

## Report of EMAGE advisory group

MRC Human Genetics Unit  
Edinburgh, UK  
18 November 2005

### **Advisory Board members present:**

Dr. David Wilkinson – National Institute for Medical Research, UK. (Chairperson)  
Dr. Graham Kemp – Chalmers University of Technology, Sweden  
Prof. Claudio Stern – Department of Anatomy, University College London, UK  
Dr. Sarah Wedden – MRC Technology, UK  
Dr. Alvis Brazma – European Bioinformatics Institute Hinxton, UK

### **Apologies:**

Prof. Steve Brown – MRC Mammalian Genetics Unit, Harwell UK.  
Dr. Martin Ringwald – GXD Database, Mouse Genome Informatics, USA  
Dr. Janan Eppig - Mouse Genome Informatics, USA

The aim of the EMAGE project is to establish a database of gene expression patterns during mouse development that will be a definitive resource for the international community. The database is based upon accurate spatial mapping of gene expression patterns onto an annotated three-dimensional atlas of mouse embryo anatomy. This approach enables gene expression data to be analysed in a number of informative ways, including the ability to search for spatially-overlapping gene expression patterns. The development of appropriate software for data input and analysis is a crucial part of the project. The strategy of EMAGE contrasts with other database projects that are either text based, provide images but no annotated spatial mapping, or are limited to gene expression patterns found in specific screens.

The EMAGE project is continuing to make very good progress in establishing a database of gene expression patterns during mouse development. The major priorities for the project fall into the following categories:

### ***(A) Data Entry***

Much of the potential value of the database requires that sufficient genes are represented. Rapid expansion of the database is therefore essential for it to become widely recognised as the international standard for spatial expression data. The current rate of entering data is as fast as can be achieved given the limited resources, and should be continued. A key strength of EMAGE is that it is creating a definitive spatial map of expression patterns onto the anatomy. We strongly support this strategy, although it does limit the rate of data entry. However, to greatly increase the amount of useful data available to the community via the EMAGE site, unmapped expression patterns from large scale screens can also be made available.

### **Recommendations:**

1. Data should continue to be entered at a rate of at least 1000-1500 spatially mapped patterns per year.

2. Provide pictures in EMAGE of in situs from large-scale screens, and other reliable sources, that have yet to be spatially mapped.

### ***(B) Software Tools***

The further development of software tools that enable sophisticated analysis of the data is very important. Impressive progress is being made in several areas, including Boolean analysis and clustering, that will greatly enhance the ability of users to extract important information. To accommodate the needs both of curious novices and of more experienced users with specific interests, we recommend that specific aspects of software development are prioritised, as follows:

#### Recommendation

- 3., It is recommended that the software be upgraded to include the following:
  - Boolean analysis, for example of spatial overlaps
  - Searches by similarity with painted domain
  - Cluster analysis of similarities
  - Simpler Google-type search of the database

### ***(C) Publicising the database***

Establishment of the EMAGE database as an internationally recognised resource is being promoted by it being presented at a number of meetings and courses, through publications, and by establishing links to EMAGE on sequence and genome databases.

#### Recommendation

4. These strategies for publicising EMAGE should be continued.

### ***(D) Relationship with other gene expression databases***

Coordination and/or integration of gene expression databases of different species is essential for comparative analysis, for example to find conserved and divergent expression patterns. Once appropriate databases have been established such analyses will become a routine research tool. It is therefore essential to maintain interactions with other gene expression database projects, such as in human, chick and zebrafish. It would be of great advantage for newly initiated database projects, such as for chick, to be co-located with EMAGE.

#### Recommendation

5. Maintain contacts with other database projects, including chick for which it would be a major advantage to use the same approach as EMAGE.