

2002 EMAGE Advisory Board Meeting

This short report contains a set of action points for EMAP and EMAGE to bear in mind over the coming year that were formulated by the Advisory Committee.

Also included are the minutes of the meeting are presented in Appendix I and notes on commercialisation from the HGMP-RC (Appendix II).

Action Points formulated by the Advisory Board:

- I The Editorial Office should endeavour to populate the EMAGE database primarily with
 - Data sets from large scale screening projects (eg. EURExpress) at one or two particular developmental stages;
 - Data sets of marker genes for several particular organ systems (eg. limbs, eye) at several different developmental stages and
 - Data sets of gene expression of components of particular pathways (eg. Wnt)

On launching it will be therefore be possible to conduct meaningful searches of EMAGE data by both space and time which will be necessary to demonstrate the usefulness of EMAGE to the community.

It was felt unprofitable to worry about single submissions before this date.

- II Anatomical delineation on the TS15-19 and TS21-25 EMAP models is currently lacking and funding for this activity is due to cease in the Summer of 2002. This feature is important for the usefulness of both EMAP and EMAGE. The advisory committee felt it necessary to urge for extended funding of these activities.
- III The textual anatomy descriptions used by EMAP and EMAGE will be impacted upon by phenotype ontologies derived from large scale mutagenesis screens and by the differences between molecularly defined anatomy vs. morphologically visible anatomy. EMAP is currently revising and extending the mouse anatomical delineations in conjunction with numerous sources. Review of the textual anatomy descriptions should be an ongoing process.
- IV EMAP is currently collaborating with others to develop the anatomical ontologies for other species (human, chick and zebrafish). For maximum use of the EMAGE data, future search mechanisms should allow interoperable searches of expression databases of other species also using links to the GO gene ontologies database. This would allow conserved or divergent aspects of gene expression to be detected between species.

APPENDIX I - Agenda and Minutes of EMAGE Advisory Committee Meeting

Date: Monday 18 March 2002

Present: Alvis Brazma
Steve Brown
Mike Dalrymple
Phil Gardner
Graham Kemp
Martin Ringwald
David Wilkinson

Richard Baldock
Jeff Christiansen
Duncan Davidson

Agenda and Minutes:

Welcome by Duncan Davidson and Richard Baldock

Election of Steve Brown as Chairperson

Presentation on Background and Current Status of EMAGE by Richard Baldock
Discussion raised during presentation...

SB - How will development of phenotype ontologies impact on the anatomical terms currently used by EMAP/EMAGE?

MR - GXD and EMAGE are committed to full integration of the anatomical dictionary.

DD - This would concern both the parts affected and the nature of the effect. There are already limitations in the current usage and we and others, including the Jackson Labs are modifying the framework.

RB - XSPAN project to link interspecies ontologies by Albert Burger (HGU).

SB - Did the proposals to the MRC have a specific number of deliverables by launch?

RB - No but we have said that we would like 1000 gene expression patterns in EMAGE by launch date.

DD - EURExpress is a source of many expression patterns at E9.5 (wholemound) and E17 (section).

DD - Explains 2D warping to wholemount.

DW - How will the EO tell, and be happy with authors saying expression is in a tissue from wholemount data? Do we also require substantiating evidence (eg tissue sections?).

DD - Yes if for example authors denote 'head mesenchyme' but no if they just denote 'head'.

SB - How does the EO envisage world wide sourcing of data?

JC - Collection of data from large scale screens such as EURExpress, approaching labs with specific interests in the development of particular organs (in order to obtain data sets of marker genes for these organs) or particular pathways (eg. Wnt pathway) and protein groups (eg. Hox, Fgf, Bmp etc). Will probably focus on collection at one or two stages so meaningful searches can be performed by launch date.

MR - What about anatomical delineation in 3D for the stages of development that currently don't have this? This is very important for annotating expression in the 3D models.

RB - Funding for this project is due to end this summer.

SB - If in-house funding from EMAP project is required for this, this would be a drain on the resources for other areas of the project. Should be underlined in the report to the board to recommend that funding for the tissue painting should be extended. The group agreed.

DW - Anatomy - would the anatomical terms be modified to reflect gene expression that marks tissues prior to any morphological evidence of the anatomy?

DD - Probably not, but this would be a linked expression pattern.

DW - This would affect lineages too - something to bear in mind.

Tour of Editorial Office (Demonstration of Submission and Query Interfaces)

Discussion raised during tour of EO

GK - Who are the primary users of the database? Searchers or submitters?

JC - Searchers would probably be non-registered and submitters registered. There will probably always be more searchers.

General Discussion - Board Room

DD - EO so far has been ascertaining what the average quality of submitted data is - both ancillary data and image data. Should we ask for annotated images?

AB - What about conflicting data submitted between two authors?

SB - The model about how these genes act may have to be changed - the contradiction may have to be resolved.

DW - Depending on staining protocols different data would be expected anyway.

AB - Should more details of the protocol be included in a submission to allow interpretation of these differences?

DW - For example hybridisation temperature.

JC - Only compulsory 'experimental' fields for a submission are probe definition and the original image.

MR - To overcome a lack of appropriate information in published papers, the GXD has asked several journals to modify their instructions to authors to ensure the information that is required for GXD is available in the publication.

DD - What about resolution of mapping?

DW - Mapping must be an approximation of the truth. Normal experiments require extra evidence such as double in situ with known marker genes to map to a particular tissue.

AB - The database will only be used as tool to retrieve information anyway. It's a tool to build models of how genes act/interact.

Particular Discussion Points

What data should be in the database before we go public?

Which gene to focus on?

Marker genes?

Focus on which organs?

Rapid population with EURExpress data - any other?

DW/SB - Functional groups, components of pathways and tissue markers should be the priority.

DW - Would it be possible to search for the whole functional group? (eg. Wnt pathway associated)

DD - We could use GO for this.

MR - Gene families are being developed at Jackson Labs.

DW - Would it be manual or interoperational?

DD - Envisage interoperational. Use info from single specialised databases.

DW - What about regulatory and downstream molecules? Would it be possible to find members of a synexpression group? Would these be annotated as such?

JC - Synexpression might only be apparent after entering of data.

DD - several interoperable resources might be able to deal with this.

RB - This is a challenge for the bioinformatics.

SB - Note genes with very similar expression patterns and similar nucleotide sequences.

MR - One aspect not covered so far is that areas of no expression will also be able to be noted (not currently denoted in (any) studies)

GK - Is raw data or 'added value' such as 'Wnt pathway' going to be in place by launch?

SB - This might require database changes.

Group consensus that this is probably a future direction.

OPT and 3D gene expression

DD - With very few OPT machines available generally, the EO and perhaps the HGMP-RC could act as a service for others to scan data or we could run a hotel service for researchers to use the OPT scanners themselves.

DW - What about other OPT machines in MRC units? NIMR would be happy to host a machine.

SB - As would Harwell.

MR - Is there funding for the EO to provide a service for scanning others' specimens by OPT?

DD - Probably but limited by numbers of machines.

RB - Warping in 3D is currently not available.

MR - Maybe we should wait on OPT until 3D warping is in place.

SB - What is MRC policy with disseminating OPT machines?

MD - MVM to develop this technology, but MRC can still distribute to other MRC units.

SB - What about training for OPT?

RB - Currently limited to HTML pages.

Other data types - proteins and reporters

SB - Protein screens - what about Allan Bradley's antibody screening?

DD - We have had discussions with Allan Bradley and John McCafferty about this.

DW - Reporters? Transgenics: If you could get all of the required information to define the sequence driving expression it could be possible. Knock-ins: again, enough information is required to detect possible artifacts.

MR - The GXD policy is to focus on endogenous gene expression patterns.

DW - Maybe EMAGE EO should also focus on curating only gene/protein expression but allow users to use the interface for reporters etc if they like (ie. EMAGE to support the appropriate interfaces (non-

curated).

DD - What about aligning RNA in situ and RNA microarray data?

AB - Waiting for data from Tom Freeman.

SB - Limus tool being developed between MRC Microarray Programme and HGMP-RC - should EMAGE also be involved in talks with these groups?

Linking to Journals

MR - Development and Developmental Biology make images available to GXD - GXD gets author approval as these are copyright. Hoping for journals to take electronic submissions of experimental data which will allow easier transfer of data to GXD.

AB - EMAGE could provide a service to journals to publish 3D images.

DD - What about the idea that data could be deposited in the database if peer-reviewed - only an abstract would appear in a journal (similar to Mechanisms of Development Gene Expression patterns section- has been in contact with Claudio Stern (Editor of Mechanisms of Development)).

DW - The Mechanisms of Development Gene Expression section specifically wants no function anyway therefore could be OK - unless it's too much effort for authors.

Links to the GXD

DD - Apart from anatomy browser, what should be the next link? Platform compatibilities?

RB - Would like to see complete platform integration so its not obvious to a user that two databases are being searched.

SB - Perhaps an issue for the future.

Embargoes

MR - If in conjunction with publication is not a problem but sometimes this can be an issue - GXD deals with these situations on a one by one basis.

Computation and Interfaces and Query requirements and Interoperability - see above

Requirements for local versions of the database?

AB - This would be a drain of resources. Only do this if large sets of data will come out of the collaboration.

Strategy for commercial users?

MD - If they were using this for data management they would want a private database.

PG - The HGMP-RC has a framework of issues to consider with respect to commercial users - see Appendix III.

Other questions

DW - What about links between mouse and human and chick databases?

Will it be possible to search several databases concurrently to find conserved/divergent aspects of gene expression?

DD - The collaboration with Tom Strachan's group in Newcastle is already in place for human, a loose

collaboration with Fons Verbeek (zebrafish) exists and talks have begun with Claudio Stern, Andrew Lumsden and Cheryll Tickle to apply for funding to set up a chick atlas and gene expression database. Interoperability is certainly envisaged. These links will apply to spatially mapped data as well as anatomy.

APPENDIX II

Notes on commercialisation issues from HMGP-RC (from Phil Gardner)

- * what 'marketing' should be carried out? (e.g. flyers in the HGMP-RC marketing pack, posters at conferences/workshops, purchase a mailing list, use HGMP-RC email list of commercial users etc.)
- * how to charge for a unique product (as there is no similar product on which to base charges)? (e.g. something measurable like disk space usage, CPU usage, megabytes transferred, connect time, Editorial edit-time, Editorial query-time, local 'seat' licences etc. etc. - or something more arbitrary like deciding what the market may bear...??)
- * important to avoid the academic mind-set (!) along the lines of "Well, these commercial companies often poach our staff, luring them away with high salaries, so let's sting them for all we can...!!" (i.e. can't assume commercial companies will pay high charges)
- * what would commercial charges be for? (e.g. to recoup development costs (unlikely?!), to cover running costs (only a bit less unlikely!), to get a bit of money back in rather than no money at all (likely?!))
- * multi-user licence with sliding scale? (e.g. company has 10-user licence with named users - who would maintain such a list? What about small-startups who might only have 1 user?) Offer a 30-day trial licence?
- * invoicing in advance, per annum? Pro-rata charging? (e.g. each invoice period starts 1st April, so charge company less if it registers part-way through the financial year...)
- * office overhead of maintaining licence user lists, chasing unpaid invoices, etc.
- * discounts to companies who are collaborators? (i.e. companies may add value to the project, therefore may want to negotiate a discount)
- * security issues? (e.g. companies may want assurances that mechanisms are in place to maintain secrecy of their searches and retrievals). Delayed release of data (e.g. pending patents, publishing of papers etc.)
- * sell access to the central HGU system, or sell the complete system for company to run at their site? Distribution and update implications, system support issues (e.g. when they update operating systems, or update Java/CORBA versions)
- * pricing of training courses? (i.e. course costs more for commercial attendees)
- * setting up a hotel facility? (e.g. companies come to HGU to use equipment and software, bring own samples and take away their own data)
- * commercial sponsorship of facility? (e.g. company logos and adverts discreetly in front-end software. Or donation of hardware...)