Int. J. Dev. Biol. 43: 397-403 (1999)

A bioinformatics approach to investigating developmental pathways in the kidney and other tissues

JONATHAN B.L. BARD*

Genes and Development Group, Department of Biomedical Science, Edinburgh University, Edinburgh, United Kingdom

ABSTRACT Over the past few years, large amounts of data linking gene-expression (GE) patterns and other genetic data with the development of the mouse kidney have been published, and the next task will be to integrate these data with the molecular networks responsible for the emergence of the kidney phenotype. This paper discusses how a start to this task can be made by using the kidney database and its associated search tools, and shows how the data generated by such an approach can be used as a guide to future experimentation. Many of the events taking place as the kidney develops do, of course, also take place in other tissues and organisms and it will soon be possible to incorporate relevant information from these systems into analyses of kidney data as well as the new information from microarray technology. The key to success here will be the ability to access over the internet data from the textual and graphical databases for the mouse and other organisms now being established. In order to do this, informatic tools will be needed that will allow a user working with one database to query another. This paper also considers both the types of tools that will be necessary and the databases on which they will operate.

KEY WORDS: *anatomical atlases, bioinformatics, databases, gene expression, interoperability, kidney development*

Introduction

There is an interesting dichotomy in the evidence that is being acquired as the techniques of molecular genetics are being applied to the phenomena of developmental biology. On the one hand, the amount of data being generated is simply awesome, at least as compared to our expectations of even a few years ago. On the other hand, most of this information simply floats freely and we find it hard to integrate, say, the expression patterns of an individual gene into a coherent framework that can explain the molecular basis of some event.

There are two types of reasons for these difficulties. The first and more important is that we simply do not have enough solid and reliable data yet available for most systems to provide that framework. The second is that we find it hard to access all of the data currently published and which might be relevant. The reason here is that the data is usually inadequately archived, particularly if it is only stored as a journal article. This is partly because that GE data that is published is an often incomplete sample of what is in the notebook (editors are becoming less attracted by such "facts" and now prefer what is viewed as experimental data). A more important reason, however, is that the literature is only indexed to a very limited degree and is therefore hard to query, even in on-line databases such as Pub-Med (see Table 1).

The obvious solution, and one that is now being implemented in a number of systems (see Bard, 1997; Bard *et al.*, 1998a), is to store all this molecular data in web-accessible databases that can be queried on the basis of gene name, anatomical tissue and other relevant key words. With this done, it becomes possible to query any of these databases, provided that the appropriate informatic tools needed to link them are available.

These issues are brought into strong focus by a consideration of kidney development. There is now a great deal of experimental data on how this tissue develops and this now being integrated with the increasing amounts of molecular data that are being published and stored within the *Kidney Development Database* (see Table 1 and Davies *et al.,* 1997). What makes the kidney such a fascinating tissue, at least to those who work on it, is that its morphogenesis requires a particularly wide set of developmental phenomena, that include induction, epithelialisation, branching morphogenesis, stem-cell maintenance and aggregate formation (for review, see Davies and Bard, 1998). If we are to have a molecular understanding of kidney development, we are clearly going to need data on these phenomena from other organisms, both to find candidate genes to

Abbreviations used in this paper: MM, metanephric mesenchyme; GE, gene expression.

*** Address for reprints:** Genes and Development Group, Department of Biomedical Science, Edinburgh University, Edinburgh EH8 9AG, United Kingdom. FAX: 44 (0) 131 650 6545. e-mail: j.bard@ed.ac.uk

0214-6282/99/$15.00
© UBC Press
Printed in Spain
www.ehu.es/ijdb

fill gaps in the kidney data and to analyse homologies between the kidney and these other systems involving the same phenomena.

This paper sets out to consider these issues by describing the various textual and graphical databases for the mouse and other species that are currently available or that are now being constructed, and that someone interested in nephrogenesis might need to access. It ends by considering the types of interoperability tools that are needed if users are to be able to access one database from another on the basis of their anatomical details and gene names.

The molecular data on kidney development

As it is now well known, the kidney forms from reciprocal inductive interactions between the ureteric bud and the metanephric mesenchyme (MM) located within the metanephric blastema which is situated in the caudal region of the intermediate mesoderm. As a result of these interactions, the MM induces the bud to grow and bifurcate, while the bud initially induces the MM to partition itself between stem-cells located at the metanephric periphery and stromal cells in the medullary region. Induction does however continue almost until birth (in the mouse) with the stem cells inducing the bud tips to continue bifurcating, while these tips induce groups of stem cells to aggregate, epithelialise, and then form elongated tubules which fuse to the ducts at their so-called distal end and interact with capillaries at their proximal end to form the glomerulus.

There are now almost 500 genes known to be expressed as the mouse metanephric kidney develops. Most of this data comes from GE studies, and here, it is usually possible to localise this expression in space and time on the basis of relatively few sections because the developmental anatomy of the developing kidney changes little from E14 to birth, other than to enlarge and have ever-more mature nephrons. It is thus easy to localise GE patterns to a relatively few classes of tissues, and this in turn makes archiving simpler than in most other tissues (see below). Such GE data does not, in the very great majority of cases, give any clue about their function, other than by the nature of the type of gene under consideration (e.g. signal, receptor, transcription factor, etc.), or its established role in other tissues or organisms.

The exact function of most of these 500 or so genes thus remains obscure, but there is relevant and important functional data on a few of them, and it comes from two essentially different approaches. Because the metanephros will develop so well in culture, it has proven possible to investigate *in vitro* the role of some of the signalling molecules and cell-surface proteins and associated molecules expressed during nephrogenesis. This work has been based on the use of enzymes to destroy them and on the addition of ectopic protein either through the medium or through localised beads that can block receptors or give signal to a region of the developing kidney that might not usually receive it (for review, see Davies and Bard, 1998).

The other approach to investigating function has been to analyse transgenic mice where the expression of a gene has been down-regulated (through gene targeting, or addition of loss-of-function constructs) or upregulated (through gain-of-function constructs). The use of this approach has been less successful than one might have liked as, of the almost 60 mutations in the literature where a phenotype has been sought in the kidney (for review, see Davies and Bard, 1998, and http://www.ana.ed.ac.uk/anatomy/database/kidbase/kidhome.html) the great majority either have no phenotype, an abnormality that only appears very late in development, a loss of kidneys (implying that the genes are important too early in nephrogenesis to be useful in analysing kidney development), or such a major effect that it is hard to analyse the actual role of the gene.

There are, however, a few genes where the transgenic data points fairly clearly to the processes in which the gene is involved. *WT1* (metanephric mesenchyme competence, Kreidberg *et al.,* 1993), *BF-2* (the role of the stroma in regulating aggregate formation; Hatini *et al.,* 1996), *Wnt-4* (the initiation of epithelialisation in nephrogenic aggregates; Stark *et al.,* 1994), *bcl-2* (cell death; Veis *et al.,* 1993) and the *BMPs* (the regulation of growth, e.g. Dudley *et al*., 1995) are obvious candidates. These genes can thus clearly be assigned to classes of genetic network, but they are in a minority

The databases

Kidbase

Those who study kidney development are extremely fortunate that there is a textual database in which all of the expression data for these 500 or so genes is stored, together with references and transgenic data (*Kidbase*, Table 1; Davies *et al*., 1997). Because the essential geometry of the developing kidney is so simple, there are only a few types of tissue in which the genes are expressed (duct domains, stroma, stem cells and the various component of nephrons) and this makes linking GE patterns to anatomical domain names far easier than for most tissues.

A key strength of this database is that it can be searched in the following ways:

- gene name (linked to expression pattern and the literature)
- expression pattern (lists genes expressed in specified structures)
- gene type (signal, receptors, transcription factors etc.)
- unique tissue markers
- mutants

The use of these search tools provides a list of candidate genes and their functions (based on gene type and transgenic data) that are likely to be involved in the morphogenesis of any specified tissue. Even candidate initiating signals can be identified on the basis of local geometry.

The inevitable limitation of the database is that, because not all of the genes involved in kidney development have been identified, it cannot yet contain sufficient data to elucidate genetic networks. It is because the data that may be needed for constructing such networks on, for example, induction, growth, death and epithelialisation may be available from other systems that one needs to be able to search other databases.

Textual databases for the mouse

In addition to the Kidney Development Database, there are now several textual databases containing mouse gene expression and function data that are now online, or soon will be, that may contain information relevant to kidney development. GE databases include the relatively small ones for ducted glands and teeth as well as the more ambitious one for the embryo as a whole (GXD, Table 1), while the key database dealing with function is *TBase*. which

includes information on targeted mice and other mutants (see Table 1).

The smaller GE databases link a limited set of anatomical terms to a more or less up-to-date set of genes expressed in them, as well as including associated links and other useful material. GXD has a far more ambitious aim and this is to make available all of mouse embryo GE data in a format that directly links it to mouse developmental anatomy at quite a fine level of resolution (Ringwald *et al.*, 1997). The large relational database has to incorporate several components, the key ones being;

- a literature index containing all the core references
- a mouse developmental anatomy database
- mouse cDNA data
- expression patterns of various types linked to the anatomical nomenclature
- pictures of gene-expression data

One item of particular note here is the database of mouse developmental anatomy (DMDA, Table 1), which provides the keywords for inputting, archiving and searching the data and is now essentially complete (Bard *et al.*, 1998b; DMDA, Table 1). For this, mouse developmental anatomy for each of the 26 Theiler stages (E1-E17.5) is given as a tree, with tissues being grouped in a hierarchy based on common, intuitive roots that mainly centre around major organ or functional names (e.g. cardiovascular and respiratory systems) and common tissue types (e.g. glands and muscles). There can be some difficulty in defining when a tissue can be recognised in a developing embryo, but, by and large, the anatomy database seems robust and generates some 5000 pigeon holes for storing GE data, although many of its names represent the same tissue at different developmental ages. It should, however, be said that its resolution for the kidney is not as fine as that in *Kidbase*, nor does it yet contain data on as many genes expressed in the kidney.

Major parts of GXD are now in place and new GE data is continuously being archived by its editors. The database is now sufficiently mature that large amounts of material are now coming online and the rest should, by the time of this article's publication, be available (the current status is available at the website, see Table 1) and searchable for kidney data.

Graphical databases for the mouse

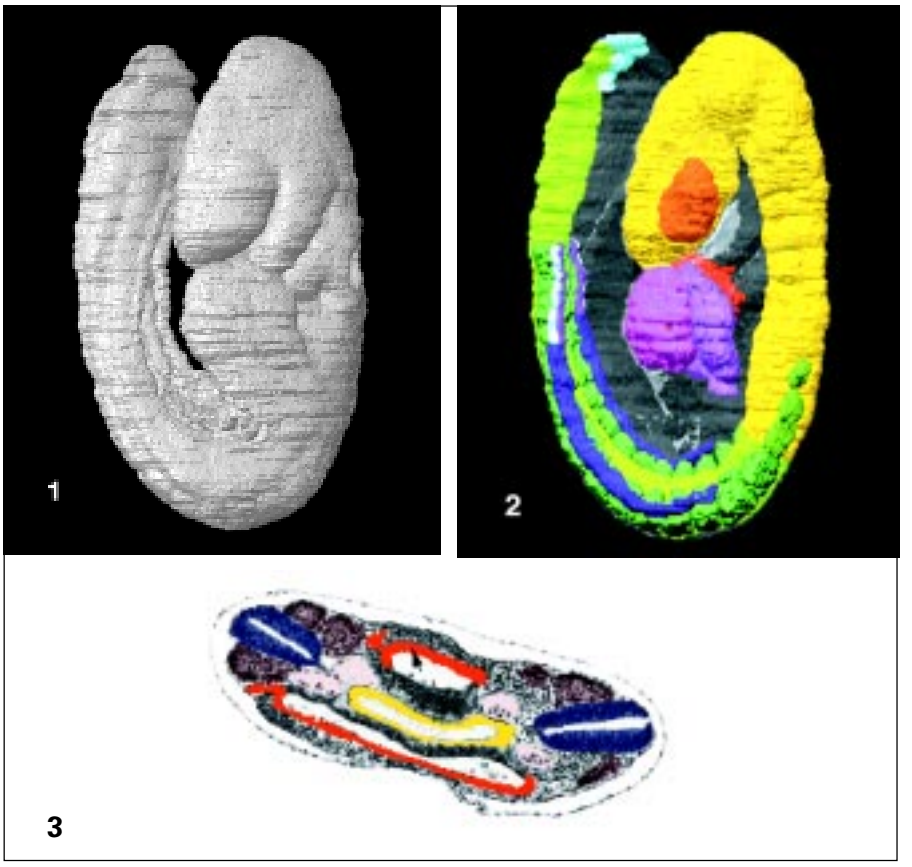
The anatomy database is, surprisingly, both a strength and a weakness for archiving data: its strength is that it provides a useful nomenclature for handling GE data; its weakness is that GE patterns frequently do not respect tissue boundaries. This weakness is particularly important for those genes whose role is to partition tissue domains (e.g. *Hox* genes). In a text database, one handles this difficulty by the use of notes, but this solution is not very satisfactory as such notes are hard to search automatically.

A better solution is to map GE data directly onto digital representations of the embryo and so produce a true graphical representation of the data. This approach has some distinct advantages over the textual approach of GXD, but has its own difficulties. The key advantages are that GE data is mapped to embryo space and not to the tissues which also of course occupy their own defined domains of embryo space. Where a gene is expressed within a

TABLE 1

SOME KEY WEBSITES FOR DEVELOPMENTAL DATABASES DEALING WITH GE AND FUNCTION (ONLY DMDA CURRENTLY INCORPORATES CORBA)

Resource	URL
Embryo databases	
A *C elegans* DataBase (ACEDB)	http://www.sanger.ac.uk/Software/Acedb/
The Nematode Expression Pattern Database	http://watson.genes.nig.ac.jp:8080/db/readme.html
Flyview	http://pbio07.uni-muenster.de/
The zebrafish database (ZFIN)	http://zfish.uoregon.edu
Mouse text gene-expression database (GXD/MGI)	http://www.informatics.jax.org/gxd.html
Mouse graphical gene-expression database (GGED + DMDA)	http://genex.hgu.mrc.ac.uk/
Kidney and ducted-gland gene-expression databases	http://www.ana.ed.ac.uk/anatomy/database/
The tooth database	http://honeybee.helsinki.fi/toothexp/toothexp.htm
TBASE (targeted mutations in the mouse)	http://tbase.jax.org/
Xenopus molecular marker resource	http://vize222.zo.utexas.edu/
Axeldb (*Xenopus* gene-expression database)	http://www.dkfz-heidelberg.de/abt0135/axeldb.htm
Other databases	
Swis-Prot	http://www.ebi.ac.uk/swissprot/
Entrez	http://www.ncbi.nlm.nih.gov/Entrez/
PubMed	http://www.ncbi.nlm.nih.gov/PubMed/
NHGRI Microarray Home page	http://www.nhgri.nih.gov/DIR/LCG/15K/HTML/
MRC HGMP Resource Centre (houses several databases including DHMHD and OMIM)	http://www.hgmp.mrc.ac.uk/



Figs. 1-3. Digital images of a TS14 (E9) mouse embryo reconstructed from a complete stack of digitised sections. *For this embryo, all the individual tissues have been delineated and each can be shown in 2 or 3D in any orientation (For further details, see Davidson et al., 1997. Pictures courtesy of Renske Brune and Richard Baldock).* **(1)** *The embryo reconstructed from a stack of histological sections and surface-rendered to show the ectoderm.* **(2)** *The same embryo with the ectoderm removed and displaying some of the underlying tissues, particularly the early nephric duct and intermediate mesoderm (dark and light blue). Other tissue shown include presumptive brain (brown), neural tube (light green), somites (dark green) , eye placode (red) and heart (purple/ mauve).* **(3)** *A section from the digital stack showing the coelomic cavity and surrounding tissues (from Bard et al., 1998). The expression pattern for the WT1 gene (Armstrong et al., 1992) is shown in red. (Tissues illustrated include the neural tube, blue; the gut, yellow; the somites. brown; and the dorsal aorta, pink).*

whole tissue, textual and graphical databases contain the same information, but the latter has one additional advantage. Because the anatomy database contains no spatial information (e.g. all the arteries are grouped together), GXD and any other textual database cannot be searched on the basis of tissue proximity. In contrast, graphical databases can be queried in this way.

The kidney provides a good example here. We still do not know how the metanephric blastema forms in the caudal region of the intermediate mesoderm, but, with a complete graphical GE database, it becomes fairly easy to ask questions of the following form "which signal receptors are expressed in the caudal intermediate mesoderm?" and "which signalling genes are expressed within 100 μm of this domain, and where are they expressed?". In this way, it becomes possible to generate candidate genes for blastema induction (if that is how it forms).

The difficulties associated with making such a database provide the reason why it is not yet available. The infrastructure for a graphical database of GE requires a set of digital embryos (Kaufman *et al.,* 1996) and the effort here is large: each has to be reconstructed from digitised serial sections that have to be electronically warped to remove distortions and onto which are "painted" all the tissue boundaries. Ideally, every voxel (3D pixel) in embryo space is assigned to one or another tissue so that the whole of embryo space is allocated to unique anatomical domains.

A database that can handle these reconstructions is also not simple to construct, as standard relational databases which incorporate data in tables do not handle graphical information in any natural way. It turns out that a sensible choice here is to use an object-oriented database as this is able to integrate all types of data (such databases use internal pointers to link one self-contained object with another), and our choice has been *ObjectStore* (Davidson *et al.,* 1997).

The only graphical GE database currently being assembled is for the mouse (GGED). Thus far, many of the early embryos have been reconstructed (Figs. 1,2), the database structure is in place and linked to the GXD text database. It can thus, in principle at least, display GE data such as that for WT1 (Armstrong *et al.,* 1992) in graphical format (see Fig. 3), although the database does not yet contain its own graphical GE data.

Interoperability with databases of other species

It is now clear that gene function data applicable for one organism is likely to be relevant to another, even if the species are not close (e.g. mouse and *C. elegans*, Kumar *et al.*, 1994). For mouse kidney development, therefore, it is more than likely that there will be novel information held in the databases of other organisms that will be helpful in constructing the genetic networks underpinning mouse kidney development.

An important aspect of the bioinformatics enterprise is, of course, that it allows users to access information from a wide range of sources, and the property that allows one database to interrogate another transparently is known as *interoperability*. At the moment, however, none of the major embryo databases (see Table 1) is interoperable with another; this is because each was developed independently for the benefit of its own cohort of workers. While there are protocols that facilitate interoperability (CORBA is probably the best known of these; Object Management Group, 1995), these are not generally in place yet for developmental databases.

The major embryo databases at the moment are for the mouse, zebrafish, *Drosophila* and *C. elegans* (see Table 1 for websites of these and other databases,) and, in the case of the kidney, for example, there is gene data held in databases for the mouse meso- and metanephroi (GXD), for the zebrafish pronephros and for the Malpighian tissues of *Drosophila*. It is also worth noting that kidney

data is held in the databases that cover human congenital diseases (e.g. OMIM, *Online Mendelian inheritance in man* and DHMHD, *the Dysmorphology human-mouse homology database*, see Table 1).

Provided that a user knows the appropriate anatomical terminology, it is clearly easy to visit any of these databases and explore what GE data they hold about kidneys and their homologues. Then, provided that the user recognises the genes expressed in the second organism and can identify their homologues in the original organism, the database search may yield new candidate genes that may be involved in nephrogenesis and whose role can be investigated experimentally (see discussion).

It should be emphasised that, although these databases do contain pictures of GE data, they are essentially textual as they can only be searched by word strings and not graphically. There are currently no plans to make graphical databases for these organisms, although it is worth pointing out that the essential 3D reconstructions have been made for *Drosophila* (Campos-Ortega and Hartenstein, 1997), and these could readily be incorporated into a graphical database (Fig. 4 shows the development of the Malpighian tubules, the *Drosophila* homologue of the vertebrate kidney).

Finally, it is worth mentioning a new class of gene expression database that is being assembled and it is from *microarray data*. Here, full sets of cDNA standards for an organism such as the mouse are prepared and individual clones then surface-localised in an array of microdrops on a glass slide (e.g. Marra *et al*., 1999). Once these are in place, any cDNA library can be assayed against these standards and its entire contents identified using automated immuno-labelled *in situ* hybridization technology, with the resulting data being stored in a standard database system (see Table 1)

In the first instance, such technology will be used for libraries from cell cultures from which it is possible to make a reliable cDNA library. In the longer term, however, the technology for making such libraries from the minute numbers of cells in specified organs in early embryos should become available. This will facilitate the identification of all the genes expressed in that tissue, and the next problem will be to identify which of these are involved in a specific network. It should however be pointed out that the technology for this is still in its earliest stages and we should not assume that working databases for the developing will be accessible for some time yet.

Discussion

This analysis of the databases and their use in analysing kidney development shows that bioinformatics has much to offer for developmental biologists seeking to assemble gene networks for nephrogenesis and for other systems. While tissue-linked GE data is now accessible for the mouse embryo, it is clearly not going to take a great deal of time to construct equivalent facilities for the other embryos. All that is required for a text database is an appropriate anatomical hierarchy lined to the gene-expression data, and the former at least is currently under construction for *C. elegans* and the zebrafish, while a sophisticated *Drosophila* anatomical database has already been assembled (and is accessible in *Flybase*, see Table 1).

The next serious challenge is *interoperability* so that one database can query another. This can, in principle, be achieved in two ways: through direct interfaces or through web tools. The former requires that the different databases present common queryable interfaces, the latter only requires that there be facilities for generating the search terms that allow a second database to be accessed for manual searching.

As all the major embryo databases have been generated independently, the interfaces (e.g. CORBA) necessary for direct interoperability are not yet in place and, given the difficulties both in putting the protocols in place and in co-ordinating the appropriate query languages, it may take some time for them to be implemented. A simple short-term expedient therefore is to produce the anatomical and gene tools that will enable a user to link tissues and genes in one system with those in another.

The sort of question that these tools will help answer from GE databases is as follows. Given a tissue (e.g. the kidney) in one organism (the mouse), are there genes expressed in the functionally equivalent tissue in a second organism (e.g. the Malpighian tubules in *Drosophila*) whose homologues might be expressed in the first organism. To handle such a query, two expert tools are needed, one that links the anatomies of the various organisms, and a second that identifies gene homologues.

The anatomy and gene tools would thus be used together in the following way to find genes that might be involved in, say, *C. elegans* on the basis of the mouse data. First, one uses the anatomy tool to search the mouse anatomy database to find the first instance of the tissue which is homologous to that in *C. elegans*. Second, one searches the mouse GE database for genes expressed in that tissue. Finally, one uses the gene tool described above to find the homologous genes in *C. elegans*.

Producing such tools is a non-trivial exercise and some would doubt whether it is in general possible to produce something that has sufficient expert knowledge for either task. Some aficionados indeed take the view that a great deal of knowledge is needed to operate these tools properly and to interpret their output and that the use of such tools should be restricted to those who understand them. Others take the simpler line that enough expert knowledge should be embedded in the tool for it to be helpful to the naïve user, and that, while its use may not do everything that the expert may want, it will be helpful to that user. This is an important point as the creator of a tool never knows who will need it and it is important to provide help to someone who is not an expert. The key point, however, is that every user must be aware of the limitations of the tool that they are using.

Gene homologue search tool GHoST

An example clarifies this point. A literature search for genes expressed in the Malpighian tubules in *Drosophila* shows the *kruppel* is expressed in the tip cells (Hoch and Jackle, 1998), as well as a variety of other intriguing genes (e.g. Liu *et al*., 1999); someone interested in mouse nephrogenesis will immediately want to know the name of any mouse homologue of this gene. The obvious search facility here is *Entrez* which is a large database maintained by the National Centre for Biology Information (Table 1) that, *inter alia*, contains a fairly complete dataset of all published nucleotide and proteins sequences, archived under a wide variety of headings, with each having links to other database entries.

Searching Entrez for *kruppel* yields 317 "hits" that cover a wide range of organisms, far too many to be readily searched, and a cull is needed to produce only those which are realistic candidate genes in the mouse. The first function of this tool is just to filter this output

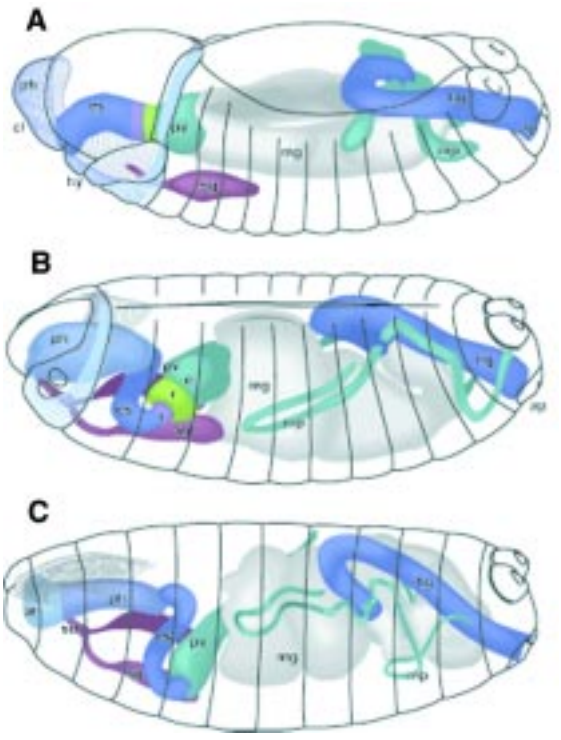


Fig. 4. Schematic drawing based on 3D histology showing the development of the *Drosophila* gut and the growth of the Malpighian tubules (mp) at the border between the midgut (mg) and hindgut (hg) from small evaginations of the endoderm (A, stage 13-14) that elongate and narrow (B, stage 16; C, stage 17). *Other tissues shown include the proventriculus (pv), the pharynx (ph), the somatic mesoderm (sm), visceral musculature (vm), oesophagus (oe) and gastric coeca (gc). (Courtesy of Volker Hartenstein; from Campos-Ortega and Hartenstein 1997; with permission from Cold Spring Harbor Press).*

to give only the mouse homologues. The next task is to align these against the *Drosophila* kruppel sequence so that those which failed to meet some (user-defined) homology standard were also filtered out. The remainder, with hot links to the databases, would then be presented on the screen to the user. Such a tool (GHoST) is currently being constructed in collaboration with Val Curwen and Gary Williams of the MRC Human Genome Mapping Project Resource Centre (Table 1), and, once completed, will be made publicly available.

Even at this stage, however, it is worth pointing to the limitations of GHoST: it will only generate homologues that have been stored in Entrez and we do not know if the list is complete, or if the search tool is completely adequate. Users who feel that more detail is required will need to explore the protein databases (Swiss-Prot etc., Table 1) in the usual way, and do their own filtering of the output.

An anatomy tool

It is rather more difficult conceptually to construct a tool that generates the tissues in a second organism that are homologous to the first as such a facility demands an enormous amount of anatomical data. Indeed, the detailed homology that might link very different anatomies (e.g. the mouse and *C. elegans*) certainly has not been done, and, indeed, it is not easy to see how it could be. Even in organisms where there are strong similarities between the anatomies (e.g. mouse and human), there are sufficient differences in nomenclature for there not to be good one-to-one identities for some tissues. In short, it is not going to be possible to make a perfect matching tool for anatomy as the organisms are simply different.

It is therefore worth considering what might actually be possible here. A sensible starting point is that every animal whose development has been studied has a well-known anatomy at each stage of its development that becomes more complex as development proceeds. The anatomy at each stage can be expressed in a tree format (e.g. Bard *et al*., 1998b) that includes developmental-stage and functional information as well as anatomical data. In the case of the mouse, for example, shows that the first occurrence of the mesonephros is in the E9.5 embryo and the database generate a hierarchy of the form:

```
TS15 (E9.5)
   embryo
      organ system
         visceral organ
            urogenital system
               mesonephros
```

while the earliest kidney entry for the early zebrafish is pharyngula (prim 5; 24h):

```
   embryo
      organ system
         visceral systems
            urogenital system
               pronephros
```

and it is obvious where one might search one or the other GE databases. For a slightly more complex example, one might consider a ganglion in the nervous system and the search might be between the mouse embryo and *C. elegans*. One such ganglion in the mouse is:

```
TS19 (E11.5)
   embryo
      organ system
         nervous system
            central nervous system (CNS)
               ganglion
                  cranial
                     vagus X
                        inferior (nodose ganglion);
```

(where bracketed terms can be viewed as synonyms), while that for *C. elegans* will be:

```
mature worm
   nervous system
      ganglion
         head ganglion
            anterior ganglion
```

If the purpose of the search is to find the mouse tissue that might be considered analogous to that in *C. elegans*, then the use of the anatomy search tool is, in principle, simple: one takes the lowest level *C elegans* term (anterior ganglion) and look for the same term in the mouse anatomy. If this fails, one goes up the tree to head ganglion, and then up to ganglion where there is a direct match. The first occurrence of "ganglion" in the mouse database turns out to be at TS14 and refers to the neural crest that will be the future

trigeminal ganglion and these two terms provide sufficient information to search GXD, the mouse gene-expression database, for candidate genes.

The key point about this approach to comparative anatomy is that it is readily and rapidly computable as programmes that handle trees are relatively easy to write. This advantage probably outweighs the obvious disadvantage, that the only knowledge here lies in the individual anatomical trees and that the matching is based on word similarities rather than on any attempt to provide a detailed anatomical comparison between any two organisms. Nonetheless, provided that the user is aware of the limitations of the approach, useful information on both tissue types and developmental time may be obtained that might not otherwise be easy to access.

Time scale

One of the problems about discussing GE and developmental anatomy at the moment is that for the sorts of systems discussed in this article to work, more databases and tools have to be in place than are currently available.

At the moment, the only substantial gene-expression databases in place are for kidneys, teeth and for the mouse embryo and these are all textual. Anatomical databases have been constructed for the mouse, human and for *Drosophila* (although this is not currently in a hierarchical format), while those for the zebrafish and *C. elegans* are currently being put together. It is of course not possible to build a GE database until the anatomy component is in place to provide the input, storage and search terminology.

All of the anatomical databases should be in place soon into the new millennium, but the GE databases for embryos other than the mouse will take longer to assemble and integrate. Indeed, the time for these to be in place will probably depend on the amount of funding the enterprise can attract. Similarly, for all the grant support that the databases that are to be filled by micro-array data (see above) are attracting, they should not be expected to be publicly accessible for at least another five years.

Once all these databases are in place, however, the enormous amount of work that has been done in eliciting the patterns of GE data in developing organisms can begin to be made accessible to those involved in trying to discover the major genetic networks responsible for generating developmental anatomy. Already, however, those who work on the development of the kidney are fortunate in having *Kidbase*, and ready access to all the GE data for nephrogenesis. Indeed, this resource provides yet another reason for working on this best of model systems. It is however unlikely to be adequate to the task of assembling all the genetic networks that underpin kidney development and this task will need every tool and database that bioinformatics can make available.

Acknowledgements

I thank Val Curwen and Gary Williams of the MRC Human Gene Mapping Resource Centre for discussions on the gene homology tool, and the MRC for partially funding this work.

References

ARMSTRONG, J.F., PRITCHARD-JONES, K., BICKMORE, W.A., HASTIE, N.D. and BARD, J.B.L. (1992). The expression of the Wilms' tumour gene, WT1, in the developing mammalian embryo. *Mech. Dev. 40:* 85-97.

BARD, J. (1997). Making and filling gene-expression databases. *Semin. Cell Dev. Biol. 8:* 455-458.

BARD, J.B.L., BALDOCK, R.A. and DAVIDSON, D.A. (1998a). A bioinformatics approach to analysing the genetic networks of development. *Genome Res. 8:* 859-863.

BARD, J.B.L., KAUFMAN, M.A., DUBREUIL, C., BRUNE, R.M., BURGER, A,. BALDOCK, R.A. and DAVIDSON, D.A. (1998b). An internet-accessible database of mouse developmental anatomy based on a systematic nomenclature. *Mech. Dev. 74:* 110-120.

CAMPOS-ORTEGA, J.A. and HARTENSTEIN, V. (1997). *The Embryonic Development of Drosophila melanogaster*, (2nd edition). Springer Verlag, Berlin.

DAVIDSON, D., BARD, J., BRUNE, B. BURGER, A., DUBREUIL, C, HILL, W., KAUFMAN, M., QUINN, J., STARK, M. and BALDOCK, R. (1997). The mouse atlas and graphical gene-expression database. *Semin. Cell Dev. Biol. 8:* 489-498.

DAVIES. J.A. and BARD, J.B.L. (1998).The development of the kidney. *Curr. Top. Dev. Biol. 39:* 245-301.

DAVIES, J.A., BRANDLI, A.W., HUNTER, D. and NIEMINEN, P. (1997). Design considerations for small, special-system developmental databases. *Semin. Cell Dev. Biol. 8:* 519-525.

DUDLEY, A.T., LYONS, K.M. and ROBERTSON, E.J. (1995). A requirement for bone morphogenetic protein 7 during development of the mammalian kidney and eye. *Genes Dev. 9:* 2795-2807.

HATINI, V., HUH, S.O., HERZLINGER, D., SOARES, V.C. and LAI, E. (1996). Essential role of stromal mesenchyme in kidney morphogenesis revealed by targeted disruption of Winged Helix transcription factor BF-2. *Genes Dev. 10:* 1467-1478

HOCH, M. and JACKLE, H. (1998). Kruppel acts as a developmental switch gene that mediates Notch signalling-dependent tip cell differentiation in the excretory organs of *Drosophila*. *EMBO J. 17:* 5766-5775.

KAUFMAN, M.H., BRUNE, R.M., BALDOCK, R.A., BARD, J.B.L. and DAVIDSON, D. (1996). Computer-aided 3-D reconstruction of serially-sectioned mouse embryos: its use in integrating anatomical organisation. *Int. J. Dev. Biol. 41:* 223-233.

KREIDBERG, J.A., SARIOLA, H., LORING, J.M., MAEDA, M., PELLETIER, J., HOUSMAN, D. and JAENISCH, R. (1993). WT-1 is required for early kidney development. *Cell* 74: 679-691.

KUMAR, S., KINOSHITA, M., NODA, M., COPELAND, N.G. and JENKINS, N.A. (1994). Induction of apoptosis by the mouse Nedd2 gene, which encodes a protein similar to the product of the *Caenorhabditis elegans* cell death gene ced-3 and the mammalian IL-1 beta-converting enzyme. *Genes Dev. 8:* 1613-1626

LIU, X., KISS, I., and LENGYEL, J.A. (1999). Identification of genes controlling malpighian tubule and other epithelial morphogenesis in *Drosophila melanogaster*. *Genetics 151:* 685-695

MARRA, M., and 27 other authors. (1999). An encyclopedia of mouse genes. *Nature Genet. 21:* 191-194.

OBJECT MANAGEMENT GROUP (1995). *The Common Object Request Broker Architecture and Specification.* OMG Publications, Framinghan Ma, USA.

RINGWALD, R., DAVIS, G.L., SMITH, A.G., TREPANIER, L.E., BEGLEY, D.A., RICHARDSON, J.E. and EPPIG, J.T. (1997). The mouse gene-expression database GXD. *Semin. Cell Dev. Biol. 8:* 509-518.

STARK, K., VAINO, S., VASSILEVA, G. and MCMAHON, A.P. (1994). Epithelial transformation of metanephric mesenchyme in the developing kidney regulated by Wnt-4. *Nature 372:* 679-683.

VEIS, D.J., SORENSON, C.M., SHUTTER, J.R. and KORSMEYER, S.J. (1993). *Bcl-2*-deficient mice demonstrate fulminant lymphoid apoptosis and abnormal kidney development in *bcl-2*-deficient mice. *Am. J. Physiol. 268:* F73-F81.