GoMiner Quickstart Guide
Revised May 3, 2004
Contact barry@discover.nci.nih.gov

INTRODUCTION

- In gene expression microarray and other omic studies, one generally obtains a list of dozens or hundreds of genes that differ in expression between samples and then asks, "What does all of this mean biologically?" GoMiner displays the genes within the framework of the Gene Ontology hierarchy in two ways: (i) as a tree-like structure similar to that in AmiGo; (ii) as a compact "directed acyclic graph." The program then provides numerical analysis and links to important publicly available information sources. The twin aims are to provide functional coherence to the observations on individual genes and to generate hypotheses to guide future research directions. We provide here a practical sample "recipe" for performing most of the operations. It includes loading of the required files, interaction with the various graphics, and generation of quantitative and statistical results.

WHAT THIS IS

- A walkthrough using a *specific* example of real data
    - o GoMiner supports all organisms and identifier types that are provided in the GO Consortium database, as well as HUGO names
    - o This examples assumes that HUGO names are used as gene identifiers
    - o Details for using the Data Source, Organisms, and LookupSettings menus will differ for other organisms
- NOT a *general* user's manual
- NOT an installation guide
- NOT a substitute for getting a conceptual overview by reading the article.
- NOT an absolute indication of the order in which every step must be performed, but an example of one possible order of steps that does work

REQUIREMENTS

- Computer with Java1.3 or higher installed. See web site (http://discover.nci.nih.gov/gominer) for implementation-specific details.
- Internet browser such as Internet Explorer or Netscape with the SVG plug-in
- High speed internet connection
  (If this is too slow, instructions for installing the database locally are available on the web site)
- Cn3D 4.1 or higher, as a web browser helper application for viewing NCBI molecular structures. (Instructions for installing and configuring are available on the web site.)

INVOKING PROGRAM

- Open command line in Windows (Start->Run->Cmd) or Macintosh OS X and type: java `-Xms256M -Xmx256M` –jar gominer.jar

- Or double-click on gominer.jar icon in Windows or Macintosh OS X

TROUBLESHOOTING
- If you have difficulty installing or operating GoMiner, you can find troubleshooting and contact information on the GoMiner home page at http://discover.nci.nih.gov/gominer.

LOADING GO DATABASES
- Select "File" Menu and click on "Load GO Terms"
- Click on "Ok" at bottom of pop-up dialogue box
- Wait several seconds until pop-up dialogue box disappears

SELECTING ORGANISM SOURCE DATABASES [This selection enables the user to avoid the ambiguity that arises if the same gene symbol is used for multiple organisms, but with different GO annotations]
- Select "Data Source" Menu and select "UniProt (H. sapiens et al.)" by clicking
- Select Organisms Menu and select "H. sapiens" by clicking
- Select "LookupSettings" Menu and confirm that "Enhanced Names (UniProt only)" is selected

LOADING DATA FILES RELATED TO THE SPECIFIC EXPERIMENT
- In bottom left corner, click on "Browse" in "Dialogue" panel
- Choose total.gene [or substitute your own file containing the HUGO name, one per line, for each gene in the microarray]
- Click on "Query Gene File" on "Dialogue" panel
- Wait several seconds until progress indicator shows completion of loading and gene list appears in left panel
- Resize left panel if necessary
- Click on "Browse" in "Dialogue" panel
- Choose "under.over.2col" [or substitute your own file for genes with altered expression level in the microarray containing either (1) the HUGO name ("under.over.gene") or (2) the HUGO name[tab]integer (one per line, where integer = -1 for underexpressed, or +1 for overexpressed) ("under.over.2col")]
- Click on "Query Changed Gene File" on "Dialogue" panel
- Wait several seconds until numerical values in center panel are displayed

BROWSING THE GENE/GO HIERARCHY
- Click on small box containing "+" next to "biological_process"
- Click on small box containing "+" next to "death"
- Continue this process with each new category that appears until "apoptosis" has been opened
- Note that specific genes with altered expression level appear as members of "apoptosis" GO category

- Note that blue number indicates enrichment of proportion of changed genes in that GO category relative to overall microarray (the green and red numbers indicate the enrichment of under- and overexpressed genes, respectively). The enrichment is computed as

$$\dfrac{\left(\dfrac{changed\,genes\,in\,category}{total\,genes\,in\,category}\right)}{\left(\dfrac{changed\,genes\,in\,chip}{total\,genes\,in\,chip}\right)}$$

- This number may or may not be statistically significant, and care should be taken not to over-interpret it. For the overall enrichment (ie, the blue number), the statistical significance is computed and displayed as the one-sided nominal unadjusted p-value from Fisher's exact test.
- Select "View" Menu, and click to select "View All Genes" (this allows toggling between changed genes only view and total chip view)
- Wait several seconds for this toggle switch to take effect
- Repeat sequence of clicks in the middle panel to view both the unchanged and changed genes within the "apoptosis" GO category

CUSTOMIZING CATEGORY AND GENE VIEWS

- The default situation is to display only changed genes and the categories that contain them. Other views of genes and categories are customizable:
- Click on the text word "apoptosis
- Select "View" Menu, and select "View All Genes and their Categories" (this allows toggling between default changed genes only view and total chip view)
- Wait several seconds for this toggle switch to take effect
- Navigate as in previous section to view desired changed plus unchanged genes and GO categories
- Select "View" Menu, and select "Hide Genes" (this allows viewing of only the GO categories in without viewing the genes they contain)
- Select "View" Menu, and select "View All GO Categories" (this allows viewing of all GO categories, even those that contain no genes from the current experiment; these empty categories are indicated by gray rather than black lettering)

VIEWING DETAILED INFORMATION

- Click on the text word "apoptosis" [note that right panel indicates all occurrences of the text word, in this case just one instance of "apoptosis", in the GO hierarchy]
- Click on the gene name "BAD" just under "apoptosis" [note that right panel indicates all occurrences of the gene name, in all positions in which it occurs in the GO hierarchy]

- **OBTAINING A SUMMARY VIEW**
  - **Click on the " Summary View" tab in the right panel of the GUI**
  - **Click on the top of the column by which you want to sort the results**

- **Click on "p-values Changed" to sort categories by p-values for changed genes, in ascending order. The most significant categories will appear at the top**
  - **Click on a category of interest to navigate to this category in the middle panel of the GUI**

STRATEGIES FOR SYSTEMATIC SEARCHING

- Click on the gene name "ABL1" in the alphabetized gene name list in the left panel [note that right panel indicates all occurrences of the gene name, in all positions in which it occurs in the GO hierarchy]
- Click on "Search" in "Edit" Menu
- Type "*apoptosis*" in third window
- Click "Find" in right-most window [note all occurrences of GO terms containing "apoptosis" are highlighted in middle window]

QUERYING AN EXTERNAL DATABASE

- Click on "BAD" in middle panel so that the gene is now highlighted
- Right click on "BAD" in middle panel (a pop-up menu will appear)
- Click on "LocusLink," "PubMed," "MedMiner," "GeneCards," "CGAP," "NCBI_STRUCTURE," or "Entrez" in the pop-up menu
  - MedMiner will extract and organize relevant sentences in the literature based on a gene, gene-gene or gene-drug query. There are options for searching from either one year or from five years prior to the current date. The knowledgeable user can customize the search dates by editing the URL line in the web browser.
  - LocusLink will provide a relationship between GO and chromosome visualizations (accessed by clicking on the chromosome designation below "Position" in LocusLink)
  - LocusLink will also provide a wealth of other information, such as SNPs (click on purple square containing "V") and MGC (select Mm under "LocusID", click on "View Loci" and then on purple box containing "MGC")
  - CGAP will provide a relationship between GO and BioCarta and KEGG pathway maps
  - NCBI_STRUCTURE will present a 3D dynamic structure and associated color-coded amino acid alignment as implemented by NCBI. The helper application Cn3D is required, as described above in "System Requirements." Those genes that encode for proteins for which an associated 3D is available are indicated by a "3D" icon following the gene name in the middle panel of the GUI

GENERATING A DIRECTED ACYCLIC GRAPH (DAG) VIEW
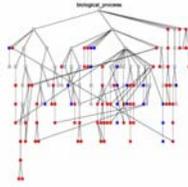
- Click on "biological_process" in middle panel so that the category is now highlighted
- Right click on "biological_process" in middle panel (a pop-up menu will appear)
- Click on "DAG Graph of Changed Genes" in menu
- Right click anywhere in browser window
- Click on "Zoom Out" in the pop-up menu
- Drag to center using simultaneous Alt-click
- View GO category name and changed genes by mousing over a node
- Click on a node to highlight all paths back to root
- Click on original GoMiner window to return to GoMiner

EXPORTING AN OUTPUT FILE WITH SUMMARY DATA TABLE
- Click on "Gene_Ontology" in middle panel
- Right click on "Gene_Ontology" in middle panel
- Click on "Export summary data to text file" in menu
- Select directory and output file name in popup dialogue box
- Click on "Save" (This may take several minutes, depending on the size of your file)
- Import tab-delimited data file into a spreadsheet program
- Examine quantitative results or apply your choice of statistical analysis
- Alternatively, click on "Export DAG of changed genes to file" to save an SVG file representing the DAG
- Alternatively, click on "Export Genes by Category" to save a tab-delimited text file of the category-gene relationships for the changed genes

EXITING
- Click on "Exit" in "File" Menu
- Click on "Yes" in popup dialogue box