Bioinformatics Framework at Organism Level

Traditional bioinformatics is mostly about some 1-D linear sequences of gene, protein. A lot of efforts have been done to establish the databases, to design better searching algorithm or match system, etc. Many things in biology, however, cannot be expressed in 1-D sequences, such as the organism. It has a 3-D pattern. The frameworks at organism level can help us to understand the function of gene. We can "compare precisely the patterns of expression of different genes" by them. Some groups have done a lot of work about it [1]. I will focus on the digital frameworks for embryonic development.

Structure:

Since we need to describe the organism itself, we need not only the text, but also some image. The framework should include time, location of organisms in the embryo and the common concepts of developmental biology. The development of every organism is divided into some standard discrete, defined stages of development. The text is appropriate to relate data to these concepts. Some databases use the text as the index for the image and limit the searching in the textual framework because of difficulty in the comparing of images. One interesting example is the Gene Ontology (GO) framework [2]. However, the graphics model has many advantages because it is much easier to be searched with experimental raw data. And with the recent advances in image processing and computer graphics, some graphical atlases have been developed. But Pattern-recognition and image-processing methods, which are used to establish and query these databases, are only beginning to be developed. Only some very simple searches can be done now. We will focus on these atlases.

Some Example Atlas Frameworks:

1. Mouse Genome Informatics Gene Expression Database (GXD) [3]
   It stores images and annotates them by text. Searching remains limited by the textual framework.

   Frameworks below can be search by both text and image.
2. The Edinbugh Mouse Atlas Project (EMAP) [4,5]
   It combines textual and spatio-temporal approaches. It is an on-going project. The atlas is build by histological sections. One aim of this project is to develop a general tool to build similar frameworks for other species.
3. MRI atlas of mouse development.
   The atlas is build by high-resolution magnetic resonance imaging [6].
4. Other
   There are some frameworks in other organisms, such about larval brain (the FlyBrain project), zebrafish, human embryo, etc.

Compromises and Limitations:

1. The framework must have the sufficient resolution to index spatio-temoral and quantitative relationships in the raw data and in every entry of database. We need more details to map a pattern, which needs a higher resolution. Meanwhile, we only have limited resources to index huge large numbers of data, which needs a lower resolution. We have to optimize the balance between them.
2. The choice of image material for building the models is crucial. A 3-D voxel image is much better than a surface image or a wire-frame image.
3. In the atlas above, each model is based on data from an individual specimen, which subject to variation. So we need to make models accommodating the variations between specimens.

Use:

1. A reference framework
   The most obvious use of these atlases is as a bench tool to identify and name parts of the embryo and to explore their morphological development in embryo. [7]
2. Interaction with other data sources
   After designing a good interface with other databases, we can query them from both sides. We will extend our work to some spatial analysis and visualization.
3. A tool for local analysis
   The most important use of these frameworks is to query it with some raw image data. Mapping the data to the framework is a very useful tool for comparing small numbers of gene expression patterns and for analyzing individual gene expression result. It can also help to collect other information related to gene function, such as cell proliferation, apoptosis.

   However, this area needs a lot of effort. The resolution is limited and varies a lot in both the atlas and the experimental raw data. Furthermore, it is hard to recognize the models in the 3-D images, to collect enough information for every model and to compare them, which is unlike the 1-D sequence of the gene and protein. Some methods for 3-D to 3-D image transformation has been used. The trade-off between precision and time is crucial.

Future direction:

1. The better methods on pattern-recognition and image-processing are need desperately.
2. The first generation of frameworks contains the stage-by-stage views of development. A framework, which can present the continuous process, needs to be developed for further research.
3. The interface between all kinds of databases is needed. Some problems, such as gene function, will be more easily studied by better understanding of the relation between these databases.

Reference:

[1] Bioinformatics Beyond Sequence: Mapping Gene Function in the Embryo, Duncan Davidson & Richard Baldock, Nature Reviews Genetics, 2(6), 409-417, Jun. 2001

[2] The Gene Ontology Consortium. Gene Ontology: tool for the unification of biology Nature Genet. 25, 25-29 (2000)

[3] GXD: a gene expression database for the laboratory mouse: current status and recent enhancements. Ringwald, M., et al Nucleic Acids Res. 28, 115-119, (2000)

[4] A real mouse for your computer, Baldock, R., et al, BioEssays 14, 501-502, 1992

[5] A three-dimensional model of the mouse at embryonic day 9, Brune, R. M. et al. Dev. Biol. 216, 457-468, 1999

[6] In vivo visualization of gene expression using magnetic resonance imaging. Louie, A. Y. et al. Nat. Biotechnol. 18, 321-325 (2000).

[7] The Mouse Atlas Database: a community resource for mouse development. Davidson, D., Bard, J., Kaufman, M. H. & Baldock, R. Trends Genet. 17, 49-51 (2001).