Overview

The Throughput Ranking by Iterative Analysis of Genomic Enrichment (TRIAGE) platform is designed to facilitate prediction, analysis, and hypothesis generation from genome-wide perturbation studies like those designed with RNAi and CRISPR technologies.

TRIAGE is best utilized as tool for selecting candidates from high-throughput studies to be further validated by rigorous low throughput follow up studies. TRIAGE integrates analysis from gene to gene predicted interactions and statistical enrichments of known gene sets to correct for the false positive and false negative error rates that are associated with normalization methods that require arbitrary cutoffs. By sorting the data into low, medium and high confidence tiers the TRIAGE platform integrates the experimental readout of the analyzed study with *a priori* knowledge about biological networks and interactions.

Getting Started

Uploading Your Data:

To upload your data, ensure that your document is in .csv format and includes at least one of these two columns:

1. A column titled “GeneSymbol” with the HGNC gene names of your gene targets.
2. A column titled “EntrezID” with the NCBI Entrez ID of your gene targets.

Your document should also include a column with the numeric values by which medium and high confidence values will be selected. Each gene target should only have one value associated with it and listed in the document only once. The name of this column is up to the user. (Additional columns can be included in the document and they will be included in the output file at the end of the analysis.)

Running Your Analysis:

To run your analysis, select the parameters that describe your data and the databases you want to use for your analysis.

Select your organism: A drop down menu provides the option to select “Human” or “Mouse”
Select a pathway to enrich: Select a database of biological pathways to be used to analyze your data for gene set enrichment. Currently the Kyoto Encyclopedia of Genes and Genomes (KEGG) is the only available database and the default option.
Select a PPI database to use: Select a database of Protein-Protein Interactions (PP!) to use and the confidence score for the interactions to be included. Currently the experimentally predicted interactions from the STRING database are available with the option to select high confidence

predicted interactions (Confidence score ≥ 0.7) or medium confidence predicted interactions (Confidence score ≥ 0.4).

Choose an input file to upload: Upload your .csv file. A progress bar will inform you when your upload is complete, the data from your file will appear in a table under the "Input" tab.

Cutoff Type: One your file is uploaded a dropdown menu will appear with a list of the column names in your document. Select the column that contains the numeric values to be used for the high confidence/medium confidence cutoffs of your targets.

High-conf Cutoff Value: Enter a numeric value to be used as a cutoff for high confidence hits from your screen. Gene targets with values in the "Cutoff Type" selected column *equal to or greater than* the High-conf Cutoff Value entered will be used as High Confidence hits for the analysis.

Med-conf Cutoff Value: Enter a numeric value to be used as a cutoff for medium confidence hits from your screen. Gene targets with values in the "Cutoff Type" selected column *equal to or greater than* the Med-conf Cutoff Value entered and *Less than* the High-conf Cutoff Value entered will be used as Medium Confidence hits for the analysis.

It is crucial that not only the medium and high confidence hits from the screen be included in the upload document but all genes observed in the analysis be included as the background genes (i.e. low or no confidence hits) are critical for accurate statistical analysis of gene set enrichment.

Analyze my data: Once you have selected and entered the parameters for your analysis click the "Analyze my data" icon and your analysis will begin. A progress bar at the bottom right corner of your screen will show the progress of your analysis.

Reading Your Results:

<u>Enriched Pathways</u>

When the analysis is complete a list of enriched pathways will appear in a table under the "Enriched Pathways" tab. The list includes all pathways that have a statistical score of 0.05 or less in a two-tailed fisher's exact test.

The table provides the following information:
Pathway: The name of the enriched pathway. Clicking on the pathway will open a new tab from the KEGG database showing a schematic of the genes in the pathways with the gene hits from the analysis highlighted. Genes that were marked as high confidence at the start of the analysis are highlighted in blue and those marked as medium confidence are highlighted in red.

pVal: The p-values based on a two-tailed fisher's exact test for the enrichment of each pathway are listed.

pValFDR: The p-values with added correction for False Detection Rate are listed.

pValBonferonni: The p-values with the Bonferroni corrections for multiple testing are listed.

TotalGenes: The total number of genes in the pathway.

HitGenes: The number of hit genes as selected by the TRIAGE analysis that are in the pathway.

HitGeneNames: The HGNC Gene Symbols of the hit genes in each pathway, genes that were marked as high confidence at the start of the analysis are highlighted in blue and those marked as medium confidence are highlighted in red.

Gene Hits

Lists of Gene Hits: A table with the input document and the genes added or dropped at each iteration listed in new columns. Genes counted as high confidence hits in an iteration are given a value of 1, genes counted as medium confidence are given a value of 0.5. The last Network.class.iteration_ column includes the gene hits that are counted as final hits in TRIAGE.

Gene Hits By Iteration: A graph showing the number of medium confidence hits and high confidence hits at each iteration of the TRIAGE analysis.

Network

To drive further exploration of the data the TRIAGE platform makes it possible to view which of the gene hits identified by the TRIAGE analysis that are not known members of specific gene sets ("Novel" genes) have predicted interactions with known members of the gene sets ("pathways")

To explore this feature the user can select up to three pathways by clicking on the box near the pathway name, once the pathways of interest are selected click on the "Create Network Graph!" icon. A progress bar at the bottom right of the screen will appear

PathNet

Once the graph has been generated the interface will switch to the PathNet tab with two viewing options.

1st Dimension Network: A circular graph showing the hit genes from each selected pathway (separated by group and highlighted by different colors) and only the "novel" genes from the analysis that have predicted interactions with any of the "pathway genes". Hovering with the

cursor over a gene name will highlight its interaction path and the name of the gene(s) it is predicted to interact with. This feature can be used to generate exploratory hypothesis of novel mechanisms and interactions and to identify "missing links" between predicted biological processes generated by the analysis to be further validated by subsequent research.

2nd Dimension Network: A circular network like the 1st Dimension Network that also includes "novel" genes that don't show 1st degree connectivity with genes in the pathways of interest but show 2nd degree connections to the pathways via predicted interactions with other novel genes identified by TRIAGE that have predicted direct interactions with the pathways of interest. This feature can be used for my complex exploration for possible network reconstruction and to broaden the targets for further exploration.

Both graphs can be downloaded as interactive HTML files from the "Download" tab. An excel version of the information contained in the graphs with columns for sorting by number of predicted interactions to select or all pathways of interest is also available in the "Download" tab under the title "Ranking_HumanTNFScreen.csv"

Network D3

This page contains network graphs constructed based on the selections made in the "Network" tab using the D3 platform. This feature is still under development.

Download

The "Download all files" icon downloads a zipped folder of the analysis generated by the TRIAGE platform.

TRIAGEoutput_ALL.csv: a table listing all the genes from the analysis with their ranking in each iteration, the final iteration showing the list of genes that are selected by the analysis.

Chimera_STRINGHi_against_selectedPathways_1st.hits.html: An HTML file of the graph generated in the 1st Dimension Network analysis.

Chimera_STRINGHi_against_selectedPathways_2nd.hits.html: An HTML file of the graph generated in the 2nd Dimension Network analysis.

KEGG_TRIAGE_enrichment_final.csv: A table of enriched pathways identified by the analysis with their associate p-values, FDR values, Bonferroni values, and hit genes included.

Ranking_HumanTNFScreen.csv: A table of all the hit genes selected by TRIAGE analysis with their predicted membership or interactions with genes in the pathways selected under the "Network" tab.

Help

Contact us: You can contact the TRIAGE team with any issues you encounter by filling out the form under this tab.

Documentation: A detailed user guide and associated information (a updated version forthcoming)

Updates: (forthcoming)