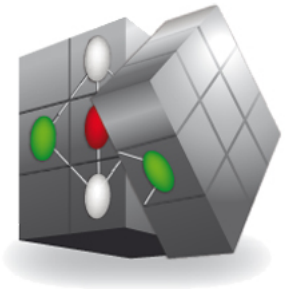


cubist

Your Business Intelligence

Initial mockup

CUBIST Deliverable D7.1.2

Kenneth McLeod
HWU
2011/06

SEVENTH FRAMEWORK
PROGRAMME

Table Of Contents

1. Administrative Details
2. Introduction
3. Mockup
4. Conclusion

Table Of Contents

1. Administrative Details
2. Introduction
3. Mockup
4. Conclusion

Summary

Mockup	
Abstract: In close collaboration with the creation of the overall CUBIST mockup in D1.1.3, this report contains a set of initial mockup screens for the expected user interfaces and a description of their functionalities.	
Type	Other
Document ID:	CUBIST D7.1.2.
Workpackage:	WP7
Leading partner:	HWU
Author(s):	Kenneth McLeod (HWU)
Dissemination level:	CO
Status:	Draft Version
Date:	17 June 2011
Version:	1.0

Versioning and contribution history
Reviewers

Versioning and contribution history

Version	Description	Contributors
0.1	Document content and basic content	Kenneth McLeod (HWU)
0.2	Extended gene based workflow	Kenneth McLeod (HWU)
0.3	Moved to powerpoint template	Kenneth McLeod (HWU)
0.4	Minor changes & corrections	Kenneth McLeod (HWU)
0.5	Clarification added following group teleconf	Kenneth McLeod (HWU)
0.6	Feedback from SAS	Kenneth McLeod (HWU)
1.0	Feedback from SHU	Kenneth McLeod (HWU)

Reviewers

Name	Affiliation
Simon Andrews	SHU
Martin Ursik	SAS

Table Of Contents

1. Administrative Details
2. Introduction
3. Mockup
4. Conclusion

Introduction

This is a use-case-dependent mockup concentrating on the biological use case. Earlier the use case was reviewed in detail in D7.1.1.

The current mockup is inspired by the prototype provided by SAP – see D1.1.3.

Mockup Example Setting

Although this document in not intended to be read in isolation and should be considered in parallel with D7.1.1, this section shall briefly review the key terminology and concepts from this use case.

A gene is a unit of instructions that directs, the body, how to do one essential task, i.e., create a protein. Gene expression information describes whether or not a gene is expressed (active) in a location. Broadly speaking there are two types of gene expression information: those that focus on where the gene is expressed, and those whose primary concern is the strength of expression. This work focuses on the former category, in particular a technology called *in situ* hybridisation gene expression.

Information on gene expression is often given in relation to a tissue in a particular model organism. Here the model organism is the developmental mouse. This organism is studied from conception until birth. The time window is split into 26 Theiler Stages (TS). Each stage has its own anatomy, and corresponding anatomy ontology called EMAP.

The result of an *in situ* assay (i.e. experiment) is recorded as an image displaying an area of a mouse (from a particular Theiler Stage) in which some subsections of the mouse are highly coloured. Areas of colour indicate that the gene is expressed in that location. Additionally, the image provides some indication of the level (strength) of expression: the more intense the colour, the stronger the expression.

Result images are analysed manually under a microscope. A human expert determines in which tissues the gene is expressed, and at what level of expression. As volume information is not the main focus of the experiment, its description uses vague natural language terms such as strong, moderate, weak, expressed or absent. For example, the gene *bmp4* is strongly expressed in the future brain from Theiler Stage 15.

Completed *in situ* gene expression experiments are published online. One of the main resources in this field is EMAGE. EMAGE documents the result of an experiment using a series of "textual annotations". Each annotation is a triple: gene - tissue - level of expression.

Table Of Contents

1. Administrative Details
2. Introduction
3. Mockup
4. Conclusion

This section shall discuss the mockup created for the biological use case. As such, it shall walkthrough the prototype, adding a commentary where appropriate.

In this mockup the starting point is a faceted search, which allows the user to select entities of interest. All facets (Assays, Genes, Level, Stage, and Tissues) are present from the beginning.

As most users' first interaction with EMAGE is by specifying a gene (or genes) of interest, the first workflow discussed in this section will do likewise.

Subsequently, a second workflow will focus on location (tissue) and time (stage) of expression.

Gene specific workflow

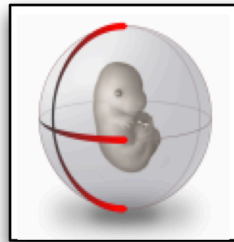
Slide 1 represents the EMAGE homepage for CUBIST. On the left is a list of the facets through which the data can be explored. The current list should be considered as minimal – it is possible to extend it by considering various factors such as quality, researcher, and experimental technique used.

Once the "Genes" facet has been selected, a window enables the user to select one or more genes in which they are interested – this is shown in slide 2. By default the option to query for all genes is selected. Notice that in this case the gene *bmp4* has been selected.

Mockup slide 1

What are you looking for?

Assays	38,418
Genes	15,587
Level	79,845
> Expressed	
> Strong	
> Moderate	
> Weak	
> Not expressed	
Stage	28
Tissues	13 381

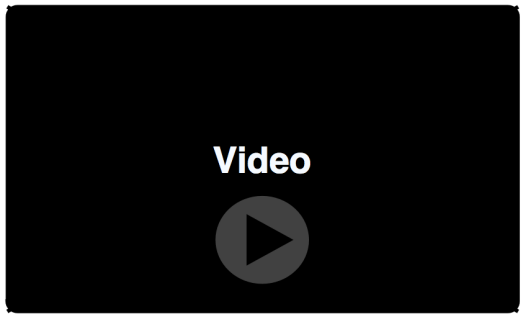


EMAGE

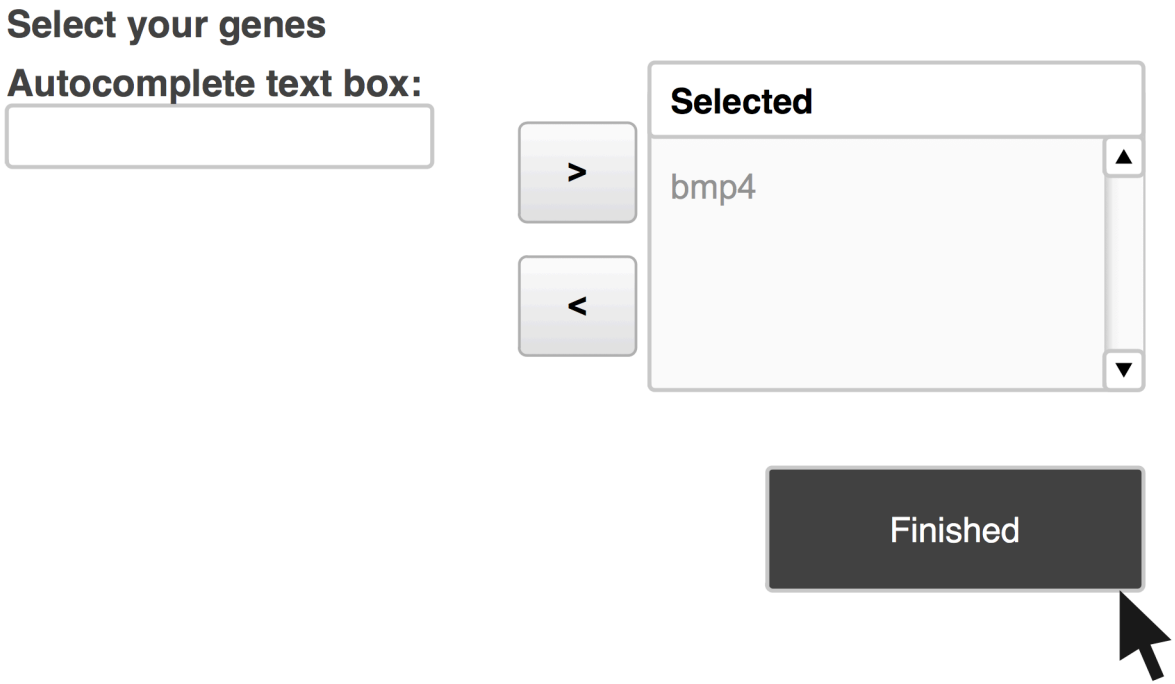
EMAGE is a database of *in situ* gene expression data in the mouse embryo and an accompanying suite of tools to search and analyse the data.

mRNA *in situ* hybridisation, protein immunohistochemistry and transgenic reporter data is included. These are sourced from the community and our curators take this data and describe it in a standardised way that allows data query and exchange. The description includes a text-based component but the unique aspect of EMAGE is its spatial annotation focus.

For more information see our introductory video below.



Mockup slide 2



Different ways of exploring the data associated with *bmp4* are demonstrated on slides 3 to 8.

Slide 3 displays a list of some of the assays that investigated *bmp4*. Each entry should be brief, but link to EMAGE's online page for the experiment enabling the user to retrieve all the available information if desired.

Mockup slide 3

Filter: genes: bmp4

VIEW: assay description assay navigator time location navigation

Explore facts

Assays

Level
> Expressed
  > Strong
  > Moderate
  > Weak
> Not expressed

Stage

Tissues

1. EMAGE:3829 - TS 17 - expressed in embryo

J Ramsay & D Fitzpatrick
Screen: MRC Human Genetic Unit

---

2. EMAGE:3826 - TS 17 - expressed in endolymphatic appendage epithelium

S Raft, S Nowotschine, J Liao & BE Morrow
Development (131):1801-12

---

3. EMAGE:3826 - TS 17 - expressed in non-endolymphatic appendage epithelium

S Raft, S Nowotschine, J Liao & BE Morrow
Development (131):1801-12

---

4. EMAGE:998 - TS 15 - not expressed in future brain

P Trainor, D Sobieszczuk, D Wilkinson, R Krumlauf
Development (129):433-42

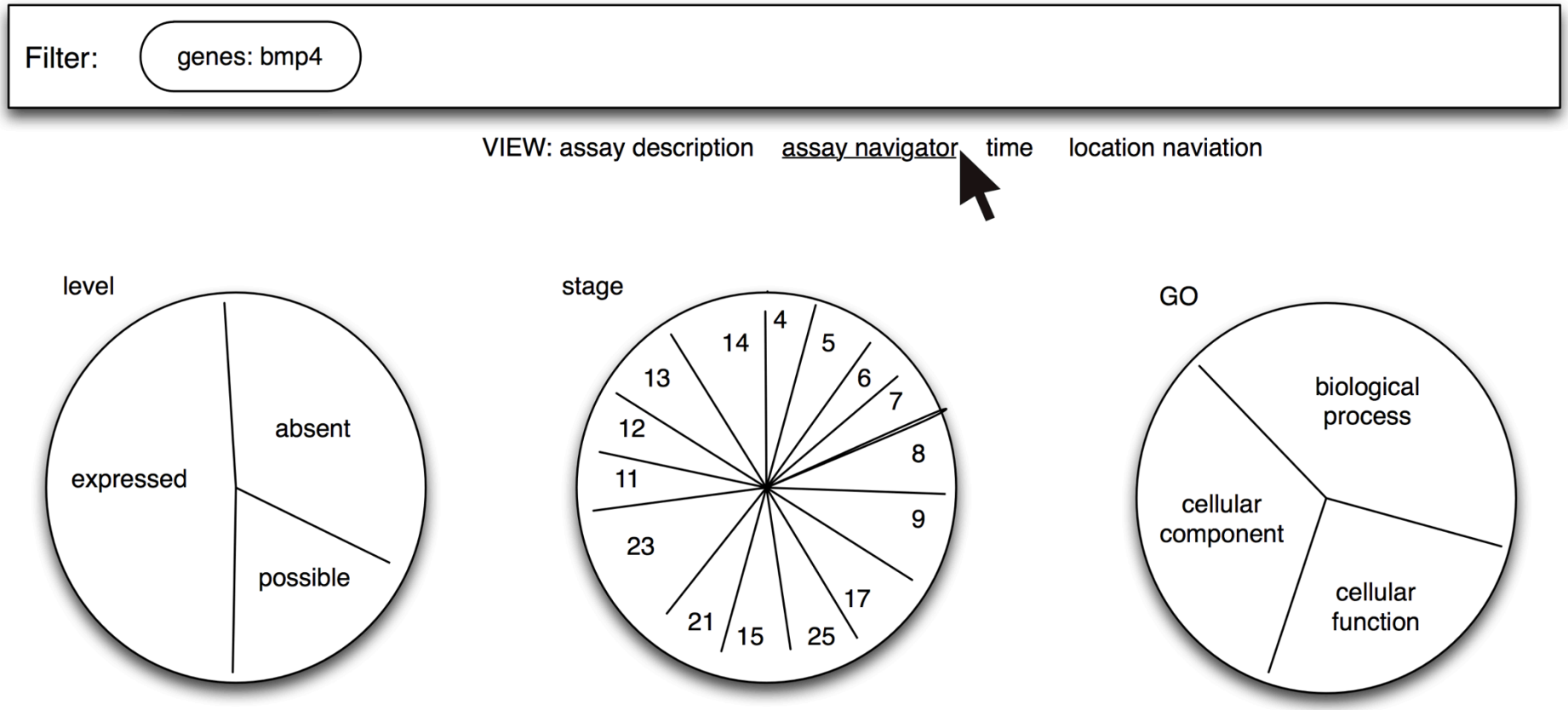
A series of complementary diagrams, which allow the user to delve into richer information, is presented on slide 4:

- Clicking on “expressed” (level chart) opens another layer in which the strong, moderate, and weak expression levels are shown. Clicking on “strong” will allow the user to see a list of all relevant experiments, or the list of tissues in which the gene is strongly expressed.
- Clicking on an individual stage, e.g. “22”, will produce either a list of experiments which are associated with the gene in the appropriate stage, or a list of tissues from that stage for which gene expression information is available.
- The third diagram is a sunburst diagram that enables the user to delve down into the Gene Ontology[1] examining the different annotations that have been applied to the current gene(s).

Only the first two diagrams are linked. Clicking on stage “22” causes the level chart to provide information for only stage “22”. Likewise, clicking on “absent” in the level chart forces the stage diagram to show only the stages in which *bmp4* is absent.

[1]An ontology that describes certain biological characteristics of genes - www.geneontology.org/

Mockup slide 4



The previous presentation mechanisms should be available regardless of whether the user is interested in one gene, or multiple genes. To illustrate this, the rest of this workflow will proceed with the genes *bmp4* and *ssh*.

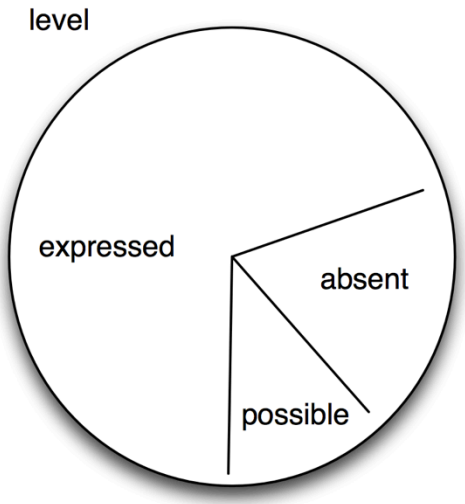
When multiple genes are featured the diagrams, from slide 4, display all the information for both genes; i.e., the GO diagram contains all the GO terms for *bmp4* and all the terms for *ssh*.

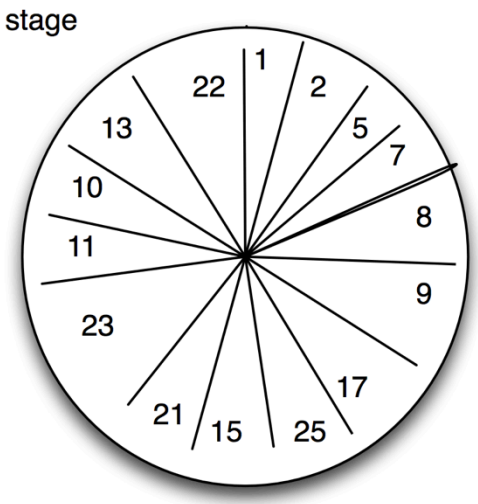
It is important to emphasize that the diagrams on slide 5 contain an aggregation, not an intersection of the underlying information from EMAGE.

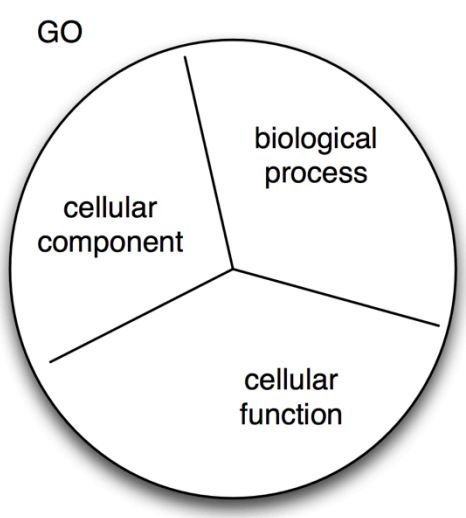
Mockup slide 5

Filter: genes: bmp4 genes: ssh

VIEW: assay description assay navigator time location navigation



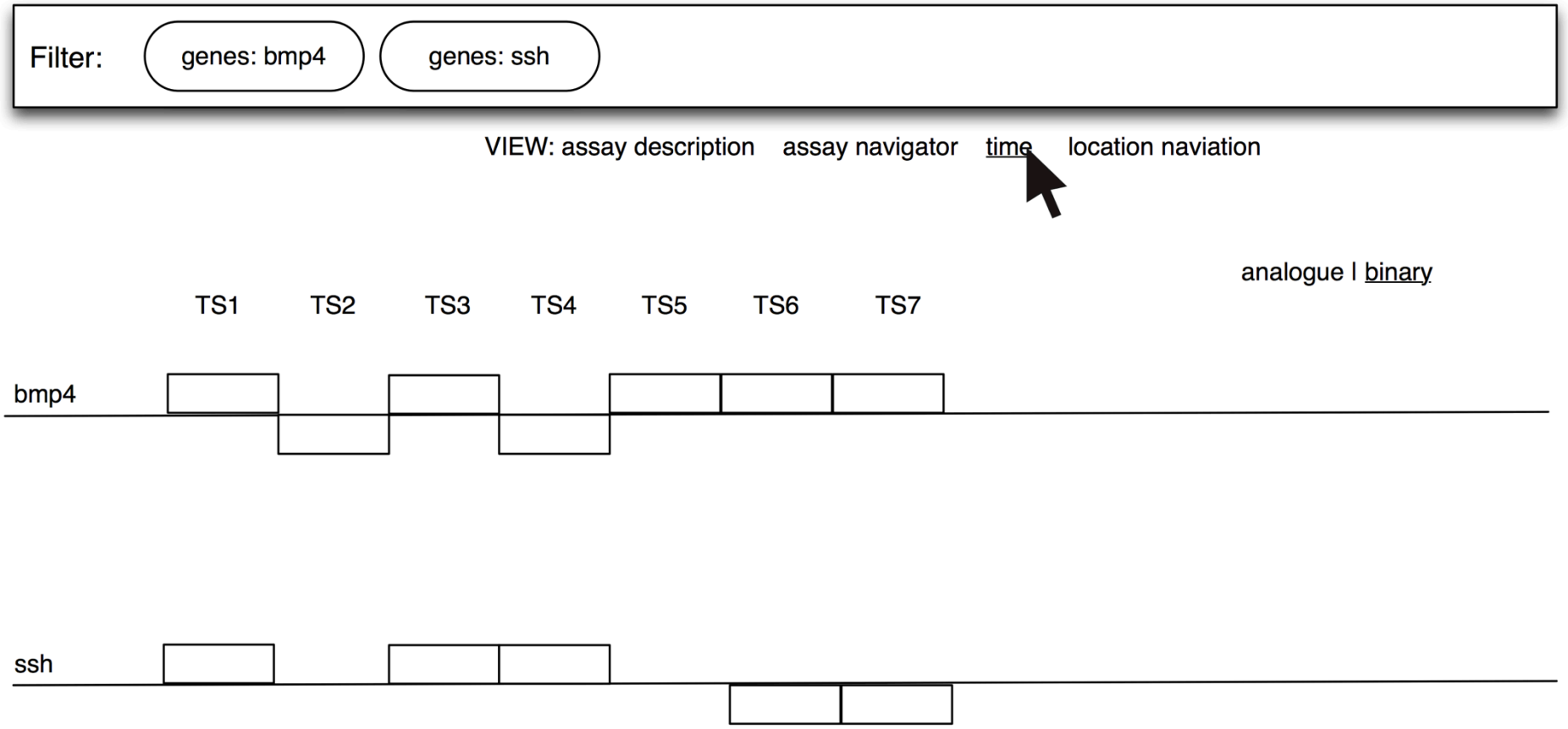




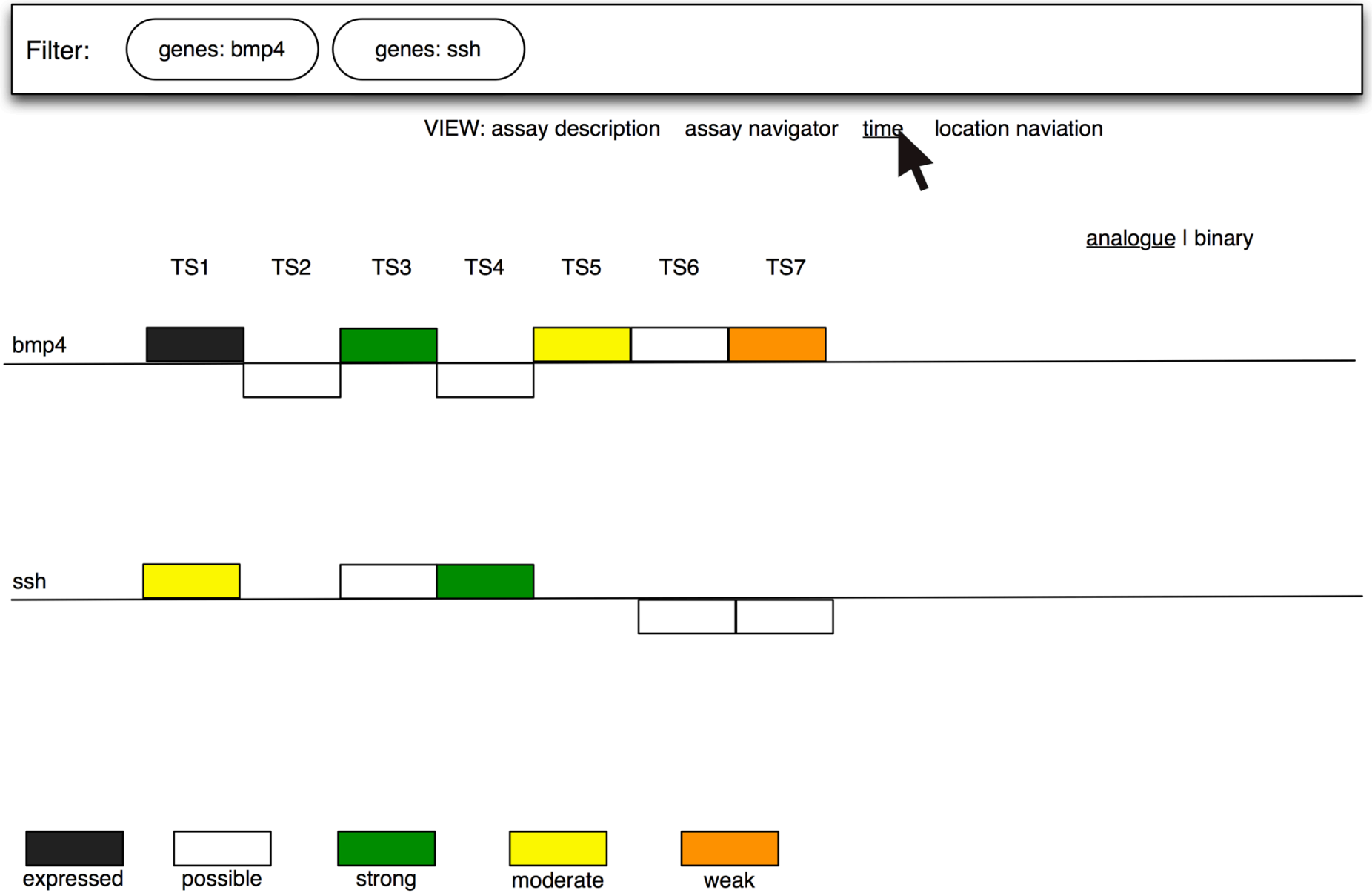
Slide 6 explores the level of expression over time. Genes are listed along the vertical axis, with the horizontal axis representing the notion of time (Theiler Stages). Positive bars indicate a gene is expressed (e.g. *bmp4* – TS1), negative bars suggest the gene is not expressed (e.g. *bmp4* – TS2), and the lack of bar demonstrates that EMAGE has no information (e.g. *ssh* – TS2). It should be possible for the user to restrict the time period displayed, or to view the entire development of the mouse (all 26 stages).

Slide 6 presents only binary gene expression. Analogue is on slide 7: coloured bars indicate the precise level of expression.

Mockup slide 6



Mockup slide 7



In order to explore the gene expression for a location, it is first necessary to specify a developmental stage as location is always w.r.t. a time. On slide 8 this is achieved through the use of a "spinner" - TS16 is chosen. A single stage is shown on slide 8; however, it may be useful to display multiple stages side-by-side.

Although a Hasse-diagram offers similar functionality, here a sunburst diagram is used to allow the user to drill down through the anatomy exploring the gene expression information. The depiction on slide 8 is an aggregation of the information for both of the selected genes. Tissues that are not related to the genes of interest, through an annotation, are not shown.

Due to propagation[2] tissues will appear in the hierarchy without an annotation directly proclaiming that a gene is expressed in that tissue. Accordingly, where an annotation suggests a gene is directly expressed in the tissue, the tissue is highlighted, e.g. ectoderm. Clicking on the tissue presents a summary of the annotation.

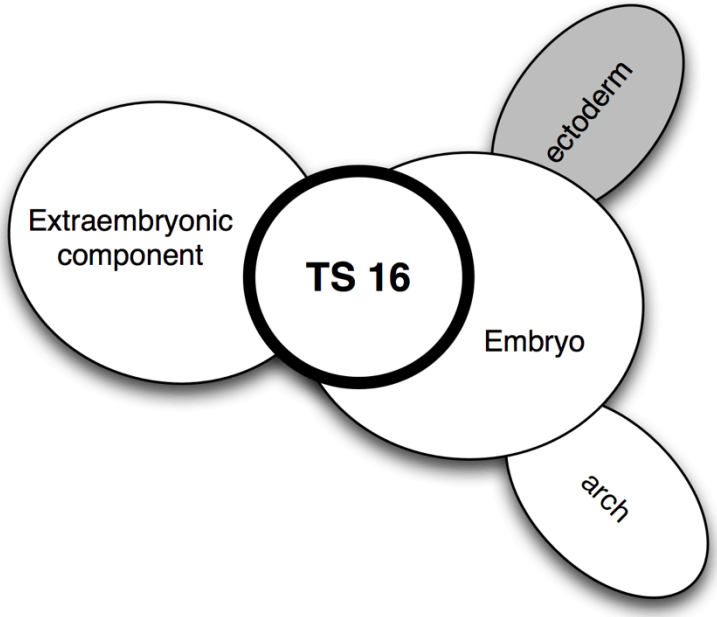
[2] If a *bmp4* is expressed in the paw, it is expressed in the limb because the paw is "part" of the limb. Thus *bmp4* is "directly" expressed in the paw and "indirectly" expressed in the limb.

Mockup slide 8

Filter: genes: bmp4 genes: ssh

VIEW: assay description assay navigator time location naviation

Stage: 16



Tissue & stage specific workflow

The focus of this workflow is location (tissue) and time (Theiler Stage). Here the user is interested in "all tissues in stage 16". Accordingly, slide 9 returns to the home screen to start a new query, and in slides 10 and 11 a tissue and stage are selected.

Subsequently, in slide 12, the anatomy of TS16 is represented as a sunburst diagram. This depiction allows the user to navigate through the anatomy of TS16, tissue by tissue.

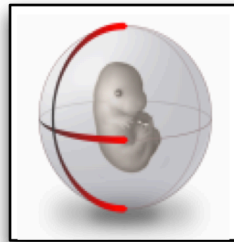
TS16 is in the centre of the diagram to emphasize that all the tissues in this diagram are from that developmental stage. Location information is always specific to a Theiler Stage.

Because the user is interested in "all tissues" the tree is complete (no tissues are missing). Right clicking the node presents the user with a menu containing a series of options.

Mockup slide 9

What are you looking for?

Assays	38,418
Genes	15,587
Level	79,845
> Expressed	
> Strong	
> Moderate	
> Weak	
> Not expressed	
Stage	28
Tissues	13 381

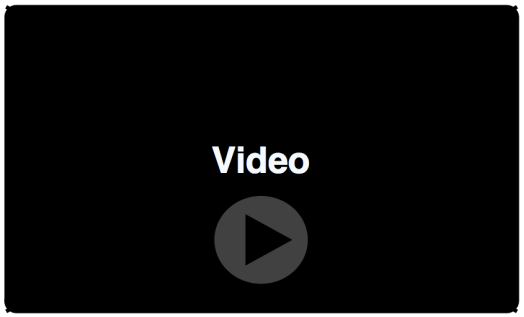


EMAGE

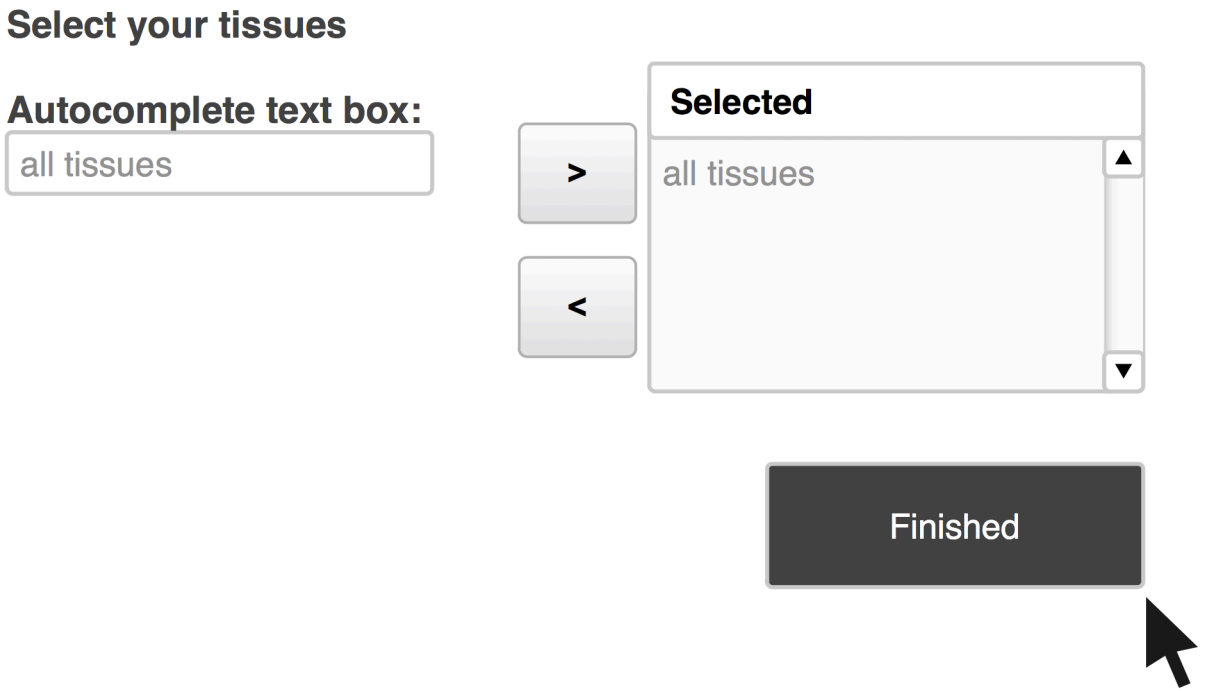
EMAGE is a database of *in situ* gene expression data in the mouse embryo and an accompanying suite of tools to search and analyse the data.

mRNA *in situ* hybridisation, protein immunohistochemistry and transgenic reporter data is included. These are sourced from the community and our curators take this data and describe it in a standardised way that allows data query and exchange. The description includes a text-based component but the unique aspect of EMAGE is its spatial annotation focus.

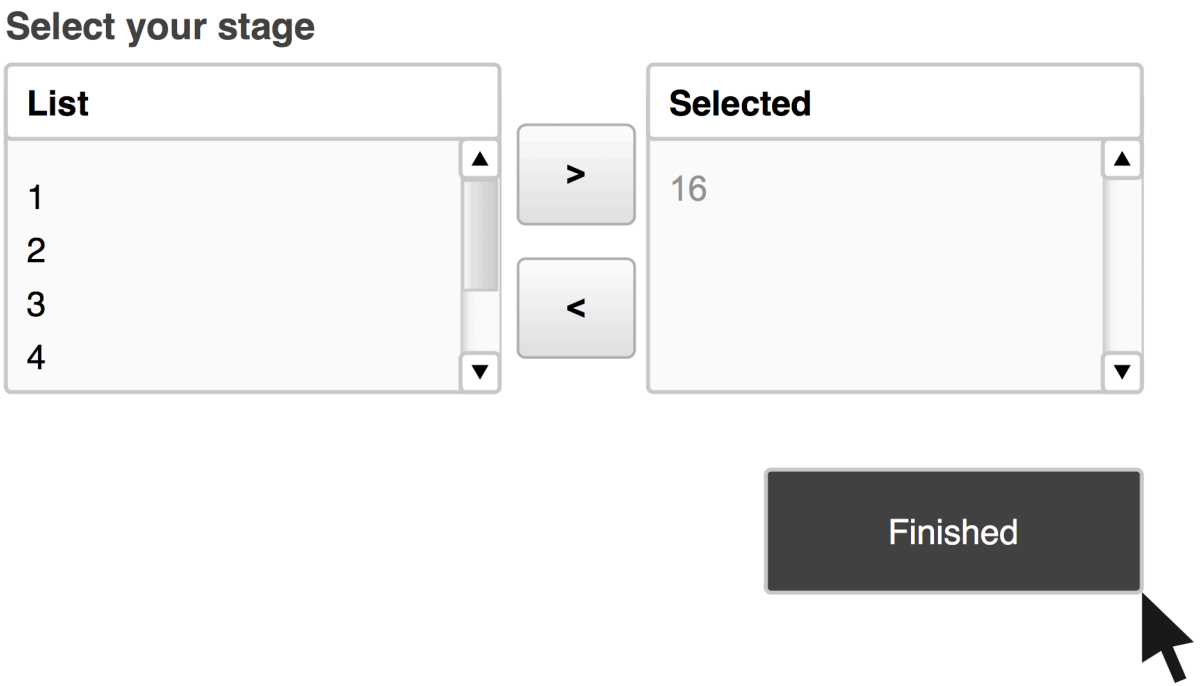
For more information see our introductory video below.



Mockup slide 10



Mockup slide 11



Mockup slide 12

Filter: tissues: all | stage: 16

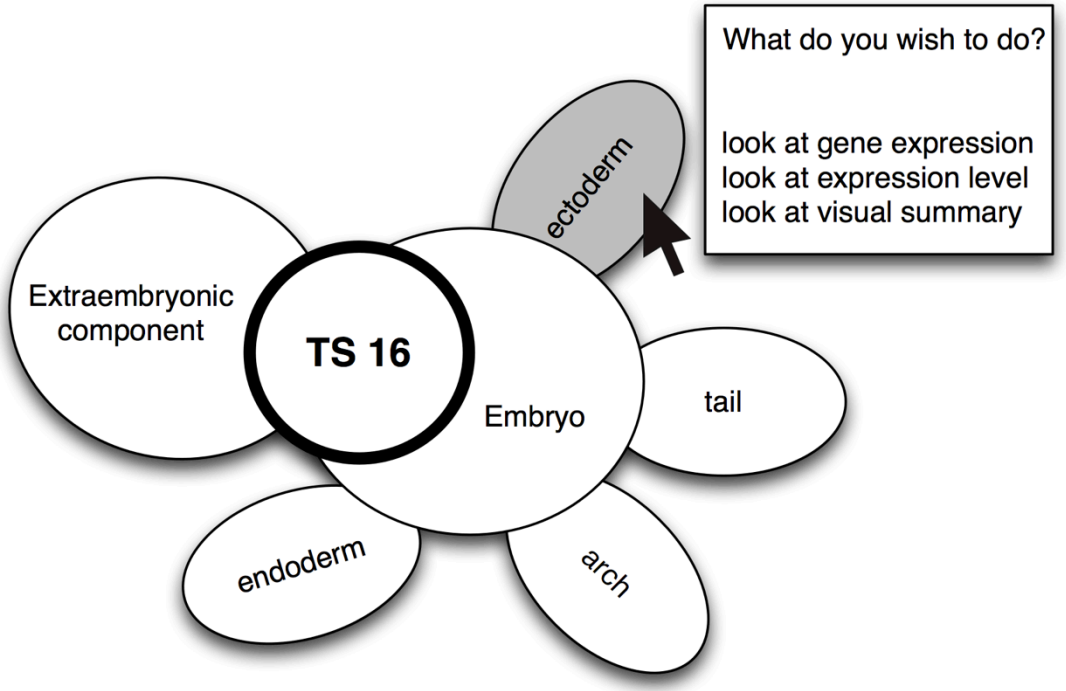
VIEW: anatomy browser visual summary

Explore facts

Assays

Genes

Level
> Expressed
  > Strong
  > Moderate
  > Weak
> Not expressed



In order to make the following slide more realistic a specific tissue has been selected. Accordingly, slide 13 focuses on gene expression for the future brain in TS16.

On slide 13 three different views of gene expression information are presented:

- The first is a simple list of the assays that have investigated the future brain in TS16. Each assay links to its web page in EMAGE.
- A level chart performs an analogously to the similar diagram contained in ``Mockup slide 5''.
- Finally, a Hasse-diagram enables the user to quickly visualise gene expression data for the sub-components of the future brain. Initially, "expressed" data is used.

The level chart is linked to the list such that clicking on a level of expression changes the list to show only the assays that provided this expression information. Note: the list does not effect the level chart.

The Hasse-diagram changes with the level chart too: e.g., clicking on "absent" causes it to display only the absent annotations. Again, there is no reverse link.

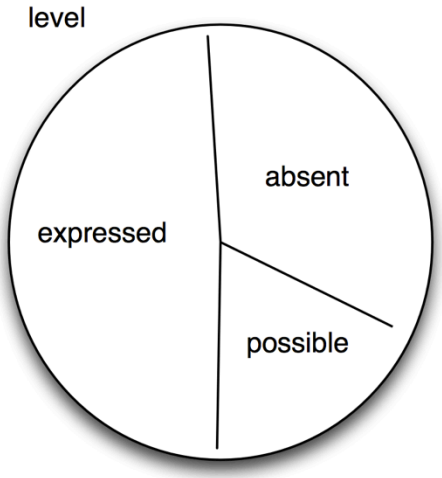
Mockup slide 13

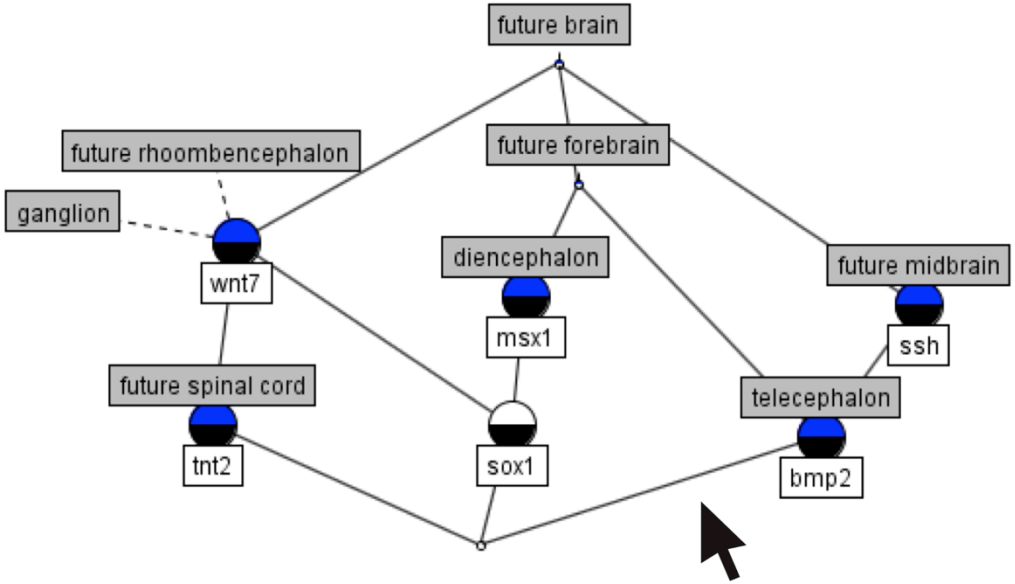
Filter: tissues: future brain | stage: 16

VIEW: anatomy browser visual summary

Assays

EMAGE:1234
EMAGE:2234
EMAGE:1765
EMAGE:6006
EMAGE:1963
EMAGE:5668
EMAGE:992
EMAGE:78
EMAGE:6007
EMAGE:6008





More complex Hasse-diagrams

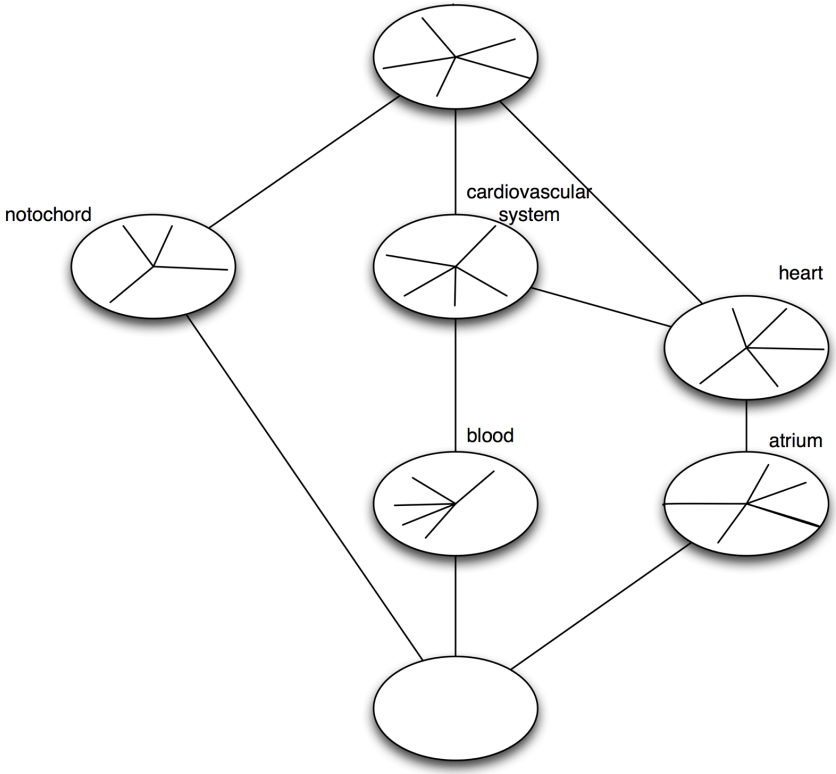
Whilst the Hasse-diagrams previously featured in this document are appropriate for biologists, another persona from D7.1.1. would prefer richer diagrams containing more information.

Computational biologists have a background in mathematics and often computer science. Unlike the biologists they are comfortable with complex graphical representations of data. Accordingly, extra information may be conveyed through the Hasse-diagrams.

For example, in the left hand diagram on slide 14 the nodes represent locations, and for each location there is a pie chart indicating the level(s) at which the gene(s) is/are expressed.

Alternatively, in the right hand diagram on slide 14 the size of a node is used to indicate the relative number of experiments linking the gene(s) to the tissue.

Mockup slide 14



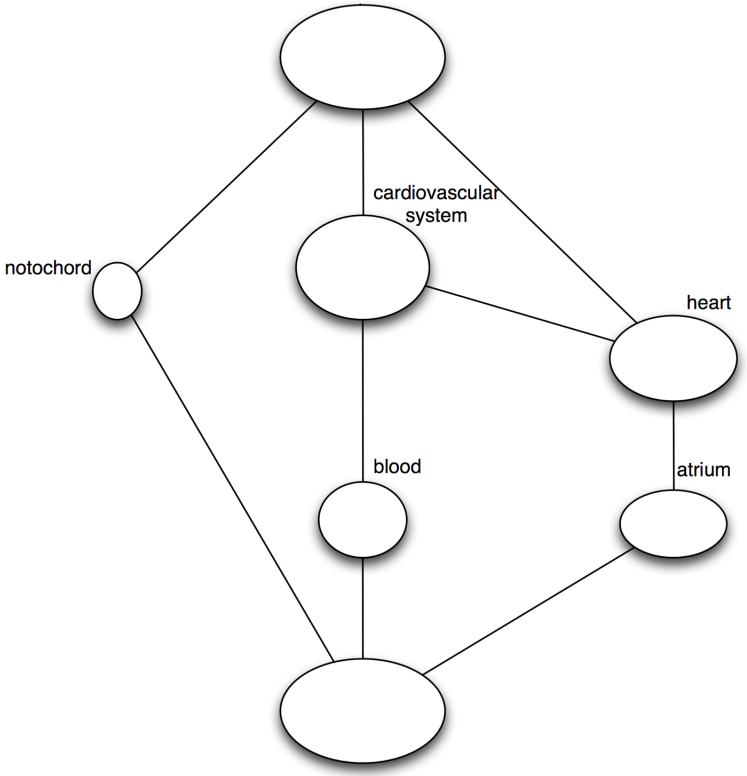


Table Of Contents

1. Administrative Details
2. Introduction
3. Mockup
4. Conclusion

Conclusion

This document has presented the so-called “mockup” of the user interface for the biological use-case. When read in conjunction with deliverable D7.1.1, this document describes the requirements for CUBIST from the perspective of the biological use case. In particular, this deliverable helps determine the visualisations required by a potential biological end user in order to fully utilise the functionality proffered by CUBIST.