WebGIVI

Version 1.0

Tutorial

Liang Sun

E-mail: sunliang@udel.edu

Introduction

A major challenge of high throughput transcriptome studies is presenting the data to researchers in an interpretable format. In many cases, the outputs of such studies are gene lists which are then examined for enriched biological concepts. One approach to help the researcher interpret large gene datasets is to associate genes and informative terms (iTerm) that are obtained from the biomedical literature using the eGIFT text-mining system. However, examining large lists of iTerm and gene pairs is a daunting task. We have developed WebGIVI, an interactive web-based visualization tool to explore gene:iTerm pairs. WebGIVI was built via Cytoscape and Data Driven Document JavaScript libraries and can be used to relate genes to iTerms and then visualize gene and iTerm pairs. WebGIVI can accept a gene list that is used to retrieve the gene symbols and corresponding iTerm list. This list can be resubmitted to visualize the gene iTerm pairs using two distinct methods. Rate-limiting genes, NCBI and eGIFT biomedical literature sentence databases have all be integrated into WebGIVI. Gene expression data can also be uploaded to color a Concept Map graph for further exploration.

The current version is "1.0".
It is freely available at http://raven.anr.udel.edu/webgivi/index.php (see Figure 1).
All the code is hosted on Github, you can fork it through the link: https://github.com/sunliang3361/webGIVI.
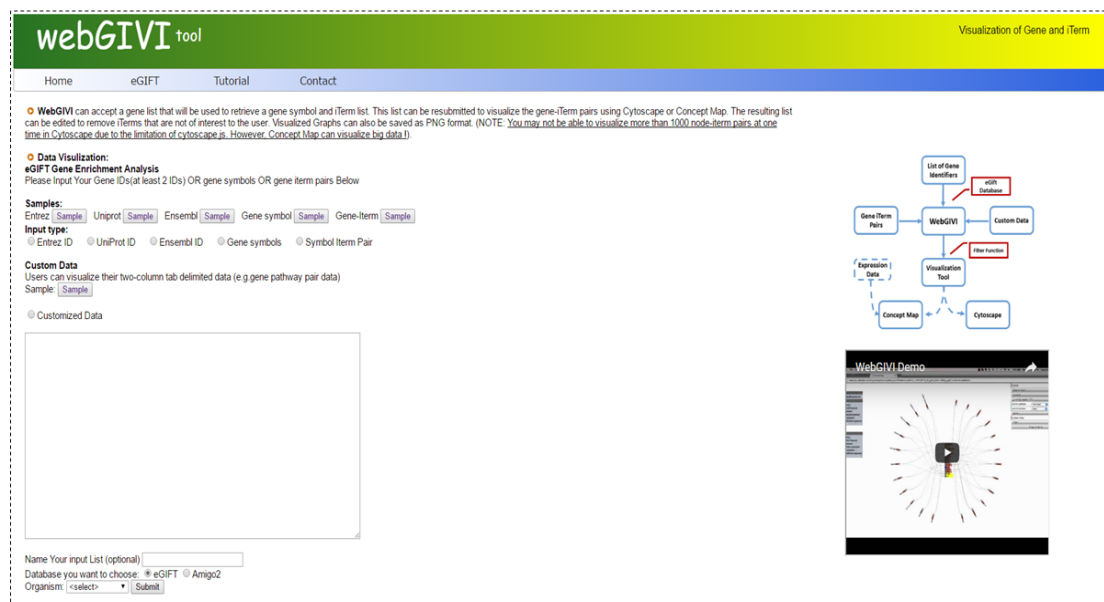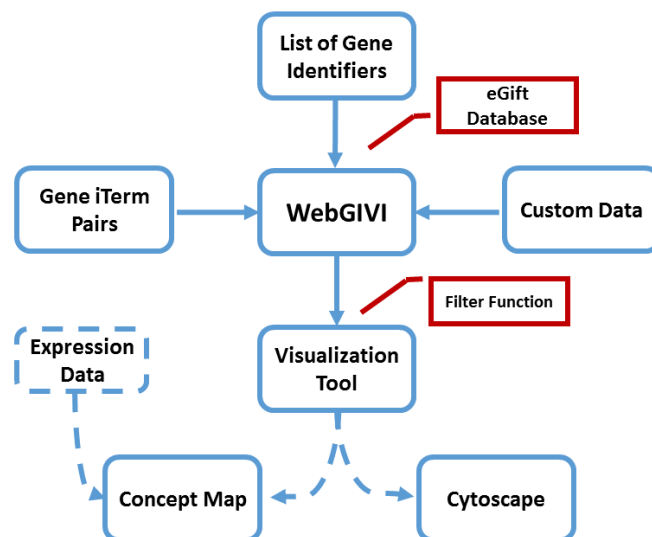

Figure 1. *Homepage of WebGIVI*

Brower requirements

All modern browsers, such as Safari, Google Chrome, and IE are supported. The recommended web browser to use is [Chrome](). Firefox is not currently supported and will result in function errors.

Introduction of WebGIVI Interface

1. **Input Data**
    Currently, WebGIVI can accept a gene list that is used to retrieve the gene symbols and corresponding iTerm list.

Entrez gene list sample: http://raven.anr.udel.edu/webgivi/sampleData/entrezID-sample.txt
Uniprot gene list sample: http://raven.anr.udel.edu/webgivi/sampleData/uniprotID-sample.txt
Ensembl gene list sample: http://raven.anr.udel.edu/webgivi/sampleData/ensemblID-sample.txt
Gene symbol list sample: http://raven.anr.udel.edu/webgivi/sampleData/symbol-sample.txt
Gene Iterm pair sample: http://raven.anr.udel.edu/webgivi/sampleData/gene-iterm-sample.txt
Custom data sample: http://raven.anr.udel.edu/webgivi/sampleData/custome-sample.txt

You can copy and paste the sample data to WebGIVI to test it.

2. **Prefilter Function**
    eGIFT provides a precomputed text mining database that has extracted all gene associated informative terms (iTerm). Not all iTerms are informative in all use cases, but could be important to others. For example, 'in situ hybridization' is an irrelevant iTerm to our case scenario but will be interesting to researchers who might want to apply this experimental method to their own work. It is beneficial to filter out irrelevant iTerms in the context of the

interests of the current study, and only save iTerms of direct relevance. Filter function with the following options:

- *Hover* function: display all associated genes when hovering your mouse on the top of iTerms (See Figure 2 (1))
- *Sort* function: Will sort all iTerms based on p value, alphabetical order or frequency (See Figure 2 (2))
- *Cut-off* defined: Highlight iTerms that fall into the user defined frequency cut-off (See Figure 2 (2))
- *Clear* function: Will clear up all selections (See Figure 2 (2))
- *Remove* function: Delete all highlighted iTerms (See Figure 2 (3))
- *View* function: Open all gene:iTerm pairs on the browser (See Figure 2 (3))
- *Download*: Download all gene:iTerm pairs to your local computer (See Figure 2 (3))
- *Concept Map* view: Visualize all gene:iTerm pairs via Concept Map method (See Figure 2 (3))
- *Cytoscape* view: Visualize all gene:iTerm pairs via Cytoscape method (See Figure 2 (3))

Figure 2 *Filter function in WebGIVI.*

3. Visualization

Either Concept Map (see Figure 3.2) or Cytoscape (see Figure 3.1) methods can be used to visualize gene and iTerms. For large input data set, using Concept Map method is highly recommended. Concept Map graph is mainly composed of three parts: Navigation panel (see Figure 3.2 (1)), data graph (see Figure 3.2 (2)) and GUI (See Figure 3.2 (3)).

Figure 3.1 *Cytoscape Graph. The red circles represent iTerms; the yellow circles represent genes. Right clicking nodes can link out to NCBI database; left clicking edges will link out to eGIFT sentence web page.*

Figure 3.2 *Concept Map Graph. Gene symbols are located in the center of the circle. eGIFT iTerms are located on the circle. The bar chart on the node represents the frequency of this node. The blue color on the gene symbol represents the rate-limiting gene.*

4. Click on the graph

Left and right clicking nodes or edges can delete, link out to NCBI and eGIFT databases and resize the node labels.

4.1 ***Link-out*** function

Left or right clicking nodes or edges can link out the genes or iTerms or gene:iTerm pairs of your interest to NCBI or eGIFT sentence databases (see Figure 4.1(1) & 4.2).

4.2 *Remove* function

Delete the iTerms or genes that are not informative (see Figure 4.1(2)).

4.3 ***Resize*** node label

If label names of nodes are too long and therefore overlapping with other nodes, you can adjust the length of node labels in the graph via setting up how many characters you want to display (see Figure 4.1(3)).
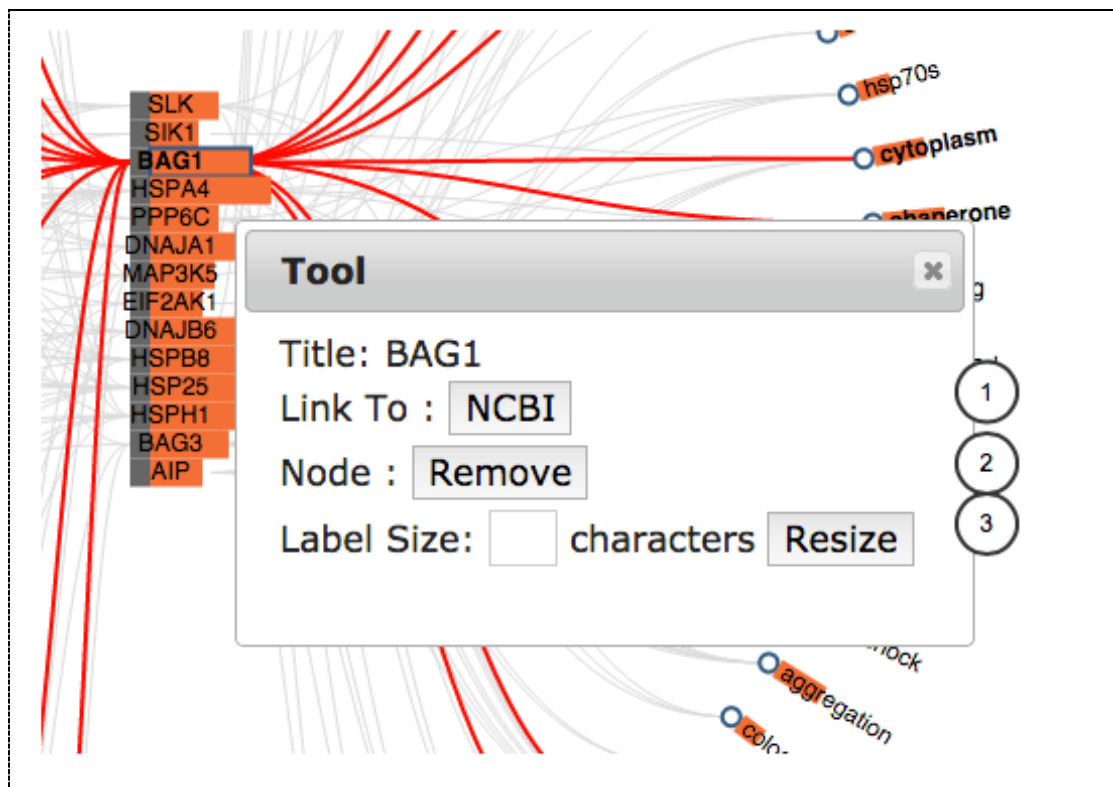
Figure 4.1 *Right-click functions. Gene 'BAG1' can be linked out to NCBI, removed from the graph, and the label name of it can be resized via right clicking the gene 'BAG1'.*

Figure 4.2 *eGIFT Sentence. All literature sentences in eGIFT databases when click the edge of gene 'BAG1' and iTerm 'chaperone' on the Concept Map graph*

5. GUI Function

- ***Switch***

  You can switch the position of the outer nodes and inner nodes by using the switch function in the GUI. Adjusting the position of the outer and inner nodes may create a more visually appealing graph.

- ***Color*** (e.g. expression data)

  WebGIVI can also color gene nodes based on gene's expression data. Please check 'Load Expression Files' function and test by using the samples
  (see sample data:http://raven.anr.udel.edu/webgivi/sampleData/symbolRegulation.txt). Note: you need to use the tab delimited data and **your header row should be the same as the sample file** (symbol \t regulation). The first column is the gene symbol and the second column is the expression regulation ('Up' or 'Down')
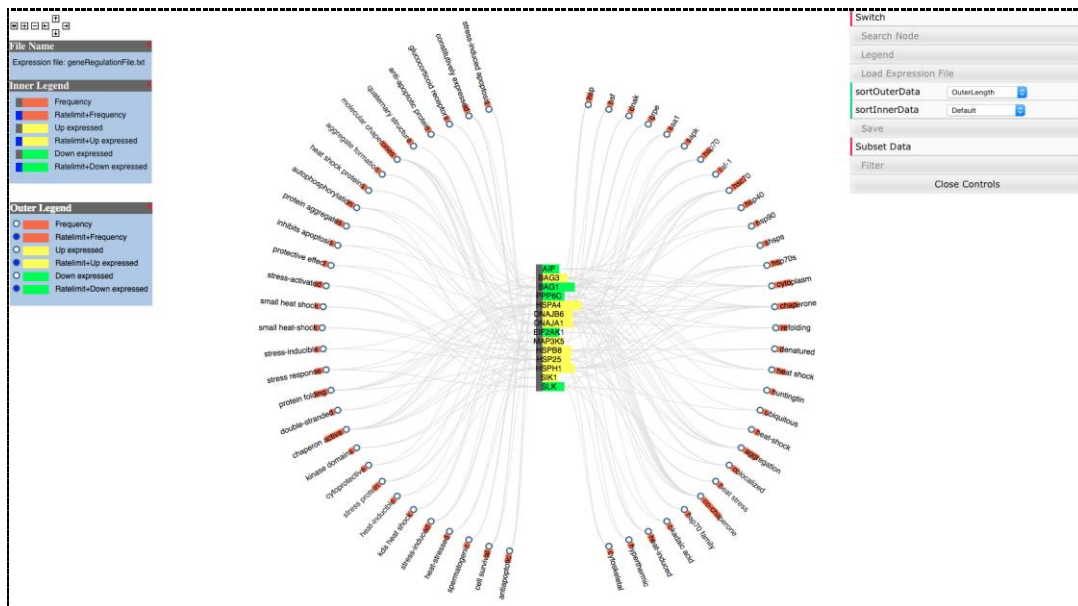
Figure 5.1 *Concept map with expression data. The yellow bar chart represents up-expressed gene and the green bar chart represents down-expressed genes.*

- *Subset*

  You can subgroup the selected nodes and edges into a simple graph. You can also go back to your original graph via clicking 'Previous Graph' button.

- *Sort*

Function	Description
OuterLength	Sort based on length of node names
OuterFrequency	Sort based on frequency of nodes
OuterRateLimited	Sort based on whether rate limited
InnerLength	Sort based on length of node names
InnerFrequency	Sort based on frequency of nodes
InnerRateLimited	Sort based on whether rate limited

- *Save*

  You can save all the data of the graph into two-column tab delimited text file, which can be resubmitted and visualized in WebGIVI in future.

  Graph can also be saved as SVG image file via 'Save as SVG' function, SVG file will be downloaded in your local computer. This SVG file can be saved as high-resolution image via 'SVG to PNG' function.

- *Filter*

  You can filter the graph based on the frequency of nodes. You can define your own frequency cutoff.

6. Custom data

You can also visualize any two-column tab delimited data. In Figure 6, genes and their involved pathways are visualized (see sample data: http://raven.anr.udel.edu/webgivi/sampleData/custome-sample.txt).

Figure 6. *Concept Map of customized data. The inner nodes are pathway names and the outer nodes are gene symbols.*