

webGIVI

Version 1.0

Tutorial

Liang Sun

E-mail: sunliang@udel.edu

Introduction

The challenge of high throughput transcriptome studies is to present 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 is to associate genes and informative terms (iTerm) that are obtained from the biomedical literature using eGIFT. However, examining large lists of *iTerm* and gene pairs is a challenge. We have developed webGIVI, a user-friendly web interface (<http://raven.anr.udel.edu/~sunliang/webgivi/index.php>), which was built via Cytoscape and Data Driven Document (d3.js) JavaScript library and can be used to visualize gene and *iTerm* pairs. webGIVI can accept EntrezGene list that will be used to retrieve a gene symbol and *iTerm* list. This list can be resubmitted to visualize the gene-*iTerm* pairs using Cytoscape, or Concept Map methods. A variety of visualization layout methods are available in webGIVI including force directed, tree and circle. In concept map graph, gene expression data can also be accepted to color the nodes in the graph. Visualized Graph on the website can also be saved as a tab-delimited file or PNG image.

The current version is “1.0”.

It is freely available at <http://raven.anr.udel.edu/~sunliang/webgivi/index.php> (see Figure 1).

All the code is hosted on Github, you can fork it through the link: [XXXX](#).

The screenshot shows the webGIVI homepage. At the top, there is a yellow header with the logo 'webGIVI^{tool}' and the text 'Visualization of Gene and iTerm'. Below the header is a blue navigation bar with links: Home, eGIFT, Tutorial, and Contact. The main content area has a white background. It starts with a paragraph explaining that webGIVI can accept EntrezGene lists to retrieve gene symbols and iTerm lists, which can then be visualized using Cytoscape or Concept Map. A note mentions a limitation of 1000 node-iTerm pairs per visualization. Below this, there are two sections: 'Data Visualization: eGIFT Gene Enrichment Analysis' and 'Customized Data'. The eGIFT section includes a form for entering Entrez Gene IDs or Entrez-iTerm pairs, with radio buttons for 'Entrez ID' and 'Symbol iTerm Pair'. The Customized Data section includes a form for uploading a two-column tab-delimited file. At the bottom right, there is a flowchart showing the workflow: 'EntrezGene ID' leads to 'eGIFT Database', which leads to 'Data Input Interface', which leads to 'Visualization Tool', which finally leads to 'BioLayout' and 'WebGIVI'.

Figure 1 Homepage of webGIVI

Brower requirements

All modern browsers, such as Safari, Google Chrome, and IE are supported. Currently, FireFox doesn't work perfectly.

Introduction of webGIVI Interface

1. Input Data

Currently, we only accept two input data formats, NCBI Entrez gene ID (see sample:http://raven.anr.udel.edu/~sunliang/webgivi_github/sampleData/entrezID-sample.txt) and gene symbol iTerm pairs (See Table 1). Column 1 is gene symbol, and column 2 is the gene-associated iTerm (see sample:http://raven.anr.udel.edu/~sunliang/webgivi_github/sampleData/gene-item-sample.txt), or any two-column tab-delimited format data can be visualized (see sample:http://raven.anr.udel.edu/~sunliang/webgivi_github/sampleData/custome-sample.txt). You can copy the sample data and paste into the submit box in webGIVI to play with it.

Table 1 Input data Format

Gene Symbol	eGIFT iTerm
PCSK7	Centrosome
RAD21	Centrosome
RCC2	Centrosome
ORC2L	Centrosome
PCNT	Centrosome
PLK1	Centrosome
TACC3	Centrosome
CDC14B	DNA repair
CHAF1B	DNA repair
ERCC3	DNA repair
CHAF1B	DNA replication
DBF4B	DNA replication
ORC2L	DNA replication
PIF1	DNA replication

2. Filter Function

eGIFT applied text mining method to extract all gene associated informative terms (iTerm). Some iTerms may be not so informative to you. You can edit this table and then visualized all genes and iTerms. Hovering iTerms in the table will display all associated gene symbols, which will help you to judge the importance of the iTerms and delete them.

<input checked="" type="checkbox"/> 2-fold	<input type="checkbox"/> aggregation	<input type="checkbox"/> antibodies	<input type="checkbox"/> atp
<input type="checkbox"/> binding	<input type="checkbox"/> catalytic	<input type="checkbox"/> catalytic activity	<input type="checkbox"/> cell polarity
<input type="checkbox"/> characterized	<input type="checkbox"/> chromosome	<input checked="" type="checkbox"/> consists	<input type="checkbox"/> cytosol
<input checked="" type="checkbox"/> defects	<input type="checkbox"/> degradation	<input type="checkbox"/> detected	<input type="checkbox"/> distinct
<input type="checkbox"/> embryogenesis	<input type="checkbox"/> embryonic	<input type="checkbox"/> energy metabolism	<input type="checkbox"/> exons
<input type="checkbox"/> extracellular	<input checked="" type="checkbox"/> fetal	<input checked="" type="checkbox"/> formed	<input checked="" type="checkbox"/> gene
<input checked="" type="checkbox"/> generated	<input type="checkbox"/> glycolytic	<input type="checkbox"/> identity	<input type="checkbox"/> impaired
<input type="checkbox"/> implicated	<input type="checkbox"/> in_situ hybridization	<input type="checkbox"/> lrp5	<input type="checkbox"/> maintained
<input type="checkbox"/> mesenchyme	<input type="checkbox"/> metabolism	<input type="checkbox"/> mice	<input type="checkbox"/> mitochondria
<input type="checkbox"/> mitochondrial	<input type="checkbox"/> mma	<input type="checkbox"/> mma expression	<input type="checkbox"/> muscle
<input type="checkbox"/> n-terminal	<input checked="" type="checkbox"/> novel	<input type="checkbox"/> ocular	<input type="checkbox"/> origin
<input type="checkbox"/> ortholog	<input type="checkbox"/> oxidative phosphorylation	<input type="checkbox"/> performed	<input type="checkbox"/> phenotype
<input type="checkbox"/> promoter	<input type="checkbox"/> rat liver	<input type="checkbox"/> severing	<input type="checkbox"/> skeletal muscles
<input type="checkbox"/> synaptic	<input type="checkbox"/> syndrome	<input type="checkbox"/> synthesis	<input type="checkbox"/> up-regulated
<input type="checkbox"/> vascular	<input type="checkbox"/> vertebrate	<input type="checkbox"/> wnt	<input type="checkbox"/> wnt pathways
<input type="checkbox"/> wnt receptors	<input type="checkbox"/> wnt signaling	<input type="checkbox"/> wnts	

Figure 2 Filter function of webGVI1

3. Visualization

You can visualize your gene and iTerms by either concept map or Cytoscape methods. For large input data set, using concept map method is highly recommended.

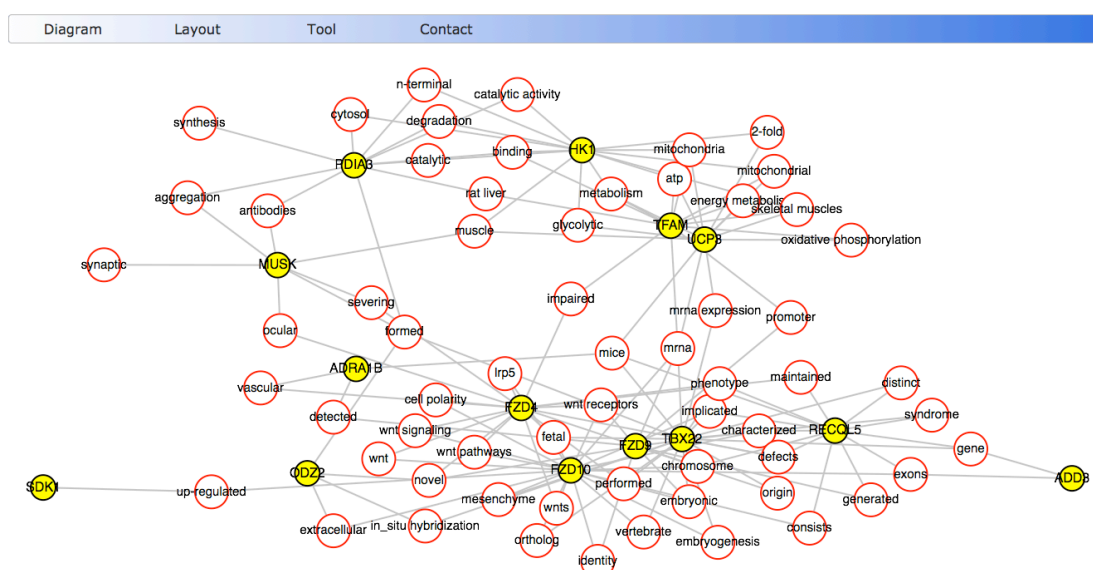


Figure 3.1 Cytoscape Graph

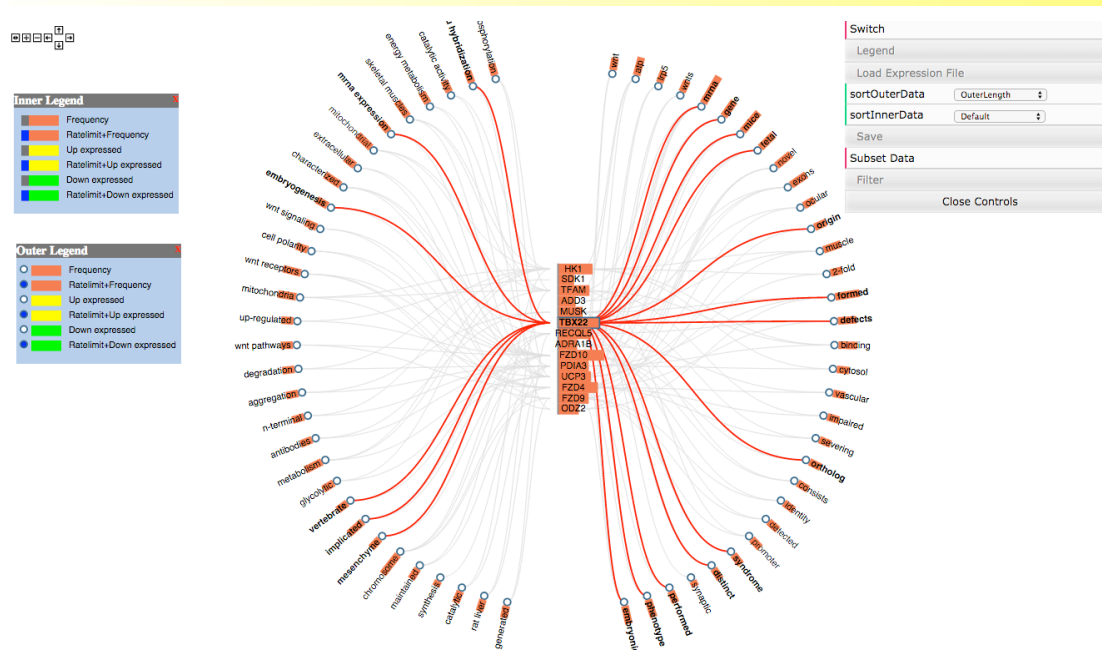


Figure 3.2 Concept Map Graph. Gene symbols are located in the center of the circle. eGIFT iTerms are located on the circle. Bar charts on the nodes indicate the frequency of connected edges. Genes will be automatically tested and colored if they are rate limiting.

4. Click in Graph

Left or right clicking nodes or edges, you can delete, or link out to other databases, or resize the node labels.

4.1 Remove nodes

User can deleted the iTerms or genes that they think is not informative.

4.2 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.

4.3 Resize node label

If label names of some nodes are too long and therefore overlapping with other nodes, users can adjust the length of node labels in the graph via setting up how many characters you want to display.

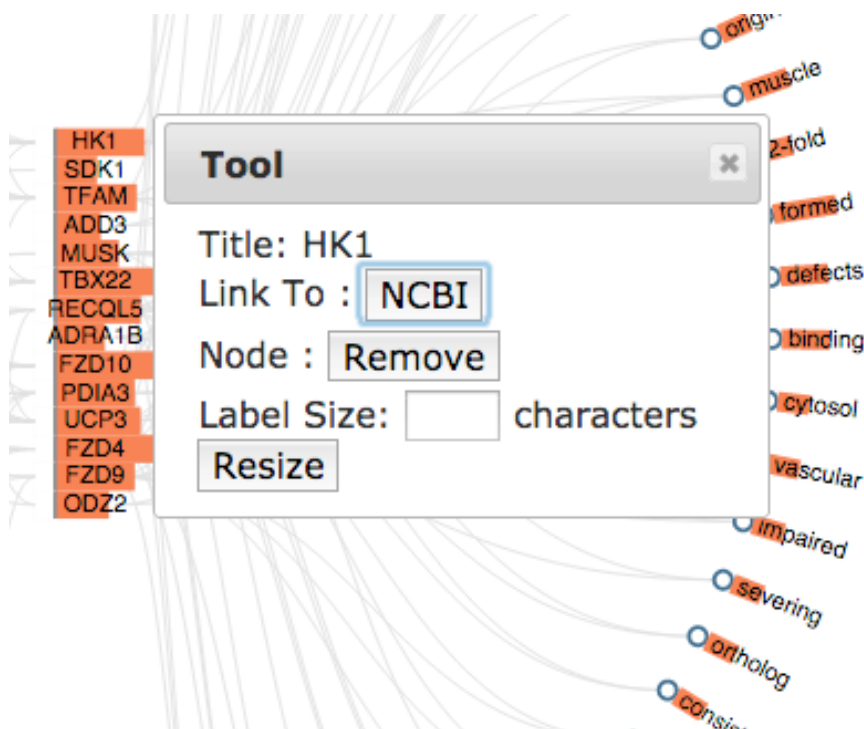


Figure 4.1 Right-click functions. Link out to NCBI, remove the node you are not interested, and resize the label if the name is too long.

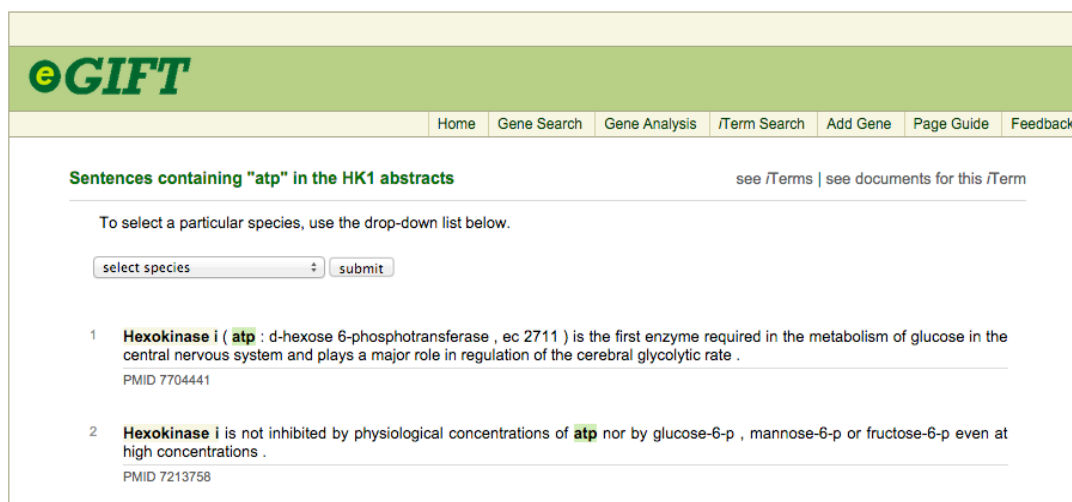


Figure 4.2 eGIFT Sentence. All literature sentences in eGIFT databases when users click the edge of gene 'atp' and iTerm 'Hexokinase i'.

5. GUI Function

5.1 Switch inner nodes and outer nodes

For outer and inner nodes, you can switch the position of outer nodes and inner nodes by using the switch function in the GUI, and make the graph look better.

5.2 Color Node (e.g. expression data)

webGIVI can also color nodes based on nodes' attributes such as expression data. Please check 'Load Expression Files' function and try to use the samples to test (see sample data:

http://raven.anr.udel.edu/~sunliang/webgivi_github/sampleData/symbolRegulation.txt). Note: you need to use tab-delimited file and your first line should be the same as the sample file (symbol \t regulation). The first column is the gene symbol and second column is the expression regulation ('Up' or 'Down', it should be case sensitive!).

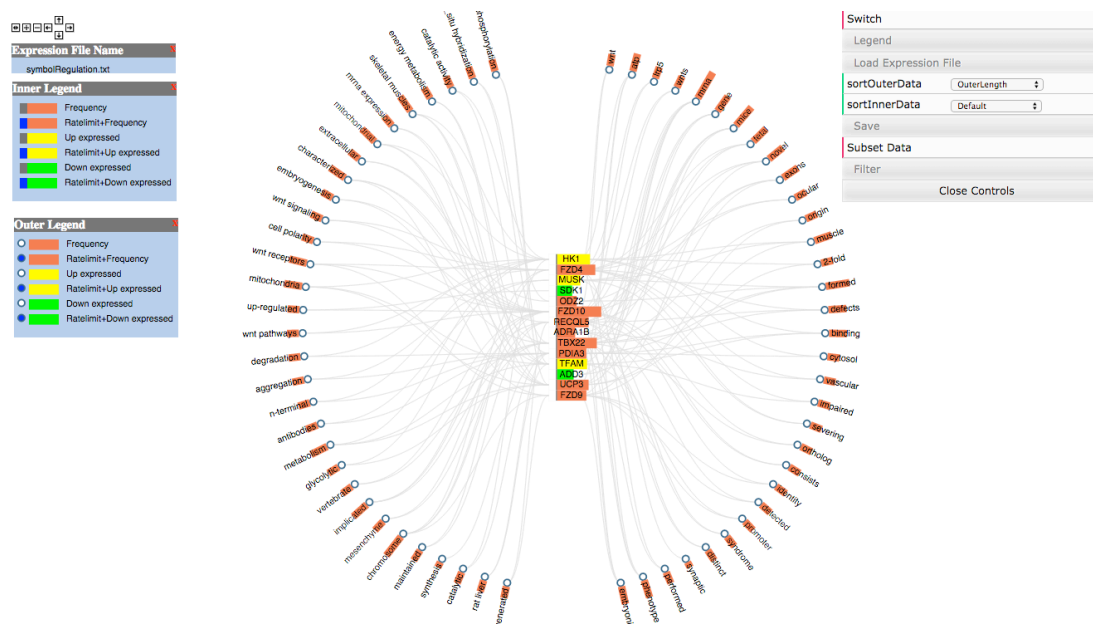


Figure 5.1 Concept map with expression data. Yellow bar charts represent up-expressed genes and green bar charts represent down-expressed genes.

5.3 Subset the graph

If your graph is too complex and you need the gene and iTerm graph of only several interested genes or iTerms, you can subset the selected nodes and edges into a simple graph. You can also go back to your original graph via clicking 'Previous Graph' button.

5.4 Sort function

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

5.5 Save

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

5.6 Filter

You can filter the graph based on the number of edges connected. Usually, more frequent of iTerm and genes in the graph, more interesting they are. You can define your own cutoff.

6. Customized data

You can also visualize your personal data if it's two-column tab-delimited format. In Figure 6, genes and its involved pathways are displayed.

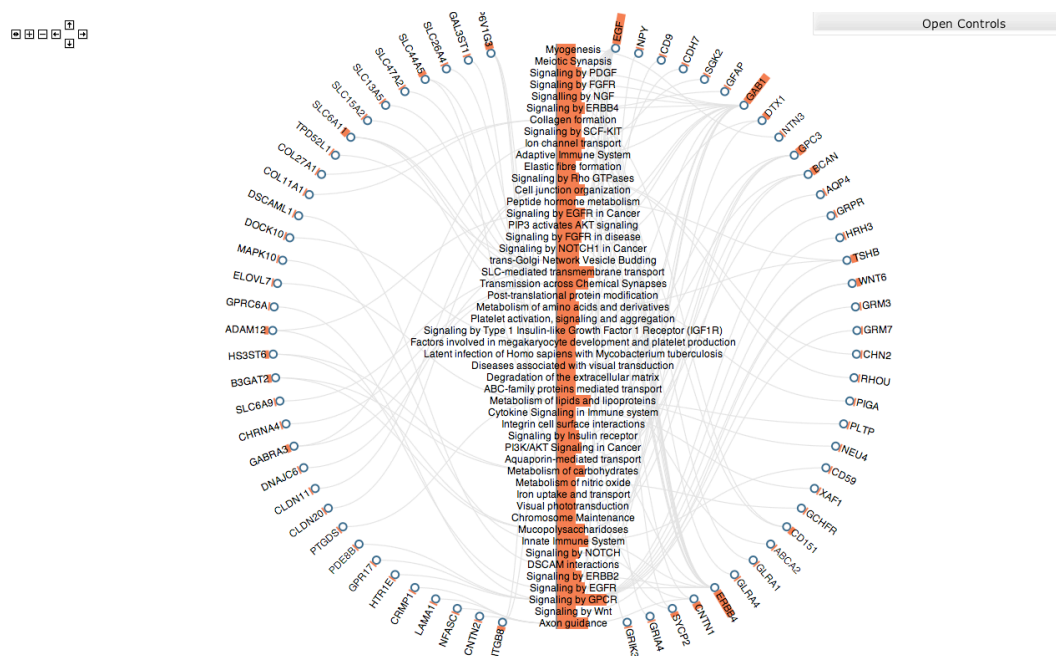


Figure 6 Concept Map of customized data. The inner nodes are pathway names and the outer nodes are gene symbols.