# *What To Do After Getting Gene Expression Results?*

Below are some suggestions.

## Interpreting the Results

AltAnalyze creates three files when the analysis is finished:
1) Gene-expression summary file (aka DATASET file)
2) Pathway/Gene Ontology over-representation results
3) GenMAPP input file (aka GenMAPP file)

These files can be found in your results folder under "ExpressionOutput" (files 1 and 2) and "GO-Elite" (file 3). You can also click the button "Results Folder" to get here. All files are tab-delimited text file that can be opened in a spreadsheet program like Microsoft Excel, OpenOffice or Google Documents.

(File 1) The gene-expression summary file reports gene expression values for each sample in your input expression file. Along with the raw gene expression values, statistics for each indicated comparison (mean expression, folds, t-test p-values) will be included along with gene annotations for that array, including Ensembl and EntrezGene associations and Gene Ontology annotations.

(File 2) A file containing all biological pathways (Wikipathways.org) or GeneOntology.org (GO) categories that contain a disproportionate number of regulated genes, you can use the free, open-source program GO-Elite (http://genmapp.org/go_elite). The primary result file from this program can be found in the user selected output directory under "GO-Elite/GO-Elite results/combined-results_z-score_elite.txt". These results consist of a highly filtered set of GO terms and pathways, non-redundant with other GO-terms, gene annotations and mean fold changes for each gene in each pathway/GO-term.

## Down-Stream Analyses

At this point you have many options but some of the most common are:
1) Filter and sort the data in MS-Excel to find interesting genes.
2) Cluster the data.
3) Look for Gene Ontology terms and over-represented pathways.
4) Load the data in a pathway analysis program to see your data on pathways.
5) Find novel gene interaction networks using Cytoscape.

**Filter and sort the data in MS-Excel to find interesting genes**

AltAnalyze will have exported specific comparisons, such as cancer versus normal. If you have a dataset with at least 2 replicates in each group, a ttest p-value will also be calculated for each probeset. Selecting the menu option Data>Filter in Excel will let you search for specific criterion. Looking for a fold change >2 and p<0.05 will give you an initial list to examine in more detail. You will also want to sort by p-value or fold change by going to Data>Sort.

**Cluster the data**

If you have multiple comparisons in your dataset, you may be interested in looking for global similiarities in expression. One way to do this is to filter your AltAnalyze results and then analyze these data in a clustering program. You can filter your data however you want, however, one suggestion is to filter only the last two columns "smallest p" and "largest fold". By filtering for a maximum p<0.05 and minimum fold>1, you will capture all gene expression changes with a fold change > 2 or less than -2 and a p<0.05 in any comparison. Next you can copy and paste the filtered list into a new spreadsheet and import into a clustering program. Note, you will only want to include probeset or gene ID and individual array expression values (no summary statistics). Once imported, you may want to cluster the data in a program such as clusterMaker (http://www.cgl.ucsf.edu/cytoscape/cluster/clusterMaker.html).

**Load the data in a pathway analysis program to see your data on pathways**
Once over-represented pathways have been found or before doing this analysis, you can see which genes on which pathways are alternatively regulated in the program PathVisio or GenMAPP 2.1. PathVisio is a cross-platform analysis program, while GenMAPP is restricted to Windows. Both tools are easy use and have access to a large archive of curated pathways. An input file for either PathViso or GenMAPP is found in the directory "ExpressionOutput" with the prefix "GenMAPP-". For making pathways, PathVisio or WikiPathways is recommended, since these resources produce superior pathway content (valid interactions between genes and metabolite IDs) in the same format (gpml). PathVisio can also export pathways to the GenMAPP format. A PathVisio tutorial can be found at http://PathVisio.org while a GenMAPP tutorial can be found at http://GenMAPP.org.

**Find novel gene interaction networks using Cytoscape**
You can also use the program Cytoscape to create literature based networks and view your data on networks. Tutorials are present at Cytoscape.org.

If you have any other questions you can email us at: alt_predictions@googlegroups.com or nsalomonis@gmail.com