

Web-based digital gene expression atlases for the mouse

Lars Geffers · Bernhard Herrmann ·
Gregor Eichele

Received: 4 April 2012 / Accepted: 4 July 2012 / Published online: 31 August 2012
© Springer Science+Business Media, LLC 2012

Abstract Over the past 15 years the publicly available mouse gene expression data determined by in situ hybridization have dramatically increased in scope and spatio-temporal resolution. As a consequence of resources and tools available in the post-genomic era, full transcriptomes in the mouse brain and in the mouse embryo can be studied. Here we introduce and discuss seven current databases (MAMEP, EMBRYS, GenePaint, EURExpress, EuReGene, BGEM, and GENSAT) that grant access to large collections of expression data in mouse. We review the experimental focus, coverage, data assessment, and annotation for each of these databases and the implementation of analytic tools and links to other relevant databases. We provide a user-oriented summary of how to interrogate each database.

Introduction

The purpose of this review is to discuss the major web-based digital gene expression atlases for the mouse that present gene expression patterns at or near cellular resolution for a significant number of genes. In addition, we review several databases that provide access to whole-

mount data of gene expression. These atlases typically annotate the sites and the characteristics (strength, pattern) of expression and offer simple or customized search tools initiated with gene names, site of expression, tissue types, gene sequences, or a combination thereof. Digital atlases that meet these criteria are still relatively limited in number because producing them is a time and cost-intense endeavor.

In addition to the digital atlases stocked with data specifically created for a particular atlas, such data resurfaces in integrated resources. Two such community-directed resources are the Gene Expression Database (Finger et al. 2011) and EMAGE (Richardson et al. 2010), which import images and associated metadata from original expression atlases and add on a wealth of information from the literature and other sources (see reviews by Ringwald et al. and Baldock et al. in this issue).

In previous reviews, the potential of large sets of spatially registered gene expression data was discussed with a focus on neuroscience (Sunkin and Hohmann 2007) or on the integration into the context of systems biology (Geschwind and Konopka 2009). An excellent and more general review of existing Gene Expression Databases with a developmental focus has recently been published (de Boer et al. 2009). These authors stressed that “despite the availability of this wealth of information, the number of references to these atlases [...] is limited, indicating an underutilization by developmental biologists.” To some extent this criticism is mitigated by the fact that, for example, the GenePaint database gets several thousand queries each month and the EURExpress publication so far has had more than 30,000 views (as of spring 2012). Nonetheless, there is a need for addressing the content and utilization of mouse digital atlases with the hope that in the future the mouse research community will take even better

L. Geffers · G. Eichele (✉)
Department of Genes and Behavior, Max Planck Institute for
Biophysical Chemistry, Am Faßberg 11,
37077 Göttingen, Germany
e-mail: Gregor.eichele@mpibpc.mpg.de

L. Geffers
e-mail: lars.geffers@mpibpc.mpg.de

B. Herrmann
Department of Developmental Genetics, Max Planck Institute
for Molecular Genetics, Ihnestr. 63-73, 14195 Berlin, Germany

Table 1 Name, data type, and content of public digital gene expression atlases for the mouse

Name Method	MAMEP Whole-mount ISH	EMBRYS	GenePaint Nonradioactive ISH	EURExpress	EuReGene	ABA	BGEM Radioactive ISH	GENSAT EGFP reporter
Strain	NMRI	ICR	NMRI, C57BL/6	C57BL/6	C57BL/6	C57BL/6	C57BL/6	FVB/N
Embryo	X	X	X	X		X		
Kidney					X			
CNS			X			X	X	X
E8.5	X							
E9.5	X	X						
E10.5	X	X	X					X
E11.5	X	X					X	
E13.5						X		
E14.5			X	X		X		
E15.5			X			X	X	X
E17.5					X			
E18.5						X		
P4						X		
P7			X				X	X
P14						X		
Adult			X		X	X	X	X
Unique genes	1,912	1,561	16,856	17,395	770	~20,000	3,430	2,311
Images	23,994	24,545	354,357	284,611	~5,500	>>400,000	~35,000	68,185
Datasets ^a	1,912	1,561	19,166	19,440	1,540	>>20,000	3,430	>>2,311
Annotated	All	All	3,062	15,775	1,222	All	1,510	1,184
Anat. Terms	516	30	96	1,420	688	103	25	74
Annotated expression level		0	0	Not detected	Not detected	0	–	Undetectable
	Possible			Possible	Possible	1		
	Weak	1	1	Weak	Weak	2	+	Weak
	Moderate		2	Moderate	Moderate	3		
	Strong	2	3	Strong	Strong	4	++	Moderate to strong
Annotated pattern	Homogeneous		U	Homogeneous	Homogeneous		WE	WE
	Regional		R	Regional	Regional		RS	RS
	Graded			Graded	Graded			
	Spotted		S	Spotted	Spotted		S	S
				Single cell	Single cell			
				Other	Other		RSS	RSS

NMRI Naval Medical Research Institute albino outbred strain, *ICR* Institute for Cancer Research albino outbred strain, *C57 BL/6* black coated inbred strain, *FVB/N* albino inbred strain, *anat.* anatomical, *R* regional, *RS* region-specific, *RSS* region-specific scattered, *S* scattered, *U* ubiquitous, *WE* widely expressed, *ABA* Allen Brain Atlas

^a The term dataset refers to a set of images. In some databases a set is derived from serial sections of a single specimen (e.g., GenePaint), while in other databases a dataset consists of multiple stages that are assembled into a single panel (e.g., BGEM)

advantage of these marvelous resources. For this reason we describe the most comprehensive databases, MAMEP, EMBRYS, GenePaint, EURExpress, EuReGene, BGEM, and GENSAT, all containing original mouse expression data generated specifically for these databases. In this review we go through each database and provide a synopsis

of data mining and data retrieval. The various databases differ significantly in the number of genes included, the coverage of stages, the method used to visualize mRNAs, and the completeness and depth of annotation (Table 1). Currently, the most complete resource is the Allen Brain Atlas (ABA; Lein et al. 2007), which is described in a

separate section of this issue. We apologize that several very valuable gene expression analyses providing digital ISH data, such as the nuclear receptor study by Gofflot et al. (2007), the genomic analysis of mouse retinal development by Blackshaw et al. (2004), or the genome-scale transcription factor analysis by Gray et al. (2004), could not be covered in this review.

MAMEP (<http://mamep.molgen.mpg.de>)

The Molecular Anatomy of the Mouse Embryo Project (MAMEP) was developed at the Max Planck Institute for Molecular Genetics in Berlin, Germany. The objective of this project was to create a resource for a directed functional analysis of developmental processes by identifying genes that are likely to play an important role in the regulatory networks controlling pattern formation and organogenesis. Accordingly, whole-mount *in situ* hybridization (WISH) was utilized in a genome-wide screen to identify genes showing a restricted expression pattern in the E9.5 embryo. The MAMEP database features around 1,900 genes fulfilling this criterion. Genes showing remarkable patterns were also analyzed at E8.5, E10.5, and E11.5. The expression patterns are presented on more than 20,000 photographs and displayed with respect to the cDNA clone(s) used for WISH analysis, since different clones may produce various results for the same gene.

The database can be queried for genes using Gene Symbol, Clone Name, or Ensemble ID. Results can be restricted to those associated with images and/or annotation. In an advanced mode the user can query for genes expressed in an anatomical structure, which can be selected from an anatomical tree based on the EMAP Theiler Stage (TS) 15 (E9.5) ontology of anatomical structures (<http://www.emouseatlas.org/Databases/Anatomy/new/theiler15.shtml>). Furthermore, a BLAST search is implemented to allow matching of a user-defined nucleotide sequence against all template sequences used in the MAMEP project.

Any result will be displayed as a list of panels (A in Fig. 1), where each panel is composed of a header with Gene Symbol, Ensemble ID, Gene Name, and an internal Clone ID, a strip of corresponding thumbnails of photographs from embryos analyzed by WISH, and a description of the expression detected in various anatomical structures at TS 15. A link to review relevant template sequences is given at the bottom of each panel (B in Fig. 1). Annotation is provided for individual structures and is color-coded based on the strength of the observed signal (possible, weak, moderate, or strong). In addition to signal strength, the pattern quality of a signal (homogeneous, regional, graded, or spotted) is recorded. Thumbnails can be opened in a separate window in order to view the image at a

larger size. The user can cycle through three levels of magnification by repeatedly clicking on the image, and the image can be saved locally (C in Fig. 1).

While the complexity of the web interface is kept at a minimum, there is little distraction from the original purpose of the database: access to high-quality WISH images of a varying number of stages and experiments per gene. Annotation of gene expression is detailed and instrumental for data mining as it helps identify genes acting in the same anatomical structures.

Data from MAMEP was used in a recent study (Qi et al. 2011) in which genome-wide expression profiling of the urinary bladder implicated desmosomal and cytoskeletal dysregulation in the exstrophy-epispadias complex, a spectrum of urological abnormalities where part or all of the distal urinary tract fails to close during development. Expression patterns of candidate genes were examined in the MAMEP database in order to assess their functional relevance during early development of the bladder primordium and its associated tissues.

EMBRYS (<http://embrys.jp>)

Since 2008 the Japanese National Research Institute for Child Health and Development has maintained a well-curated database of WISH data from mouse with a strict focus on transcription factors (TF) and TF-related proteins studied in midgestational stages (9.5, 10.5, and 11.5 dpc). Thus, EMBRYS (Yokoyama et al. 2009) is clearly following in the footsteps of MAMEP. The scope and experimental quality of the WISH data are largely comparable to those of MAMEP, and with regard to TFs some overlap in content is expected. However, because MAMEP also features genes other than TFs and their cofactors and EMBRYS provides more complete experimental data for embryonic stages E10.5 and E11.5 than MAMEP, both databases mostly complement and extend each other. What makes EMBRYS an excellent resource on the web is its straightforward design, dedicated focus, and ease of use. The comprehensive documentation makes all processes around the datasets very transparent and describes the experimental approaches, the organization of the production pipeline, and all aspects of data assessment (annotation).

The query interface is “no frills” and accepts only NCBI Gene Symbol (including alternatives) or Gene ID. Each query results in a list of candidate genes that match the query string fully or in part. Once a candidate has been selected from the list, the dataset of the gene appears in a viewer panel (A in Fig. 2). This viewer features a strip of thumbnails, links to MGI and NCBI, a larger image of a selected thumbnail, and the sequences of the primers used for template generation. The image can be zoomed in and out using the computer

mouse wheel and can be saved from its context menu (right click). All thumbnails and images are labeled to indicate the developmental stage of the specimen shown and the microscope setting (zoom factor). The saved filenames contain the Gene Symbol, the embryonic age (dpc), and the zoom factor. Datasets usually contain images at three zoom levels (10× or 12×, 20× or 24×, and 40× or 48×) for each of the three stages (E9.5, E10.5, and E11.5); however, occasionally the datasets also include additional images and stages.

A description of gene expression is available for all datasets. However, this information is not part of the dataset itself but is included in a freely downloadable spreadsheet (Microsoft Excel). All genes covered by the database are listed in this spreadsheet and gene expression is annotated using three criteria: (1) gene is expressed at any of the embryonic stages examined (yes/no), (2) statement (yes/no) in which of the three stages the gene in question is expressed, and (3) description of expression in each of 30 structures rating the relative signal intensity (0, 1, or 2). Special attention was paid to limb bud development where annotation is

more detailed. The spreadsheet allows the user global access to the annotation data, and using standard Microsoft Office tools one can further organize and shortlist genes of interest.

Typical applications of the data deposited in EMBRYOS include a systems biology approach to reveal a novel regulatory factor in the myogenesis genomic network (Yokoyama et al. 2009). Genes with spatial colocalization in myogenesis-related structures were identified according to the gene expression annotation provided by EMBRYOS. Among the few shortlisted genes with yet uncharacterized function in skeletal myogenesis, a candidate gene (RP58) was identified and further characterized as a MyoD-activated factor repressing myogenic inhibitors.

GenePaint (<http://www.genepaint.org>)

For more than a decade nonradioactive ISH data have been generated systematically from fresh frozen sections of a variety of murine specimens in a highly automated process

Fig. 1 Screenshot composite of the MAMEP database containing WISH data of midgestation mouse embryos. A query results in a list of corresponding gene entries with thumbnails and annotation of gene expression (*panel A*). Following the sequence information link provides access to template sequence (*panel B*). Thumbnails in panel A link to high-power view of the specimen (*panel C*)

The figure is a composite screenshot of the MAMEP database interface. At the top, the 'mamep database' logo is visible, along with navigation links for Home, Search, News, and Contact. A search bar contains the text 'trh'. Below the search bar, there are options for 'with image' and 'with description'. The search results display the gene 'Trh' with its full name and a brief description: 'ENSMUSG00000005892 Thyroliberin precursor [Contains: Prothyroliberin; Thyroliberin (Thyrotropin-releasing hormone) (TRH) (Thyrotropin-releasing factor) (TRF) (TSH-releasing factor) (Protirelin)]. [Source: Uniprot/SWISSPROT, Acc: Q62361]'. Below the gene information, there is a grid of thumbnails showing mouse embryos at different stages. One thumbnail is circled in red and labeled 'C'. A red arrow points from this thumbnail to a larger, high-power view of the specimen, also labeled 'C'. Below the thumbnails, there is a section for 'sequence information' which is circled in red and labeled 'B'. This section contains a sequence alignment for 'UK700' with a sequence: '>UK700 12248 TTTTGCCTGGATATATTTTATTGTC CTGAAATATCCCCAAACTGCTTAGGG GGCAGATGACAGAAGTTGGATTTTAA AGAACGTCTTGGCCAGTGTGGAAGG GGGATAGTAGG'. To the right of the sequence, there is a vertical column of 'onal' labels. At the bottom right, there is a 'save image' button and a small icon.

Fig. 2 Screenshot of the EMBRY database, one of the most comprehensive resources of mouse embryonic WISH data for transcription factors (TF) and TF-related factors. A single compact window (*panel A*) contains the query field (*top left*), list of results (*bottom left*), and details of a selected result such as thumbnails (*top*), a high-power view (*center*), and template primer sequences (*bottom*)

A

EMBRYS

19 pictures

EP. 5 x12.0 EP. 5 x24.0 EP. 5 x40.0 EP. 5 x48.0 E10.5 x12.0 E10.5 x12.0 E10.5 x15.0 E10.5 x24.0 E10.5 x38.0 E10.5 x40.0

[21385] Tbx2 : E10.5 x12.0

MGI NCBI

2D Images Search

Symbol GeneID

tbx

Search Clear

Candidates:

- TBX1
- Tbx10
- Tbx12
- Tbx13
- Tbx14
- Tbx15
- Tbx18
- Tbx19
- Tbx2**
- Tbx20
- Tbx21
- Tbx22
- Tbx3
- Tbx4
- Tbx5
- Tbx7
- Tbx8

Primer	Sequence	Length(nt)
Forward	GTAATACGACTCACTATAGGGTCTGAACTCCATGCACAAG	41
Reverse	CATTAGGTGACACTATAGCTGTCTGTCTGCACCACCAG	39

based on the GenePaint platform (Carson et al. 2002; Visel et al. 2004; Yaylaoglu et al. 2005). Almost since its inception the GenePaint database has served as an instant and web-accessible repository in which the robotics-generated expression data were uploaded shortly after their production. Now located at the MPI for Biophysical Chemistry in Göttingen, Germany, the GenePaint database offers a nearly complete genome-wide scope of gene expression patterns of the E14.5 mouse embryo (including data from the EURExpress project, see below) and, to a much lesser degree, also some datasets of the E10.5 embryo, the E15.5 head, and the P7 (Carson et al. 2005) and P56 brain. The digital images in the datasets allow visualization at near-cellular resolution. More than 3,000 E14.5 embryo datasets in GenePaint were manually annotated using a hierarchical anatomy tree with 96 unique terms. Annotated datasets include transcription factors, signaling components, transporters, and enzymes. Gene expression is judged for its relative signal intensity (0, 1, 2, 3) and for its local pattern (ubiquitous, regional, scattered), both with respect to the structure in question.

The start page of the web interface allows access to reference material such as anatomical maps, data production methods, and a description of and entry into advanced search strategies. It provides a search field to directly query

the database by NCBI Gene Symbol, Gene Name, Gene ID, RefSeq, or GenePaint set ID (database's internal ID). A more customized search can be started by clicking the "Advanced Search" button. A very useful feature is the pattern search whereby one selects any combination of structures from the anatomical tree and assigns criteria for expression intensity and local pattern. One can also blast a user-provided sequence against all template sequences in the database. Each of these query modes can be further restricted by filters for stage, strain, tissue, and annotation state.

Once a query is submitted, matching results appear in a table in which annotated datasets are marked by a star. A link to the actual dataset (Set Viewer) or a thumbnail preview (Thumbnail) is provided. The thumbnails are tagged with a direct link that leads to the Set Viewer (A in Fig. 3). The Set Viewer window itself is divided in four sectors. The upper-left sector displays gene information and experimental metadata. It has links to the NCBI RefSeq ("GenBank," red button) and NCBI Gene ID ("Gene," blue button). The RNA probe number link leads to the sequence of the template and its primers (B in Fig. 3). The upper-right sector harbors a thumbnail browser to quickly leaf through all images of the dataset using the forward and reverse double arrows. Reference maps for the embryo can

Panel A: GenePage

GenePaint Set ID: MH1862
 Gene: Slc2a1: solute carrier family 2 (facilitated glucose transporter), member 1
 Description: solute carrier family 2 (facilitated glucose transporter), member 1; Glut1; Glut-1
 Accession number: GenBank: NM_011400
 Entrez Gene ID: Gene: 20525
 RNA probe: 1523
 Orientation: Antisense
 Hybridization Temp.: 64.0
 Washing Temp.: 62.0
 Stage: E14.5
 Strain: NMRI
 Sectioning Plane: sagittal
 Resolution: 1.6 µm/pixel
 Date of Dissection: 30. Aug 2004
 Time of Dissection: 09:40

Panel B: Probe Details

Riboprobe number: 1523
 Gene: Slc2a1: solute carrier family 2 (facilitated glucose transporter), member 1
 Accession number: NM_011400
 Entrez Gene ID: 20525
 Gene Description: solute carrier family 2 (facilitated glucose transporter), member 1; Glut1; Glut-1
 Orientation of RNA Probe: Antisense
 Template Sequence (Coding Strand):
 GCTGCCTTGGATGCTCTATCTGAGCACTGGCCACTTTGGCTTGTGGCCCTTTTG
 AAGTAGGCCCTGGTCTATTCATGGTTCATTGTGGCCGACTGTTCAAGCCAGGGGC
 CCGCTCCTGGTATTGCTGGTGGCTTCCCACTGGACCTCAAACCTTCATTGG
 GGCATGCTTCCAGTATGGGAACACTGGGGCCCTCATCTTCATCTCA
 CGGTGCTCTCGTCTCTTCTCATCTCACTTCAAGTCCCTGAGACCAAGG
 CCAGAACTCGATGAGTATGCTTCCGGCTCCCGACAGGGGTCCAGCCAAAGT
 ACGAGGAGGCTTCCACCTCTGGGGCCGACTCCCAAGTGTGGAG
 CCAGCCGGGCTGCTCCCTCGAGCCAAAGGATCTCTGGAGGAGA
 GGCATGAGACTCTTCGACCGACAGATCTGGGCAAGCCGGGCTGGG
 CCGCTCAGACTGATGAGTCCAGGAGGATTCAGGACTTGTATGGCTC
 AGAATTTAATGAAAGCAAGACTGCTGCTCAGATCTTTCAGATAAGGAGGATTT
 ATAATTTTTTATGACTGTTGTATTTTTTTTTTATCAGCCACTCTCTCCACCT
 GTAGTCTTCCACTGATTGGCCAGTGCCTGAGGGTGGGACACGCCCTGCCAG
 CACTTGCCTTTGCCAAGCTACTGTAGGGCTGGACCTATGGCCAGGACACACT
 AATACCAACTCTGAGCTGAGGGTTTACCTGGAGGGCTTACTCCACCCACTT
 CCGCAGGCTGACCTGGCCACCATTAGGGGTCCGAGACTCATTTAGGATGCCCA
 TTCTGTCTTCTTCCACCAACCACTAATATCTTCTGCTGCTGAGACCAG

Forward Primer: GCTGCCTTGGATGCTCTATC
 Reverse Primer: CTGGTCTCAGGCAAGAAAG
 TemplateID: MH_OR3

Panel C: Virtual Microscope

Embryo_C1866_1_1A
 HTML Applet Maps

Panel D: Annotation

3r Nervous System (and sense organs) see(3B)
 3r Central Nervous System see(3B)
 3r Telencephalon see(4D)
 3r Cerebral Cortex see(4D)
 3r Neocortex see(4D)
 3r Olfactory Cortex see(4D)
 3r Olfactory Bulb see(3B)
 3r Piriform Cortex see(3B)
 3r Hippocampus see(3A)
 3r Basal Ganglia see(2C)
 3r Corpus Striatum (Lateral Ganglionic Eminence) see(2C)
 3r Basal Forebrain see(2A)
 3r Amygdala see(1C)
 3r Septum see(3B)
 3r Pallidum (Medial Ganglionic Eminence) see(2D)
 3r Diencephalon see(2D)
 3r Thalamus see(2D)
 3r Hypothalamus see(2D)
 3r Mesencephalon (Midbrain) see(3A)
 3r Pretectum see(3A)
 3r Superior Colliculus see(3A)
 3r Inferior Colliculus see(3A)
 3r Tegmentum see(3B)
 3r Rhombencephalon (Hindbrain) see(3A)
 3r Cerebellum see(3A)
 3r Pons see(3A)
 3r Medulla Oblongata see(3B)
 3r Spinal Cord see(4A)
 3r Other see(2D)

Image Directory

in Viewlist	Label	Position
<input checked="" type="checkbox"/>	Embryo_C1866_1_1A	0.000
<input checked="" type="checkbox"/>	Embryo_C1866_1_1B	0.150
<input checked="" type="checkbox"/>	Embryo_C1866_1_1C	0.300
<input checked="" type="checkbox"/>	Embryo_C1866_1_1D	0.450
<input checked="" type="checkbox"/>	Embryo_C1866_1_2A	0.600
<input checked="" type="checkbox"/>	Embryo_C1866_1_2B	0.750
<input checked="" type="checkbox"/>	Embryo_C1866_1_2C	0.900
<input checked="" type="checkbox"/>	Embryo_C1866_1_2D	1.050
<input checked="" type="checkbox"/>	Embryo_C1866_1_3A	1.200
<input checked="" type="checkbox"/>	Embryo_C1866_1_3B	1.350
<input checked="" type="checkbox"/>	Embryo_C1866_1_3C	1.500
<input checked="" type="checkbox"/>	Embryo_C1866_1_3D	1.650
<input checked="" type="checkbox"/>	Embryo_C1866_1_4A	1.800
<input checked="" type="checkbox"/>	Embryo_C1866_1_4B	1.950
<input checked="" type="checkbox"/>	Embryo_C1866_1_4C	2.100
<input checked="" type="checkbox"/>	Embryo_C1866_1_4D	2.250
<input checked="" type="checkbox"/>	Embryo_C1866_1_5A	2.400
<input checked="" type="checkbox"/>	Embryo_C1866_1_5B	2.550
<input checked="" type="checkbox"/>	Embryo_C1866_1_5C	2.700
<input checked="" type="checkbox"/>	Embryo_C1866_1_5D	2.850
<input checked="" type="checkbox"/>	Embryo_C1866_1_6A	3.000
<input checked="" type="checkbox"/>	Embryo_C1866_1_6B	3.150
<input checked="" type="checkbox"/>	Embryo_C1866_1_6C	3.300
<input checked="" type="checkbox"/>	Embryo_C1866_1_6D	3.450

Fig. 3 Screenshot composite of the GenePaint database. GenePaint features mainly colorimetric ISH data of the E14.5 mouse embryo but also contains additional tissues and stages. The Set Viewer (*panel A*) is the default display for an individual dataset. It provides basic gene information and experimental metadata and links to probe and

template details (*panel B*). The Set Viewer contains a thumbnail browser and links to a virtual microscope (*panel C*). Annotation of gene expression is displayed in a collapsible hierarchical anatomy tree (*field D*) within the Set Viewer and features color-coded symbols and annotation scores for 96 anatomical terms

be retrieved by pressing the “Maps” button. To access the experimental data at native resolution, either the Zoom Viewer (“Applet,” requires Java Platform; C in Fig. 3) or the HTML Viewer (“HTML”) can be launched by clicking the appropriate button underneath the thumbnail. Both viewers are virtual microscopes that permit the user to zoom in and pan the high-resolution image shown in the thumbnail. Pressing the “info” button within the HTML Viewer allows conversion of the complete image into a high-resolution JPG. This image can then be saved locally using the context menu (right click). The lower-left sector displays annotation using color- and shape-coded symbols tied to the hierarchical anatomical tree (D in Fig. 3). The symbols are defined in the Annotation of Expression document accessed from the main page. Each annotated structure in the anatomical tree links to a particular section of the dataset that represents this structure optimally with regard to gene expression. The lower-right panel gives a

list of all available images in a dataset and allows instant selection of a section of interest.

The most important feature of the GenePaint project is the systematic large-scale production of ISH data from highly reproducible sets of high-quality sections with almost gapless coverage of all organs and tissues in a specimen at near-cellular resolution. This dense information is a prerequisite for efficient atlas building. The GenePaint data production technology has been adapted by several atlas-building projects (Allen Brain Atlas, EUR-Express, and EuReGene).

GenePaint resources were used in a regulatory pathway analysis based on hierarchical clustering of 1,000 text-annotated expression patterns (Visel et al. 2007). Clustering yielded 12 distinct groups of genes with high expression pattern similarity (synexpression groups). It was suggested that members within such a group might constitute components of a common regulatory cascade.

Expression of genes from the group containing the Pax6 transcription factor was studied in a Pax6-deficient mouse. Of the 17 genes differentially expressed in these mice, 12 genes could be identified as novel potential targets of Pax6 regulation.

EURExpress (<http://www.eurexpress.org>)

In 2005 the EURExpress consortium began its work on a high-resolution anatomical atlas of the transcriptome in the E14.5 (TS 23) mouse embryo (Diez-Roux et al. 2011) based on standard operating procedures that had already been developed for the GenePaint platform. The project was completed in 2010. By that time more than 80 % of the produced datasets had been manually annotated in great detail by experts using the revised Edinburgh Mouse Atlas Project (EMAP) anatomy ontology (<http://www.emouseatlas.org/Databases/Anatomy/new/theiler23.shtml>). A EURExpress web-accessible database showing the results of the project was developed at the MRC Human Genetics Unit in Edinburgh, UK. Similar to the GenePaint database, EURExpress contains tools for looking at expression patterns and expression annotation. Detailed standard operating procedures for the data production pipeline can be downloaded from this site.

In contrast to GenePaint where a particular structure (e.g., heart or lung) is annotated, EURExpress annotates each individual image. All identified anatomical structures in a particular image are scored in terms of signal strength (strong, moderate, weak, possible, or not detected) and pattern (homogeneous, ubiquitous, regional, spotted, or single cell).

Upon opening the EURExpress web interface, the user can click on thumbnails of examples to open the particular dataset to which the thumbnail belongs. This feature allows “curiosity-driven” exploration and learning by use. A more systematic search of EURExpress is initiated by entering a term into the Quick Search field (top right) that is matched against NCBI Gene Symbol, Gene Name, MGI ID, Gene ID, Reference Sequence ID, internal Template ID, internal Assay ID, or an Annotation Summary. Alternatively, the user can choose from five advanced query modes indicated on the right side of the home page: (1) browse and query by keyword, (2) browse annotation by anatomy, (3) browse and query by sequence, (4) advanced search interface, and (5) BioMart interface. The keyword search is identical to the quick search except that it offers more filter options to restrict resulting matches. Searching annotation by anatomy is an interactive mode where the user can browse the hierarchical anatomy tree. Once a structure is selected, all genes

with annotated expression are listed. The sequence search will match the query string against all template sequences in the database and list the associated genes. BioMart (Kasprzyk 2011) is an advanced query platform that is highly customizable and allows a query across several databases (A in Fig. 4). It provides exclusive access to all textual data items in EURExpress. Results from a BioMart query can be exported in various formats making it an indispensable tool for efficient data mining.

Except for queries submitted through the BioMart interface, the standard output is a list of genes matching the query. The results page is a table that contains template- and gene-associated links, a representative thumbnail, an annotation summary, an expression/coverage levels diagram, and links to the Zoom Viewer and Montage Viewer. When clicking either of these links, the corresponding viewer opens. Both the Zoom Viewer and the Montage Viewer are headed by a tabbed information panel providing assay and gene information as well as an expression summary and links to other databases of similar orientation. While the Montage Viewer displays an overview of all the thumbnails of a dataset and provides direct access to the text annotation of each individual image, the Zoom Viewer is a highly interactive virtual microscope that is useful for inspecting images at their native, near-cellular resolution. Zoom Viewer images can be zoomed and panned, and navigation tools help to shift between sections without losing the current zoom factor and centering on the area of interest (B in Fig. 4). In addition, the Zoom Viewer provides an overview of the complete anatomical tree with corresponding annotated gene expression highlighted (C in Fig. 4). Annotated structures are linked to corresponding images of the dataset featuring the expression. Individual images (including their text annotation) can be either downloaded in the Zoom Viewer via the context menu (right click) or in the Montage Viewer by pressing the arrow button next to each thumbnail. Of note is that a selected annotated dataset is automatically matched with the expression profile of all other genes in the database. A list of similarly expressed genes (D in Fig. 4) can be retrieved directly from a designated tab (Similar Assays) located in the information panel. Template sequences and associated information can be reviewed following the Template ID link in the Assay Information tab. Template distribution on chromosomes can be visualized in the UCSC Genome browser following the instructions found under Tools and Analysis from the top menu.

The EURExpress database provides numerous high-quality ISH data of novel and established marker genes that are actively consulted as references for the anatomical organization of the developing brain. A recent review on the regionalization of the diencephalon (Martinez-Ferre and Martinez 2012) illustrates such use.

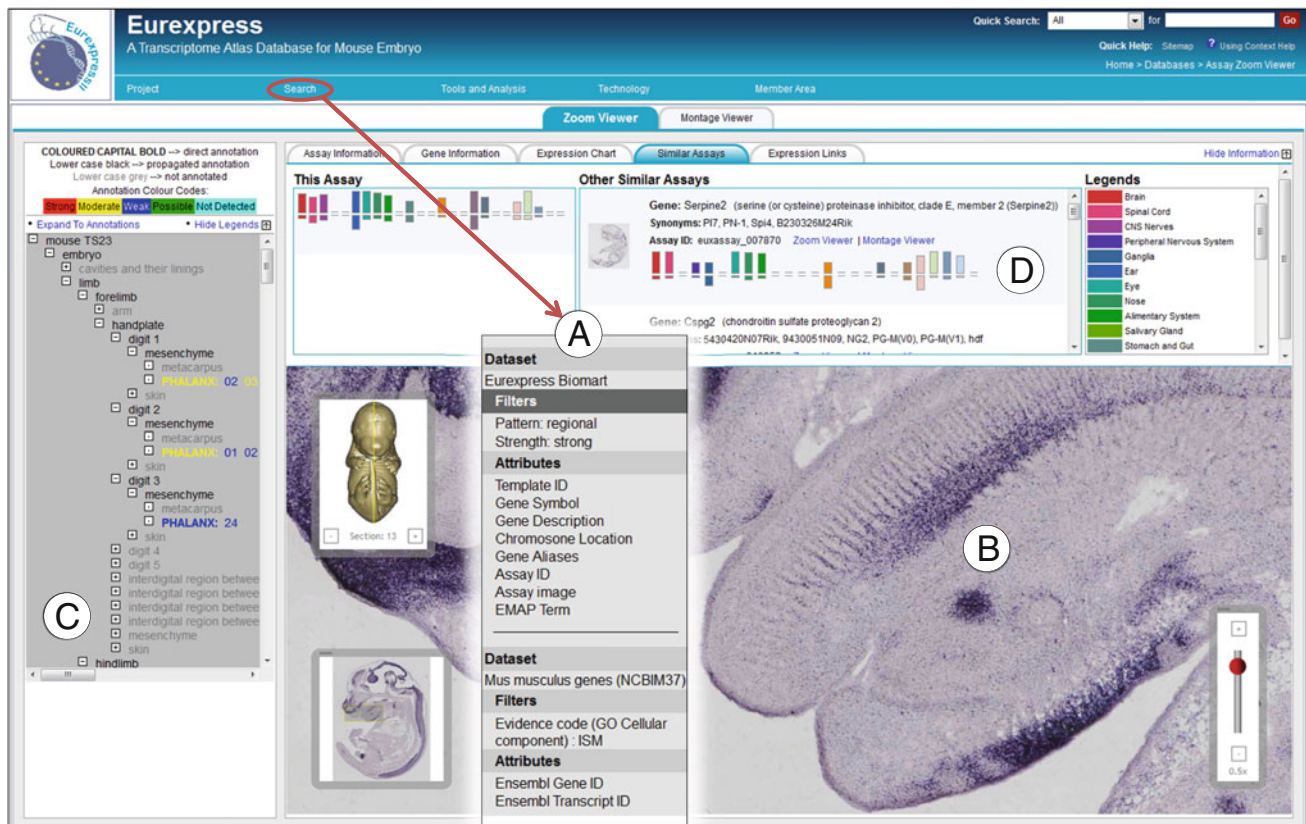


Fig. 4 Screenshot composite of the EURExpress database, a transcriptome-wide gene expression study in the E14.5 embryo. A powerful tool of the EURExpress web interface is the BioMart integration (*panel A*) that allows complex queries and access to all textual information within EURExpress. Through BioMart the user can add external datasets to the query, e.g., the Ensemble *Mus*

musculus genes (NCBIM37). The Zoom Viewer allows browsing and navigation of the digital images of an individual EURExpress dataset (*field B*) and displays the corresponding annotation of gene expression in a collapsible hierarchical anatomy tree (*field C*). EURExpress offers for each annotated dataset a list of datasets with similar expression profile (*field D*)

EuReGene (<http://www.euregene.org/portal>)

The atlas project within the European Renal Genome Project (EuReGene, 2005–2009) was aimed at the generation of the Gene Expression Database of E17.5 and P84 mouse kidney. The resulting database was developed and is maintained by the MRC Human Genetics Unit in Edinburgh. Among the gene families extensively covered by this database are the aquaporins, solute carriers, TRP (transient receptor potential) channels, claudins, carbonic anhydrases, fibroblast growth factors, basic helix-loop-helix transcription factors, GATA transcription factors, and WNT-signaling components as well as a large selection of G protein-coupled receptors and kinases. All these data are not currently publicly available as only a fraction has so far been annotated for sites and intensity of expression.

To detect transcripts, standard GenePaint procedures were applied to representative sections from paraffin-embedded embryonic and adult kidney tissues. Gene expression in kidney was elucidated using paraffin instead of frozen sections as cell morphology was critical for

proper annotation of the segments of the renal tubules. Annotation of gene expression was mapped to an EMAP anatomy ontology that incorporates additional kidney-specific terms. Expression strengths and patterns were annotated as for EURExpress. Upon entering the web interface of the EuReGene database, the user can query for specific Gene Symbols, Gene Names, MGI IDs, or Ensemble Gene IDs either individually or as a batch query. Alternatively, all public data can be retrieved at once by selecting “Browse in situ data.” The generally very useful option of viewing gene expression in anatomical structures of interest can be accessed three ways: (1) By following the link “Search [...] in anatomical structures,” the user can browse an anatomical tree, select structures of interest, and define criteria for gene expression (present, possible, not detected in any or all selected structures). (2) The link “View Kidney Structures [...]” allows the user to select a single structure from either an anatomical tree or a stack of sections comprising 400 serial sections of a histologically stained E17.5 kidney (A in Fig. 5). Upon selection of a structure, a report is generated that provides an overview of

how many genes have been annotated with signal present, possibly present, or not detected. From this overview the user can link to the corresponding lists of ISH entries in the database. (3) However, probably the most comprehensive and intuitive way to query the database for gene expression in a given set of anatomical structures is using the “Adult Nephron Graphical Query Interface” (B in Fig. 5) found on the EuReGene portal page. The schematic drawing of an adult model nephron helps those users with limited knowledge in nephron anatomy. Here one can quickly select/deselect structures by pointing and clicking and search for datasets that are either annotated in any or all selected anatomical components.

All these queries lead to a results page that contains a table of matching datasets (ISH experiments defined by gene and stage) that can be sorted and filtered by various

criteria. Each line represents an experimental condition and features a thumbnail preview. Datasets can be retrieved into a virtual shopping basket so that results selected from different queries can be reviewed later as one list. Each dataset can be opened by clicking on “EuReGene Entry Details” or the thumbnail. Entry details are listed in table format and contain links to gene and template information as well as the annotation data and experimental metadata. Annotation data can be displayed using the complete anatomical tree (C in Fig. 5) or as a table listing only annotated structures. Most important, the entry details provide thumbnails of all available images in a dataset which link to the Image Viewer (D in Fig. 5), which allows zooming, panning, and downloading of high-resolution images. In addition, the Image Viewer displays on the left side a general quality report for each image.

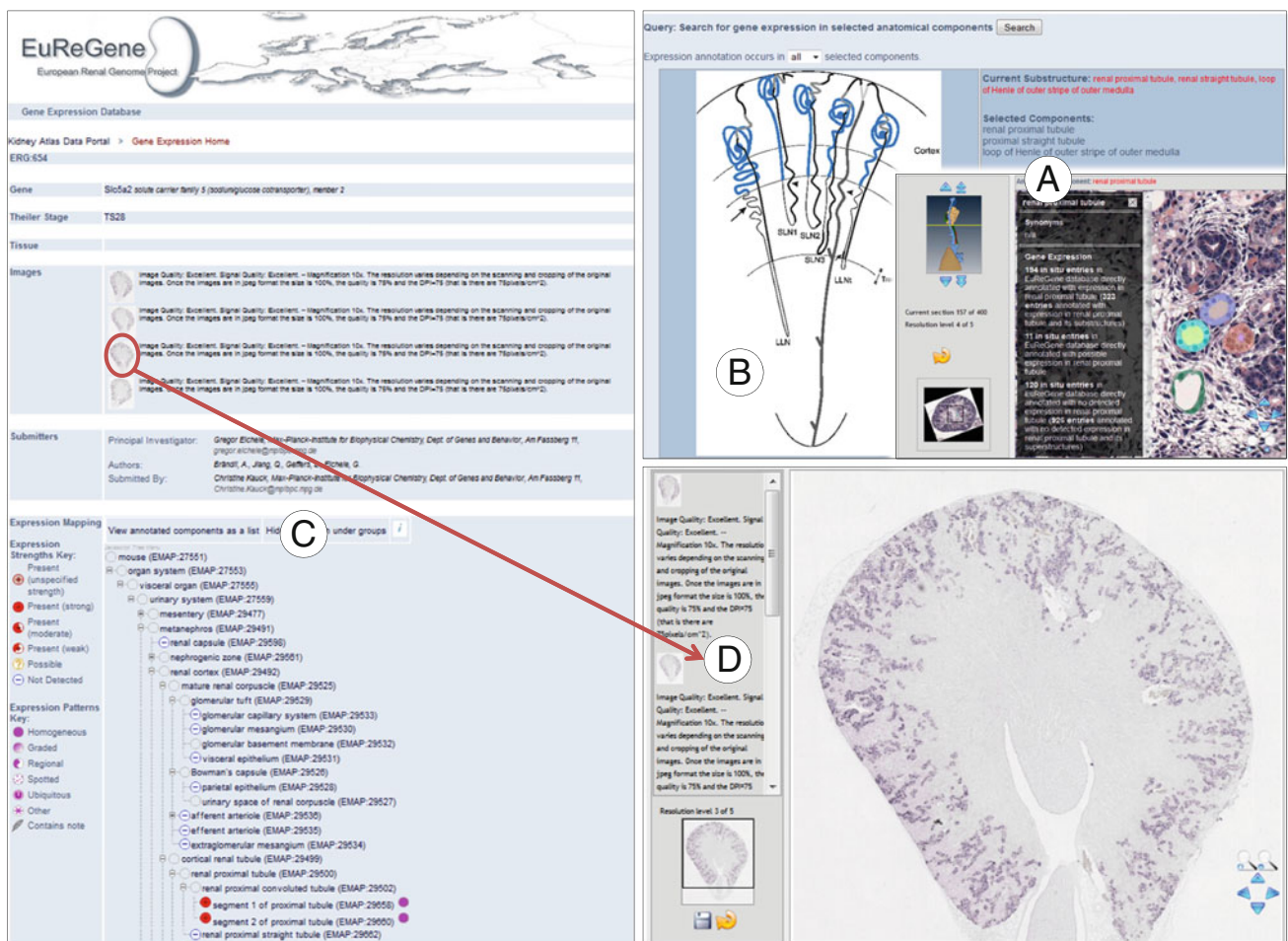


Fig. 5 Screenshot composite of the EuReGene Gene Expression Database. EuReGene covers the expression patterns of several gene families in sections of the E17.5 and adult mouse kidney. Apart from the regular text-based query input, EuReGene also provides graphical query interfaces that allow the user to select anatomical search terms from either a histological (*panel A*) or a schematic (*panel B*) browser

in order to identify genes expressed in these structures. The results page of an individual dataset displays annotation of gene expression in a collapsible hierarchical anatomy tree (*field C*) and links the representative thumbnails to an image viewer that allows inspection at high resolution (*panel D*)

Despite the limited content of this specialized database, it is a unique resource for the renal field mostly because of the quality and detail of annotation.

The EuReGene resources have been used to identify molecular markers of high specificity for a comparison of segmental organization in the *Xenopus* pronephric and the mouse metanephric nephron (Raciti et al. 2008). The study provided compelling evidence that the expression of key genes underlying inherited renal diseases in humans has been evolutionarily conserved down to the level of the pronephric kidney.

BGEM (<http://www.stjudebgem.org>)

The Brain Gene Expression Map (BGEM, generated at the St. Jude Children's Research Hospital in Memphis, TN, USA) has pioneered the concept that an analysis of gene expression needs to include embryonic, early postnatal, and adult stages (Magdaleno et al. 2006). BGEM grew in step with GENSAT and has provided input for gene selection for the latter project. BGEM allows the user to access a systematic collection of gene expression maps based on radioactive *in situ* hybridization. An advantage of radioactive ISH, besides its high sensitivity, is its high dynamic range. Autoradiographs lack histological landmarks, but BGEM provides corresponding cresyl violet-stained bright-field images for reference to anatomy. While images in BGEM lack cellular resolution, the organization and representation of datasets allow a rapid assessment of overall expression patterns throughout development, and data cover the major parts of the central nervous system (brain and spinal cord) and the eye. ISH data were generated from fixed-frozen tissue for four developmental stages (E11.5, E15.5, P7, and P42) and in three planes of sectioning. For each gene a manually annotated expression profile details signal characteristics in 25 CNS structures for all four stages side by side. Both signal intensity (–, +, and ++) and pattern (region specific, RS; scattered, S; region specific and scattered, RSS; and widely expressed, WE) are scored. All genes are registered with the corresponding Gene Ontology terms (Ashburner et al. 2000).

The start page grants fast access to a list of all genes in the database (click static brain image) and to highlighted individual datasets (click randomly appearing thumbnails). The main menu is placed within a corporate ribbon at the top of the page (A in Fig. 6). The search options as well as an exceptionally complete documentation of materials and methods can be accessed from the main menu. To query the database the user can choose from a variety of search topics and prefiltered lists. Genes of interest can be identified by keyword, sequence, chromosome, Gene Ontology annotation, or expression profile. Notable is the option to

search genes by expression profile, which, similar to GenePaint, allows the user to define complex expression constraints.

The result of any query is a list of genes from which individual datasets can be opened following the links in the list. In BGEM such lists can be exported to either MS Excel or in CSV format. The exported files contain all gene-related information stored in the database except the expression profile.

Each dataset provides a thumbnail overview of available ISH data (B in Fig. 6), and each thumbnail can be enlarged and viewed next to a histologically stained reference section (C in Fig. 6). Furthermore, all images can be accessed at print quality. If the user has registered for a free account with BGEM, images can be added to a shopping basket for later review. The images of a dataset can be conveniently downloaded as a single zipped file containing the complete dark-field data at highest resolution. In addition, the dataset provides gene information (D in Fig. 6) and links to other databases. In the gene information section, Gene Ontology terms and the chromosome number can be used to initiate a query with these arguments. Links are provided to review the template sequence (E in Fig. 6) and the annotation of the expression profile (F in Fig. 6).

The BGEM web interface puts the user into direct control of the database content by providing sophisticated search options and suitable means to extract data for offline use. The database offers a well-balanced compromise between spatial resolution and a stringent focus on the developmental trajectory of gene expression in the central nervous system. The implementation of Gene Ontology annotation in BGEM not only points at possible gene function but also serves as a valuable contextual sorting argument.

A study investigating the developmental expression pattern of *Hspb8* mRNA in the mouse brain cited the BGEM database (García-Lax et al. 2012).

GENSAT (<http://www.gensat.org>)

GENSAT, the Gene Expression Nervous System Atlas project (Gong et al. 2003; Heintz 2004) located at Rockefeller University, New York, excels in its approach to documenting gene expression in the central nervous system of the mouse. The mission of this project is to provide the scientific community with BAC-EGFP reporter and BAC-Cre recombinase driver mouse lines in order to map and manipulate gene function with cell-level resolution. The GENSAT database provides detailed information about the site of expression of the EGFP reporter and/or of Cre recombinase activity. Gene expression is indirectly monitored in transgenic mice by detection of an EGFP

Panel A: BGEM website header with navigation links and search options.

Panel B: Darkfield Expression Patterns grid for Niemann Pick type C2 at different ages: E11, E15, P7, Adult Coronal, and Adult Sagittal. A red circle highlights a specific image.

Panel C: Histologically stained reference section of the brain.

Panel D: Gene Information section for Niemann Pick type C2, including description, accession number, and functional annotations.

Panel E: Sequence section showing the probe sequence for Niemann Pick type C2.

Panel F: Expression profile table for g00412 across various brain regions and developmental stages.

Brain Regions	E11	E15	P7	Adult
Amygdala	NS	NS	ASS	RSS
Basal Ganglia	NS	NS	ASS	WE
Brainstem	NS	NS	WE	WE
Cerebellar Anlagen	NS	+	NS	NS
Cerebellum	NS	NS	WE	RSS
Cerebral Cortex	NS	NS	ASS	RSS
Forebrain	NS	NS	NS	NS
Ganglionic Eminence	NS	WE	NS	NS
Hindbrain	NS	ASS	NS	NS
Hippocampal Formation	NS	+	NS	NS
Hippocampus	NS	NS	ASS	RSS
Hypothalamus	NS	+	WE	WE
Midbrain	NS	ASS	ASS	WE
Neocortex	NS	ASS	NS	NS
Olfactory bulb	NS	WE	WE	WE
Retina	NS	RS	RS	RS
Spinal Cord	NS	ASS	WE	+
Spine	++	RSS	NS	NS
Thalamus	NS	NS	ASS	WE
Meninges	NS	NS	RS	RS
Choroid	NS	NS	RS	RS
Rhombencephalon	++	RSS	NS	NS

Fig. 6 Screenshot composite of the BGEM database. Gene expression patterns in the nervous system of the developing and adult mouse were characterized by radioactive in situ hybridization. The main menu providing links to documentation and search modes is integrated within the corporate ribbon at the top of the page (*field*

A). The results page of an individual dataset displays a thumbnail overview of available images (*field B*). Thumbnails can be enlarged and viewed next to a histologically stained reference section (*panel C*). Next to general gene information (*field D*), the page provides links to the probe sequence (*panel E*) and an annotation table (*panel F*)

reporter. To replicate the endogenous expression pattern of a gene of interest, EGFP is expressed from a corresponding BAC construct. EGFP signals in founder lines are compared to data from conventional gene expression analysis in wild type to assess the level of veracity. The veracity report details how faithful the signal from the EGFP reporter resembles the endogenous expression pattern. EGFP is detected by fluorescence and imaged as a close-up at subcellular resolution. Additionally, EGFP expression is visualized by horseradish peroxidase-based immunohistochemistry, including Tyramide Signal Amplification (TSATM)

that allows the capture of high-resolution bright-field images.

Similar to BGEM, which maintains strong links to GENSAT and essentially provides a prescreen for interesting candidate genes, GENSAT also covers four stages of mouse development (E10.5, E15.5, P7, and adult) and employs a similar annotation scheme for gene expression with respect to expression level and pattern. Annotation is attributed to 15 major anatomic structures (e.g., hippocampus, thalamus, or cerebellum) and 59 fine structures (e.g., locus coeruleus, raphe nuclei, or mammillothalamic

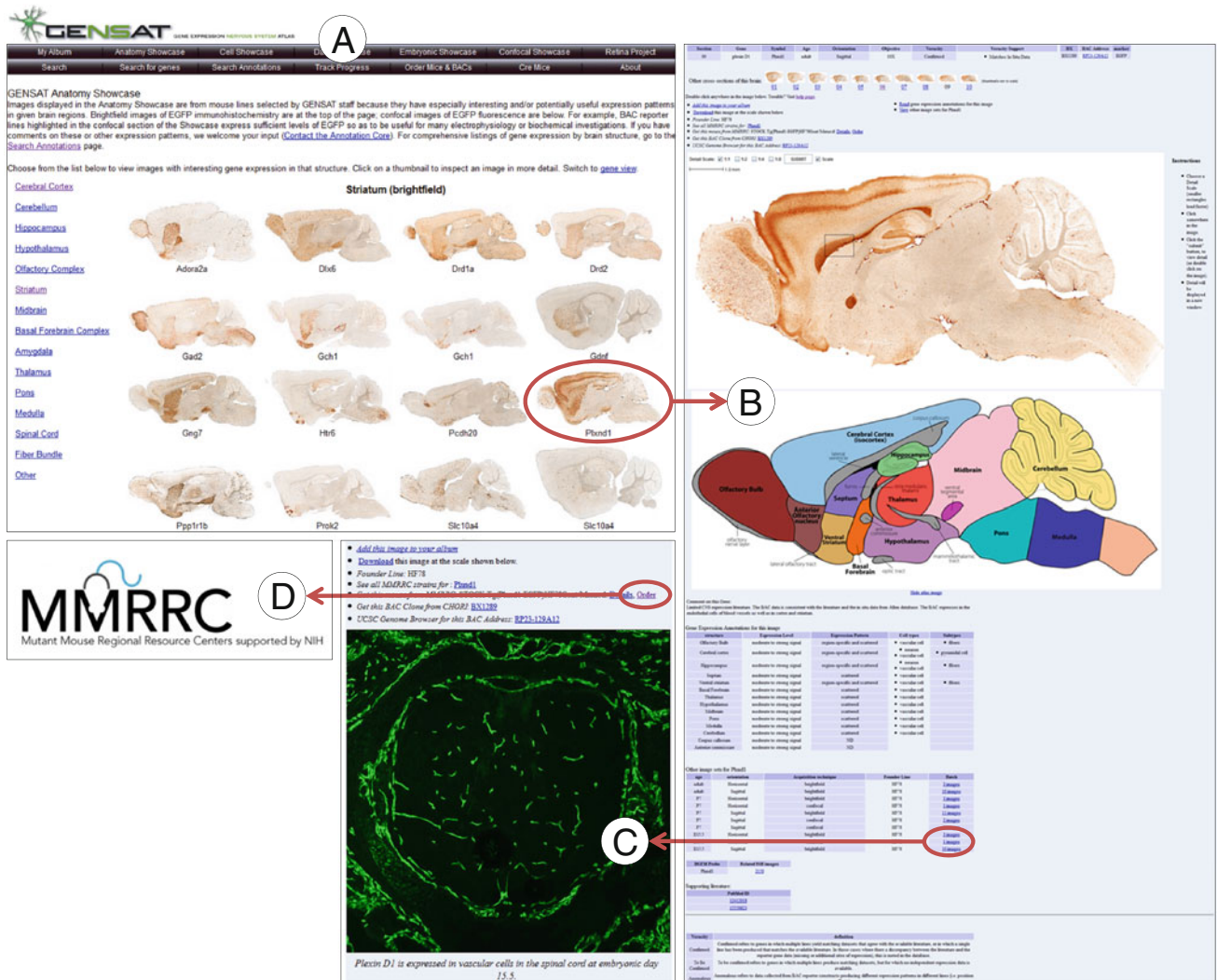


Fig. 7 Screenshot composite of the GENSAT database. GENSAT provides high-resolution characterization of publicly available BAC-EGFP reporter and BAC-Cre recombinase mouse lines generated by the project. The GENSAT web interface provides preassembled showcases that organize available datasets by topic. An example is the anatomy showcase (*panel A*) grouping datasets representing different genes expressed in the same anatomical structure. Once a dataset is selected, available images are shown as thumbnails on top

of a new page (*panel B*). The selected thumbnail is presented at higher resolution on the same page next to a reference image that can be toggled from schematic to histological (*center of panel B*). A table with gene expression annotation as well as a table with additional datasets for the same gene is also provided on this page (*bottom of panel B*). Additional datasets may illustrate GFP-reporter expression (*panel C*) and link to the Mutant Mouse Regional Resource Centers from which corresponding mouse lines can be ordered (*panel D*)

tract) and provides a catalog of cell-type-specific expression.

A key concept in GENSAT is the interactive showcase that brings attention to datasets deemed representative or interesting (A in Fig. 7). The web interface of the database features five different topics for showcases (Anatomy, Cell, Daily, Embryonic, and Confocal) and also has the option of a user-defined showcase (“My Album”). Each showcase presents thumbnails of images relevant to the topic or selectable subplot.

Search functions are divided into a general search, a search for genes, and a search based on annotation. In all search modes the user has access to a list of controlled

vocabulary terms in the form of info text or drop-down menu. Combinatorial search patterns can be easily created by applying logical operators.

Apart from the “Search for genes” option, the result of a query is always a list of Gene Symbols. While maintaining the original list of results, individual Gene Symbols can be expanded into a thumbnail overview of available images matching the search arguments. Next to Gene Symbol and age, the overview provides structural annotation information as well as the IDs of the founder line and the BAC construct for each individual image. Selecting a thumbnail opens an image navigator (B in Fig. 7) that allows the user to download or study the image in more detail. Areas of

interest from the DAB-stained sections can be enlarged at higher resolution in a new window. Usually the DAB-stained section is displayed alongside a schematic reference section that can be changed to a Nissl-stained image. The image navigator provides access to all relevant cross sections of the same series and links to other image sets collected for the same gene (C in Fig. 7). It provides an overview of the gene expression annotation and free-text comments on the veracity of the available transgenic lines. Moreover, from within the image navigator the user can order the corresponding transgenic mouse line (D in Fig. 7) or the BAC construct.

In the universe of gene expression atlases, GENSAT provides a unique resource that extends well beyond the mere documentation of gene expression. The generation and high-resolution characterization of transgenic mouse lines with a reporter system suitable for live imaging not only allows identification and tracking of cell identity throughout development but is also valuable for guided single-cell electrophysiology. Additionally, the incorporation of a Cre pipeline (Gong et al. 2007) into the existing infrastructure of the GENSAT database provides detailed characterization of 222 available Cre driver lines with expression in the central nervous system.

BAC transgenic mice derived from the GENSAT project were used in a recent screen for GFP-labeled retinal cell types (Siegert et al. 2009). This work illustrates that transgenic mouse lines with selectively marked cell types greatly facilitate the dissection of the components of complex neuronal circuits as found in the retina and in many other brain structures.

Perspectives

Much work and considerable public and private funding have gone into the production of digital atlases of gene expression for the mouse. Similar efforts have been undertaken for *Drosophila* (BDGP) and *Xenopus* (XGEbase). Additionally, there are *Drosophila* (FlyBase), zebrafish (ZFIN), *Xenopus* (XenBase), and chicken (GEISHA) databases that compile and register third-party expression patterns. Given that many biological processes are conserved, it would be desirable to link the expression patterns of orthologous genes across the various databases. This would allow generation of comparative molecular anatomy digital atlases, which would require intelligent new data-mining tools. An example of a medium-scale cross-species analysis done manually is a comparison of gene expression patterns in *Xenopus* pronephric and mouse metanephric tissue (Raciti et al. 2008).

Most annotations of structures are based on the detailed evaluation of individual embryos or images by experts.

Such textual annotations have a counterpart in “geometric” annotation in which a three-dimensional orthogonal grid is placed across an expression volume (e.g., the mouse brain) and levels of expression in each volume element are automatically calculated (Ng et al. 2007). In this approach there is no need to anatomically identify structures. Moreover, quantitation of expression is based on image analysis and thus may be less subjective than an expert-based annotation. Organisms have irregular shapes so fitting them with deformable meshes instead of orthogonal ones may be an even better approach (Carson et al. 2005). However, the realization of automated three-dimensional annotations of gene expression patterns in mouse embryos is still fiction but needs to be promoted wherever possible because of its enormous potential.

It is critical that the high-quality databases discussed here are maintained, suitable hardware migration is implemented, and, where appropriate, the data contained in them are expanded and/or revalidated. It would be desirable to maintain each database’s individual character and functionality such as the many customized search and visualization tools which are one of the hallmarks of these web resources. These goals may be a challenge since the original funding sources to create the data and databases have mostly expired. This predicament is partly mitigated by the ongoing efforts to preserve and propagate the information contained in the original expression atlases by integrating them into larger, curated databases such as the MGI Gene Expression Database or EMAGE.

Digital atlases are powerful resources complementing microarray or RNA sequence databases. Unlike the latter that consist of numerical values associated with a tissue or cell, the information in digital atlases is complex and therefore their deep mining is difficult and requires the development of new analysis tools that can process data of diverse modality ranging from global views provided by WISH to EGFP reporter transgenics that afford spectacular subcellular resolution. The decade during which digital atlases have been produced needs to be followed by a decade during which they are integrated, curated, and mined.

Acknowledgments The authors are supported by the Max Planck Society for the Advancement of Science.

References

- Ashburner M, Ball C, Blake J, Botstein D (2000) Gene ontology: tool for the unification of biology. *Nature* 25:25–29
- Blackshaw S, Harpavat S, Trimarchi J, Cai L, Huang H, Kuo WP et al (2004) Genomic analysis of mouse retinal development. *PLoS Biol* 2:E247
- Carson JP, Thaller C, Eichele G (2002) A transcriptome atlas of the mouse brain at cellular resolution. *Curr Opin Neurobiol* 12: 562–565

- Carson JP, Ju T, Lu HC, Thaller C, Xu M, Pallas SL et al (2005) A digital atlas to characterize the mouse brain transcriptome. *PLoS Comput Biol* 1:e41
- de Boer B, Ruijter JM, Voorbraak FPJM, Moorman AFM (2009) More than a decade of developmental gene expression atlases: where are we now? *Nucleic Acids Res* 37:7349–7359
- Diez-Roux G, Banfi S, Sultan M, Geffers L, Anand S, Rozado D et al (2011) A high-resolution anatomical atlas of the transcriptome in the mouse embryo. *PLoS Biol* 9:e1000582
- Finger JH, Smith CM, Hayamizu TF, McCright IJ, Eppig JT, Kadin J, Richardson JE, Ringwald M (2011) The mouse Gene Expression Database (GXD): 2011 update. *Nucleic Acids Res* 39:D835–D841
- García-Lax N, Tomás-Roca L, Marín F (2012) Developmental expression pattern of Hspb8 mRNA in the mouse brain: analysis through online databases. *Anat Rec (Hoboken)* 295:492–503
- Geschwind DH, Konopka G (2009) Neuroscience in the era of functional genomics and systems biology. *Nature* 461:908–915
- Gofflot F, Chartoire N, Vasseur L, Heikkinen S, Dembele D, Le Merrer J, Auwerx J (2007) Systematic gene expression mapping clusters nuclear receptors according to their function in the brain. *Cell* 131:405–418
- Gong S, Zheng C, Doughty ML, Losos K, Didkovsky N, Schambra UB et al (2003) A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. *Nature* 425:917–925
- Gong S, Doughty M, Harbaugh CR, Cummins A, Hatten ME, Heintz N, Gerfen CR (2007) Targeting Cre recombinase to specific neuron populations with bacterial artificial chromosome constructs. *J Neurosci* 27:9817–9823
- Gray P, Fu H, Luo P, Zhao Q, Yu J, Ferrari A et al (2004) Mouse brain organization revealed through direct genome-scale TF expression analysis. *Science* 306:2255–2257
- Heintz N (2004) Gene expression nervous system atlas (GENSAT). *Nat Neurosci* 7:483
- Kasprzyk A (2011) BioMart: driving a paradigm change in biological data management. *Database (Oxford)* 2011:bar049
- Lein ES, Hawrylycz MJ, Ao N, Ayres M, Bensinger A, Bernard A et al (2007) Genome-wide atlas of gene expression in the adult mouse brain. *Nature* 445:168–176
- Magdaleno S, Jensen P, Brumwell CL, Seal A, Lehman K, Asbury A et al (2006) BGEM: an in situ hybridization database of gene expression in the embryonic and adult mouse nervous system. *PLoS Biol* 4:e86
- Martinez-Ferre A, Martinez S (2012) Molecular regionalization of the diencephalon. *Front Neurosci* 6:73
- Ng L, Pathak SD, Kuan C, Lau C, Dong H-wei, Sodt A et al (2007) Neuroinformatics for genome-wide 3D gene expression mapping in the mouse brain. *IEEE/ACM Trans Comput Biol Bioinform* 4:382–393
- Qi L, Chen K, Hur DJ, Yagnik G, Lakshmanan Y, Kotch LE et al (2011) Genome-wide expression profiling of urinary bladder implicates desmosomal and cytoskeletal dysregulation in the bladder exstrophy-epispadias complex. *Int J Mol Med* 27:755–765
- Raciti D, Reggiani L, Geffers L, Jiang Q, Bacchion F, Subrizi AE et al (2008) Organization of the pronephric kidney revealed by large-scale gene expression mapping. *Genome Biol* 9:R84
- Richardson L, Venkataraman S, Stevenson P, Yang Y, Burton N, Rao J et al (2010) EMAGE mouse embryo spatial Gene Expression Database: 2010 update. *Nucleic Acids Res* 38:D703–D709
- Siebert S, Scherf BG, Del Punta K, Didkovsky N, Heintz N, Roska B (2009) Genetic address book for retinal cell types. *Nat Neurosci* 12:1197–1204
- Sunkin SM, Hohmann JG (2007) Insights from spatially mapped gene expression in the mouse brain. *Hum Mol Genet* 16 Spec No:R209–R219
- Visel A, Thaller C, Eichele G (2004) GenePaint.org: an atlas of gene expression patterns in the mouse embryo. *Nucleic Acids Res* 32:D552–D556
- Visel A, Carson J, Oldekamp J, Warnecke M, Jakubcakova V, Zhou X, Shaw C, Alvarez-Bolado G, Eichele G (2007) Regulatory pathway analysis by high-throughput in situ hybridization. *PLoS Genet* 3:1867–1883
- Yaylaoglu MB, Titmus A, Visel A, Alvarez-Bolado G, Thaller C, Eichele G (2005) Comprehensive expression atlas of fibroblast growth factors and their receptors generated by a novel robotic in situ hybridization platform. *Dev Dyn* 234:371–386
- Yokoyama S, Ito Y, Ueno-Kudoh H, Shimizu H, Uchibe K, Albini S et al (2009) A systems approach reveals that the myogenesis genome network is regulated by the transcriptional repressor RP58. *Dev Cell* 17:836–848