What To Do After Getting Alternative Exon Results?

Below are some suggestions.

Interpreting the Results

When AltAnalyze was running it produced a number of output files. These include:

1) DATASET-dataset_name.txt
2) comparison-splicing-index-exon-inclusion-results.txt
3) comparison-splicing-index-exon-inclusion-GENE-results.txt
4) comparison-splicing-index-ft-domain-zscores.txt
5) comparison-splicing-index-miRNA-zscores.txt
6) splicing-index-summary-results.txt

These files are tab-delimited text files that can be opened in a spreadsheet program like Microsoft Excel, OpenOffice or Google Documents. .

File #1 reports gene expression values for each sample and group in your probeset input expression file. The values are derived from probe sets that align to regions of a gene that are common to all transcripts and thus are informative for transcription (unless all probe sets are selected - see "Select expression analysis parameters", above) and expressed above specified background levels. 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 putative microRNA binding sites. This file is analogous to the results file you would have with a typical, non-exon microarray experiment and is saved to the folder "ExpressionOutput".

Results from files #2-5 are produced from all probe sets that may suggest alternative splicing, alterative promoter regulation, or any other variation relative to the constitutive gene expression for that gene (derived from comparisons file). Each set of results correspond to a single pair-wise comparison (e.g., cancer vs. normal) and will be named with the group names you assigned (groups file).

File #2 reports probe sets that are alternatively regulated, based on the user defined splicing-index score and p-value. For each probe set several statistics, gene annotations and functional predictions are provided. A detailed description of all of the columns in this file is provided here.

Files #4 and #5 report over-representation results for protein domains (or other protein features) and microRNA-binding sites, predicted to be regulated by AltAnalyze. These files include over-representation statistics and genes associated with the different domains or features predicted to be regulated. More information about these files can be found in the AltAnalyze ReadMe (section 2.3).

File #6 includes the number of genes alternatively regulated, differentiatially expressed and the mean number of protein residues differing between predicted alternative isoforms. This file is useful in comparing results between different pair-wise comparison files.

Down-Stream Analyses

At this point you have many options but some of the most common are:

1. Visualize AltAnalyze results in DomainGraph
2. Filter and sort the data in MS-Excel to find interesting genes.
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. .

Visualizing AltAnalyze Results in DomainGraph
The text file results produced by AltAnalyze can be directly used as input in the protein domain and microRNA binding site visualization program, DomainGraph. DomainGraph is a plugin for the Java program Cytoscape which can be easily loaded within Cytoscape after installation. For details, go to http://domaingraph.bioinf.mpi-inf.mpg.de.

Filter and sort the data in MS-Excel to find interesting genes
When looking at differentially expressed 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.

When looking at alternative exons, you will likely wish to identify alternative exons for validation. You may wish to prioritize based on those probesets that are predicted to alter the inclusion of protein domains/motifs or microRNA binding sites or select those that correspond to specific splicing annotations (e.g., alternative cassette-exons, intron-retention, alternative promoters). A useful way to prioritize genes is to also look at the gene level alternative exon file (file #3). Since this file has all of the exon-level data agglomerated to the gene level, you can find genes with relatively few or many alternative exons. You may also wish to sort or filter by the alternative exon score (dI) or splicing p-values (e.g., MiDAS or SI).

Look for Gene Ontology terms and over-represented pathways
To quickly determine if there are 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). GO-Elite can be run independently or from within AltAnalyze. In addition to performing typical over-representation analysis (ORA) this tools allows the user to run permutation tests on these results (to assess the overall likelihood of over-representation), filter-out redundant GO-terms and pathways for publication ready tables, summarize all AltAnalyze statistics at the pathway level and easily view gene annotations for each pathways. A tutorial for GO-Elite can be found here. Note: the most valid denominator gene IDs for ORA is the complete list of Affymetrix probesets present in the AltAnalyze gene-expression summary file (file #1).

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-". When viewing data for alternative exons in GenMAPP, you can format the data similiar to the GenMAPP file found in "ExpressionOutput". 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