

Report of Phenoscape Scientific Advisory Board

Chapel Hill, Friday 27th April 2012

General

The Advisory Board were impressed by the progress made on the project so soon after startup and by the range of expertise and competence of the members of the project. We share their enthusiasm for the project, the research it will facilitate, and the important technical innovations and applications it is creating and will create in the areas of data capture, aggregation and analysis. Following the very useful presentations and Q&A sessions the Board makes the following suggestions and recommendations.

Specific points

1. Gene expression data are being used by the project as surrogate for genomic alterations leading to changes in gene function and to changed underlying phenotypic changes in different taxa. We would like the Project to consider other large scale sources of structured data such as miRNA expression data which might be readily incorporated into the knowledgebase to enhance genome coverage and provide a first level of analysis of gene regulation.
2. We suggest that the output candidate genes might be further analysed by incorporating gene set and pathway analysis, functional analysis and for example PPI data into the Phenoscape site to provide users with enhanced value. These tools and resources are now widely available and would provide additional information for users to assess the candidates returned from phenotype queries.
3. User outreach and identification of user groups and real-life use-cases is now an important aspect of resource development. Using resources to explore these aspects of the project is now critical for project development, design of user interfaces and data integration. Whilst we are impressed by the educational aspects of outreach carried out to date we feel that more needs to be done to identify user groups and communities. Three approaches might be explored: User workshops at targeted meetings; Visits to particular individuals or institutions with an interest in what the project can do for them; and an expert workshop where specific individuals are invited in to work with

Phenoscape members and bring specific problems and queries from their research. The latter particularly should help recruit potential users and would not necessarily require the interfaces for example to be an advance state of completeness for this to work.

4. We are concerned that this project does not lead to a proliferation of cross species anatomy ontologies. Our recommendation is that the VAO and Uberon merge into one cross-species anatomy, which we think would be of benefit to Phenoscape and the wider community. There are issues of ownership, attribution and coordination involved in such a merger but we feel that there are solutions to all of these perceived problems, as with coordination and allocation of responsibilities and edit rights which can be sorted out through negotiation. We also felt that this might be a good opportunity to change the name of Uberon which is perceived to be problematical in some quarters. A new name for the merged resource would mark the beginning of the new collaborative ownership and should reflect more accurately the role and function of the site.
5. We would like some measures to be made of the quality of the phenotype data capture and annotations and some kind of quality control metric to be developed. The issues of inter-curator variability were discussed at the meeting and there may be some advantage in developing more formal rules for manual curation to reduce inter-curator variability. However a useful QC metric might be the ability to recover phylogenetic relations from the phenotype annotations using the knowledgebase, for example using ROC analysis.
6. The project does not seem to be capturing variation in the sense that most taxonomists and population geneticists use the term, i.e. the inter-individual variation within taxa and often required to characterize the taxon and/ or use the date in generating species groups of phylogenetic relationships, and to relate phenotypic effects of gene mutations to the normal phenotype. We felt that capturing such information will expand the utility of the project and recommend that where it is available it be captured formally..
7. We would additionally like the project to examine the practicality and utility of capturing data from birds and humans - particularly in humans there is direct phenotype/genotype data on limb malformations and these could be very useful. In birds the extensive data sets on gene expression patterns in the limbs could be mapped to the new VAO/Uberon ontology and incorporated into Phenoscape. Birds also would open up links to dinosaurs, providing access to another major evolutionary transition.

8. At this point we would like the project to consider the relative distribution of effort into devising and developing tools and data capture and curation. With limited resources efforts put into data capture may be more important for recruiting users and identifying use cases than the advanced development of tools. This should be examined.
9. We have some concern over the current use of ontology term brokering with the NCBO tools, which seems to be attracting a great deal of effort, when the key requirement is the response time of the ontology curators and not the sophistication of the brokering tools. We are concerned that a disproportionate amount of effort might be invested in sophisticated solutions to these problems and this should be reviewed carefully.
10. It was clear that improvement in the granularity of phenotype descriptions could be achieved with further development of PATO and we encourage Phenoscopiae to interact closely with the PATO developers to get the terms they need.

Points 1,2,3 and 9 are the most pressing issues at the moment and should be addressed over the next year of the project.

We thank the Phenoscopiae team for an excellent meeting and organisation and look forward to further progress and interactions over the coming years.

Brian Hall

Alan Ruttenberg

Peter Vize

Paul Schofield (Chair for the April 2012 meeting)