Abstract

For my summer project I will employ Bioinformatics tools to generate a resource that integrates ontologies and databases. I will be using database tools and programming languages to achieve this task. By integrating ontologies a better understanding and more thorough analysis of biological information may be attained.

Specific Objectives

The objectives for the project are as follows:

1. To map the intersection between the Cell Cycle Ontology (CCO) resources and the Gene Ontology (GO).
2. Adapting scripts and applying reasoning to investigate the potential of reasoning over the mouse data in the Gene Ontology as following recent publicized method.
3. Creating a file with all of the proposed changes to the GO and posting it on the Sourceforge.net site.

Significance

Mouse Genome Informatics (MGI) is the international database resource for the laboratory mouse, providing integrated genetic, genomic, and biological data to facilitate the study of human health and disease [1]. The Gene Ontology (GO) project provides a controlled vocabulary to describe gene and gene product attributes in any organism [2]. Mouse genes are functionally annotated in the MGI resource using the GO. The Cell Cycle Ontology (CCO) is a self-contained ontology focused on the representation of cell cycle knowledge [3]. The Cell Cycle Ontology has not been integrated with either the Gene Ontology or the MGI database. That means that the data and experiments used for the CCO have not been cross-referenced with data in the GO. It is possible that the data in the CCO could match that in the GO or even in the MGD. With the cell cycle (in CCO) mapped for humans, any orthologs of the human genes to the mouse genes help researchers understand cell mutations in mice during cell division and replication. Integrating these two databases will allow for inference and knowledge analysis on a whole new level considering the cell cycle of the mouse genome. This project is significant because it is another step in expanding and furthering the knowledge and data of the Gene Ontology and the Mouse Genome Database.

Previous Work

This project is another step in the unification of biological terms. The work already done on for this project includes the ontologies that the project will encompass: the Cell Cycle Ontology, the Gene Ontology, and the Mouse Genome Informatics Database. The OBO Foundry is an organization that helps support and enhances these and other ontologies. An article about the OBO Foundry in *Nature Biotechnology*, said the following:

The value of any kind of data is greatly enhanced when it exists in a form that allows it to be integrated with other data. One approach to integration is through the annotation of multiple bodies of data using common controlled vocabularies or 'ontologies'[4].

Currently, ontologies are very spread out in that the all have very similar terms and knowledge and there is little conjunction and linkage between them. For instance, GO has the term 'lung development,’ the Mammalian Phenotype has 'abnormal lung development,' and the Adult Mouse Anatomy has ‘lung.’ All of these terms should be linked intentionally somehow, but they are not. They are instead only semantically linked. The intent then is to attempt to use published scripts and the representation of the cell cycle to examine intersections between a variety of ontological resources.

Methods

1. Gathering bioinformatics data: This will require the review of publicly available data resources. These resources will most likely include the Jackson Laboratory Library and the scientific journals there as well as open-source ontologies on the web.
2. Employ Perl, a programming and scripting language, which is used very often in bioinformatics, for data organization and analysis tasks. Using Perl I will learn about regular expressions, an important part of scripting languages.
3. Learn and use basic database concepts to create a simple database. I will use Microsoft Access and use SQL (Structured Query Language) for simple data queries.
4. Utilize an application called “OBO-Edit” for ontology reviewing and editing.
5. Make use of other GO tools to track and save my research data.

Summary

As a result of completing this project, I will learn how to utilize and understand informatics. I will contribute to the database at MGI. I hope that the work I submit to the GO project will provide the members of MGI a new method of data analysis using the GO.

References

1. MGI-Mouse Genome Informatics-The international database resource for the laboratory mouse. 18 June 2009 <http://www.informatics.jax.org/>.

2. The Gene Ontology. 18 June 2009 <http://www.geneontology.org/index.shtml>.

3. Cell Cycle Ontology. 18 June 2009 <http://www.cellcycleontology.org/home>.

4. "The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration : Article : Nature Biotechnology." Nature Publishing Group : science journals, jobs, and information. 18 June 2009 <http://www.nature.com/nbt/journal/v25/n11/full/nbt1346.html>.