



New Directions for Postgenomic Bioinformatics

1405 Bernerd Place Rockville, MD 20851

www.Biomind.com

Enhancement of Gene Ontologies using Analytic Algorithms which Combine Microarray and DNA Sequence Similarity Data

Mining Microarray and Sequence Data to Enhance Gene Ontologies

Biomind LLC WHITE PAPER

January 2006

Objective:

Demonstrate the ability of Biomind's OntoloGenius™ software to enhance a gene ontology using advanced algorithms which mine both sequence and expression data.

Commercial Capability:

Collaborate with laboratories possessing significant microarray expression data in a model species to more accurately characterize its gene ontology.

Gene Ontology (GO) resources categorize a large number of genes according to their functions for both human and model-organism genomes. However, these ontologies are far from complete—many genes are uncategorized or incompletely categorized, and a few are even mis-categorized. Strategies for improving gene ontologies have included literature text-mining, and, more successfully, gene expression and DNA sequence similarity analysis. These software based techniques work sporadically: some GOs are predictable with DNA sequence similarity while others more so with expression similarity. Biomind has developed a novel approach for Gene Ontology enhancement which combines microarray gene expression data and DNA sequence data in a nonlinear fashion.

Biomind's software algorithm is based on using genetic algorithms (a machine learning algorithm) to learn the best way to combine microarray expression and BLAST DNA sequence alignment results for enhancing the Gene Ontology for a particular organism. The website go.biomind.com contains an enhanced version of the human GO with additional, inferred ontologies for previously classified genes. However, the software may also be used for inferring ontologies for previously unclassified and misclassified genes. In this way, the gene ontologies of many model organisms can be enhanced with the analysis of either temporal or stationary microarray data.

Biomind's commercial ontology enhancement product is OntoloGenius™. This is a command-line software toolkit which can be applied to data from any organism with:

1. An existing gene ontology;
2. Sequence data; and,
3. A reasonably large collection of microarray data covering much of the genome in a variety of experimental conditions.

In the work reported at go.biomind.com, and recently submitted for publication, Biomind showed 2,386 new links between human genes and GO categories and 1,111 new links for the yeast ontology. These new classifications span the subcategories (subontologies) of *Biological Function*, *Cellular Component*, and *Molecular Function*. The software algorithm was constructed and tested based upon its ability to accurately predict well known gene to GO links. Biomind was 74% accurate in these tests.

Figure 1 shows the inferred genes for cell death GO (8219).

☐ GO View

- ☐ [GO:0050875 \(cellular physiological process\) \[311\]](#)
- ☐ [GO:0016265 \(death\) \[119\]](#)
- ☐ **GO:0008219 (cell death) [105]**

Type: biological_process
Definition: The specific activation or halting of processes within a cell so that its vital functions markedly cease, rather than simply deteriorating gradually over time, which culminates in cell death.
Synonyms: none
[See this term at AmiGO site](#)
Associated genes:

 - ☐ Inferred associations: 14
[NM_001069](#), [NM_001626](#), [NM_002169](#), [NM_002170](#), [NM_002171](#), [NM_006064](#), [NM_006086](#), [NM_006087](#), [NM_006900](#), [NM_016264](#), [NM_020040](#), [NM_021068](#), [NM_024013](#), [NM_178012](#)
 - ☐ Original associations: 30
[NM_000484](#), [NM_000957](#), [NM_001418](#), [NM_001423](#), [NM_001424](#), [NM_001425](#), [NM_001831](#), [NM_002190](#), [NM_005166](#), [NM_005241](#), [NM_005253](#), [NM_006570](#), [NM_014410](#), [NM_014618](#), [NM_015046](#), [NM_016653](#), [NM_021980](#), [NM_198712](#), [NM_198713](#), [NM_198714](#), [NM_198715](#), [NM_198716](#), [NM_198717](#), [NM_198718](#), [NM_198719](#), [NM_198720](#), [NM_199167](#), [NM_201413](#), [NM_201414](#), [NM_203339](#)
- ☐ **Subcategories**
 - ☐ [GO:0007569 \(cell aging\)](#)
 - ☐ [GO:0019835 \(cytolysis\)](#)
 - ☐ [GO:0012501 \(programmed cell death\) \[91\]](#)
 - ☐ [GO:0010198 \(synergid cell death\)](#)

Collaboration with Microarray Partners

Biomind wishes to collaborate with research organizations in possession of significant microarray expression data for humans or model organisms. The OntoloGenius™ algorithms grow more powerful (predictive) as they are supplied with more and more microarray expression data. Applied to the human genome, this may result in discovering that previously poorly-understood genes are actually biomarkers for diseases or toxic reactions. And model organism ontologies which have not been well characterized could gain new functionality for toxigenomic, pharmacogenomic and basic research applications.