The Gene Ontology Newsletter

Issue No. 2 August 2006

Using GO to Evaluate Interactions and Networks

Interaction networks derived from high-throughput (HTP) methods, such as proteome-wide purification of protein complexes using affinity-tagged proteins, have become important resources for predicting gene function and studying complex biological networks, but the specificity of these networks has been difficult to assess. Data published in a recent paper from Mike Tyers' lab (University of Toronto) and incorporated within their BioGRID resource (http://www.thebiogrid.org) compares the results of HTP datasets versus that of individual experiments described in the literature. The paper (Reguly, *et al.*, *The Journal of Biology*, 2006;**5**(4):11) describes a comprehensive dataset of genetic and physical interactions for the budding yeast *Saccharomyces cerevisiae*. Using GO Biological Process annotations curated from the primary literature by SGD (http://www.yeastgenome.org/), the authors analyzed the extent of shared GO annotations between interaction pairs defined in the literature-curated (LC) dataset versus those defined in HTP datasets. The results indicate that inclusion of the LC dataset within their analysis significantly improves gene function prediction. The LC dataset assembled by Reguly et al. thus provides not only a valuable new resource for annotating a gene's role in biology, but an important benchmark by which the accuracy of HTP datasets can be measured. The complete LC dataset is available from BioGRID and mirrored at SGD.

Annotation Camp 2006

Discussion at the 3rd GO Annotation Camp focused on the consistent use of evidence codes and resulted in the recommendations for refining the usage for several codes. A revision of current evidence code documentation is under way in an effort to clarify the use of codes for new groups and improve annotation consistency between current groups. In addition, scientists and curators from more than 20 organism databases and scientific communities were trained to manually curate GO annotations from the primary literature. Thanks to Incyte Genomics and the Department of Genetics at the School of Medicine, Stanford University, for supporting this meeting.

OBO-Edit Released

OBO-Edit, formerly known as DAG-Edit, has been rebuilt from the ground up, with improved speed, stability, and a more intuitive interface. New features include support for OBO 1.0 and 1.2 specifications, basic reasoning capabilities, cross-product editing, full user's guide, and bug fixes. OBO-Edit can be downloaded from SourceForge:
https://sourceforge.net/project/showfiles.php?group_id=36855

Upcoming Changes

Removing 'unknown' terms: The 'unknown' terms in GO — biological process unknown (GO:0000004), molecular function unknown (GO:0005554) and cellular component unknown (GO:0008372) — will be removed from GO on October 16, 2006. Current annotations to 'unknown' terms, which are made when curators have looked at all of the published literature about a given gene product and failed to establish its function, biological process or cellular component, will migrate to the top level terms — biological process (GO:0008150), molecular function (GO:0003674), and cellular component (GO:0005575). Annotation to these terms will retain the evidence code ND [No Data].

New file in OBO 1.2 Format: Beginning September 15, 2006, the GO consortium will make available an additional ontology file in OBO 1.2 format. The main difference between the files will be that the replacement terms for obsolete terms will be specified by tags in gene_ontology_edit.obo, rather than in the comments field as currently in gene_ontology.obo. For more information see the OBO 1.2 format specification (http://www.geneontology.org/GO.format.obo-1_2.shtml). The current gene_ontology.obo file in OBO 1.0 format will continue to be produced.

Upcoming meetings

GO Users Meeting
(in conjunction with MGED 9)
September 10, 2006
University Tower Hotel
Seattle, WA
http://www.geneontology.org/meeting/mged-2006-meeting.shtml
http://www.mgedmeeting.org

Genome Informatics 2006
September 13 – 17, 2006
Wellcome Trust Genome Campus
Hinxton, UK
http://meetings.cshl.edu/meetings/infouk06.shtml

Capturing Biology

The GO Consortium has been improving the representation of several major areas of biology in the ontologies by meeting with experts in the field. The most recent meetings were:

Expansion of Central Nervous System Development: Curators from MGI (http://www.informatics.jax.org/) and ZFIN (http://www.zfin.org/) met with researchers studying central nervous system development to improve the representation of these processes in GO. This collaboration resulted in the addition of over 500 terms that reflect the development of the forebrain, the hindbrain, and the neural tube.

Representing Interactions between organisms: The collaboration between PAMGO (http://pamgo.vbi.vt.edu/) and the GO consortium continued with a jamboree to discuss new terms to capture processes such as the modification of host structures and microbe responses to host defenses. Terms representing the interactions of symbionts and hosts can be found in the 'interaction between organisms' node of the Biological Process ontology.

View Expanded Newsletter Online: http://www.geneontology.org/newsletter/archive/200608.shtml
To Receive Future Newsletters: Subscribe to the GO Friends mailing list (gofriends@genome.stanford.edu)
Contact the Gene Ontology Consortium: Please send comments or questions to gohelp@geneontology.org.