

- additional microarray platforms and data formats. *Nucleic Acids Res.*, **33**, D580–D582.
25. Boon,K., Osorio,E.C., Greenhut,S.F., Schaefer,C.F., Shoemaker,J., Polyak,K., Morin,P.J., Buetow,K.H., Strausberg,R.L., DeSouza,S.J. *et al.* (2002) An anatomy of normal and malignant gene expression. *Proc. Natl Acad. Sci. USA*, **99**, 11287–11292.
 26. Divina,P. and Forejt,J. (2004) The mouse SAGE site: database of public mouse SAGE libraries. *Nucleic Acids Res.*, **32**, D482–D483.
 27. Ringwald,M., Mangan,M.E., Eppig,J.T., Kadin,J.A., Richardson,J.E. and the Gene Expression Database Group (1999) GXD: a Gene Expression Database for the laboratory mouse. *Nucleic Acids Res.*, **27**, 106–112.
 28. Ringwald,M., Eppig,J.T., Kadin,J.A., Richardson,J.E. and the Gene Expression Database Group (2000) GXD: a Gene Expression Database for the laboratory mouse: current status and recent enhancements. *Nucleic Acids Res.*, **28**, 115–119.
 29. Ringwald,M., Eppig,J.T., Begley,D.A., Corradi,J.P., McCright,I.J., Hayamizu,T.F., Hill,D.P., Kadin,J.A. and Richardson,J.E. (2001) The mouse Gene Expression Database (GXD). *Nucleic Acids Res.*, **29**, 98–101.
 30. Hill,D.P., Begley,D.A., Finger,J.F., Hayamizu,T.F., McCright,I.J., Smith,C.M., Beal,J.S., Corbani,L.E., Blake,J.A., Eppig,J.T. *et al.* (2004) The mouse Gene Expression Database (GXD): updates and enhancements. *Nucleic Acids Res.*, **32**, D568–D571.
 31. 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.
 32. Hayamizu,T.F., Mangan,M., Corradi,J.P., Kadin,J.A. and Ringwald,M. (2005) The Adult Mouse Anatomical Dictionary: a tool for annotating and integrating data. *Genome Biol.*, **6**, R29.
 33. Reymond,A., Marigo,V., Yaylaoglu,M.B., Leoni,A., Ucla,C., Scamuffa,N., Caccioppoli,C., Dermizakis,E.T., Lyle,R., Banfi,S. *et al.* (2002) Human chromosome 21 gene expression atlas in the mouse. *Nature*, **420**, 582–586.
 34. Gitton,Y., Dahmane,N., Baik,S., RuizIAltaba,A., Neidhardt,L., Scholze,M., Herrmann,B.G., Kahlem,P., Benkahla,A., Schrinner,S. *et al.* (2002) A gene expression map of human chromosome 21 orthologs in the mouse. *Nature*, **420**, 586–590.
 35. Gray,P.A., Fu,H., Luo,P., Zhao,Q., Yu,J., Ferrari,A., Tenzen,T., Yuk,D.L., Tsung,E.F., Cai,Z. *et al.* (2004) Mouse brain organization revealed through direct genome-scale TF expression analysis. *Science*, **306**, 2255–2257.
 36. Blackshaw,S., Harpavat,S., Trimarchi,J., Cai,L., Huang,H., Kuo,W.P., Weber,G., Lee,K., Fraioli,R.E., Cho,S. *et al.* (2004) Genomic analysis of mouse retinal development. *PLoS Biol.*, **2**, e247.
 37. McKee,A.E., Minet,E., Stern,C., Riahi,S., Stiles,C.D. and Silver,P.A. (2005) A genome-wide *in situ* hybridization map of RNA-binding proteins reveals anatomically restricted expression in the developing mouse brain. *BMC Dev. Biol.*, **5**, 14.
 38. Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S., Nikaido,I., Osato,N., Saito,R., Suzuki,H. *et al.* (2002) Analysis of the mouse transcriptome based on functional annotation of 60770 full-length cDNAs. *Nature*, **420**, 563–573.
 39. Baldarelli,R.M., Hill,D.P., Blake,J.A., Adachi,J., Furuno,M., Bradt,D., Corbani,L.E., Cousins,S., Frazer,K.S., Qi,D. *et al.* (2003) Connecting sequence and biology in the laboratory mouse. *Genome Res.*, **13**, 1505–1519.
 40. Carninci,P., Kasukawa,T., Katayama,S., Gough,J., Frith,M.C., Maeda,N., Oyama,R., Ravasi,T., Lenhard,B., Wells,C. *et al.* (2005) The transcriptional landscape of the mammalian genome. *Science*, **309**, 1559–1563.
 41. Kargul,G.J., Dudekula,D.B., Qian,Y., Lim,M.K., Jaradat,S.A., Tanaka,T.S., Carter,M.G. and Ko,M.S. (2001) Verification and initial annotation of the NIA mouse 15K cDNA clone set. *Nature Genet.*, **28**, 17–18.
 42. VanBuren,V., Piao,Y., Dudekula,D.B., Qian,Y., Carter,M.G., Martin,P.R., Stagg,C.A., Bassey,U.C., Aiba,K., Hamatani,T. *et al.* (2002) Assembly, verification and initial annotation of the NIA mouse 7.4K cDNA clone set. *Genome Res.*, **12**, 1999–2003.
 43. Strausberg,R.L., Feingold,E.A., Klausner,R.D. and Collins,F.S. (1999) The mammalian gene collection. *Science*, **286**, 455–457.
 44. Gerhard,D.S., Wagner,L., Feingold,E.A., Shenmen,C.M., Grouse,L.H., Schuler,G., Klein,S.L., Old,S., Rasooly,R., Good,P. *et al.* (2004) The status, quality and expansion of the NIH full-length cDNA project: the Mammalian Gene Collection (MGC). *Genome Res.*, **14**, 2121–2127.
 45. Begley,D.A. and Ringwald,M. (2002) Electronic tools to manage gene expression data. *Trends. Genet.*, **18**, 108–110.
 46. Ringwald,M., Baldock,R., Bard,J., Kaufman,M., Eppig,J., Richardson,J.E., Nadeau,J.H. and Davidson,D. (1994) A database for mouse development. *Science*, **265**, 2033–2034.