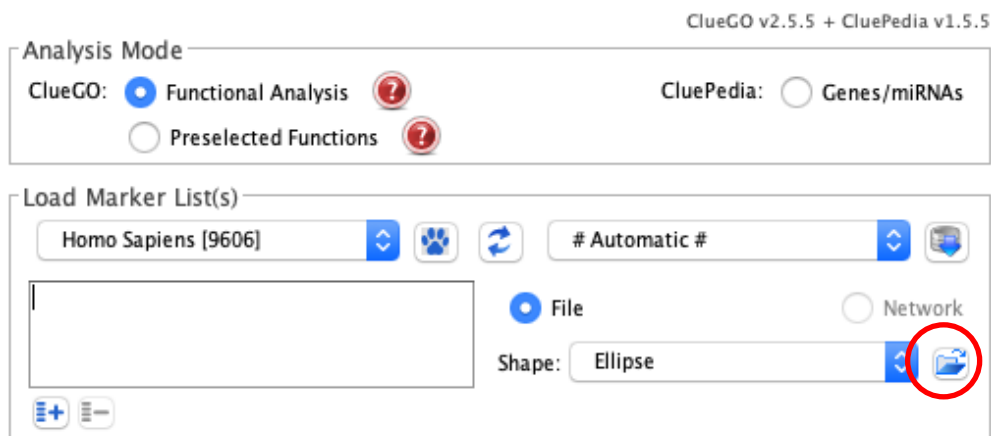


# Pathway Analysis with Cytoscape, ClueGO, and CluePedia

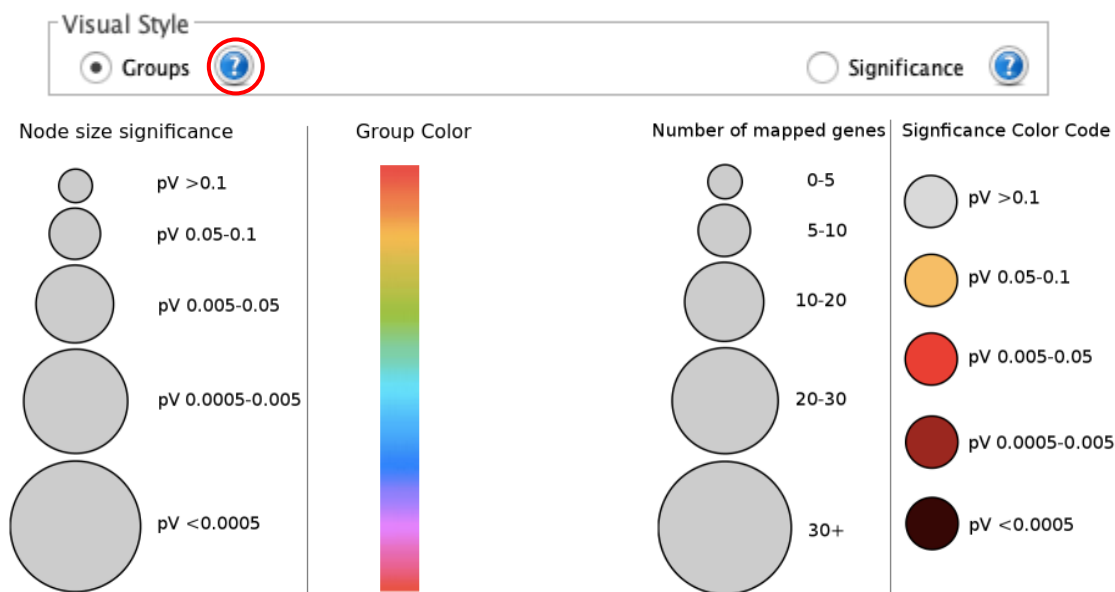
- Open Cytoscape
- Go to Apps -> App Manager and load ClueGO
- Search for and install the CluePedia plug-in: Apps -> App Manager
- Load ClueGO: Apps -> App Manager -> ClueGO
- Keep default analysis mode (Functional Analysis) and species (*Homo sapiens*)



- Open up a terminal on your local machine (without logging into logrus) and scp our list of differentially expressed genes:

```
scp -P 44111 <username>@gateway.training.ncgr.org:/home/elavelle/tutorial_genes.txt ~/Desktop/
```

- Open the document with a text editor (such as Word or Notepad). Copy the document contents and paste them into the empty box in Cytoscape. With default option “# Automatic #”, ClueGO will detect the style of gene ID.
- Selection in the "Visual Style" box decides the factors that will determine node properties. Keep the default selection, "Groups".



- Pick ontologies from among downloaded annotations. Node shape can be changed to distinguish the GO term sources.
- Clicking on the “Evidence” button will display a decision tree clarifying evidence codes. Keep “All” checked for now.

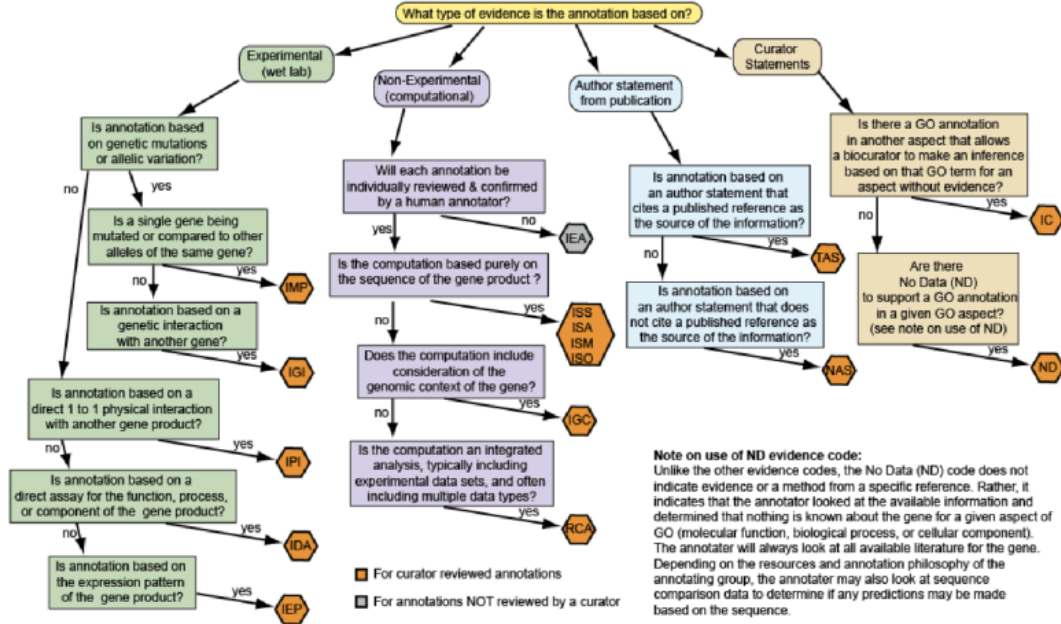
ClueGO Settings

Type	Name	#	Date	Shape	
<input checked="" type="checkbox"/>	GO	BiologicalProcess-EBI-UniProt-GOA	18361 (18090)	27.02.2019	Ellipse
<input checked="" type="checkbox"/>	GO	CellularComponent-EBI-UniProt-GOA	2026 (19089)	27.02.2019	Hexagon
<input checked="" type="checkbox"/>	GO	ImmuneSystemProcess-EBI-UniProt-GOA	1221 (3886)	27.02.2019	Rectangle
<input checked="" type="checkbox"/>	GO	MolecularFunction-EBI-UniProt-GOA	5287 (18043)	27.02.2019	Octagon
<input checked="" type="checkbox"/>	INTERPRO	ProteinDomains	5525 (12039)	27.02.2019	
<input checked="" type="checkbox"/>	KEGG	KEGG	321 (7512)	27.02.2019	v

Evidence

Code
<input checked="" type="checkbox"/> All
<input type="checkbox"/> All_Experimental_(EXP_IDA,IPI,IMP,IGI,IEP)
<input type="checkbox"/> All_without_IEA
<input type="checkbox"/> EXP (Inferred from Experiment)
<input type="checkbox"/> IBA (Inferred from Biological Aspect of Ancestor)
<input type="checkbox"/> IBD (Inferred from Biological Aspect of Descendant)

### GO Evidence Code Decision Tree



- Check the “Show only Pathways with  $pV \leq$ ” box and change the value to 0.05. Skip over “Advanced Term/Pathway Selection Options” for now and expand the “Statistical Options” menu. In the first box, change the selection from the default (“Enrichment/Depletion”) to “Enrichment”.
- A left side to the hypergeometric test will include significantly depleted pathways, a.k.a. GO terms with significantly few genes found, based on the number of genes in the pathway, the reference list and the input list. If you want the details on how p-values are calculated for GO terms, click the “?” icon.

Statistical Options

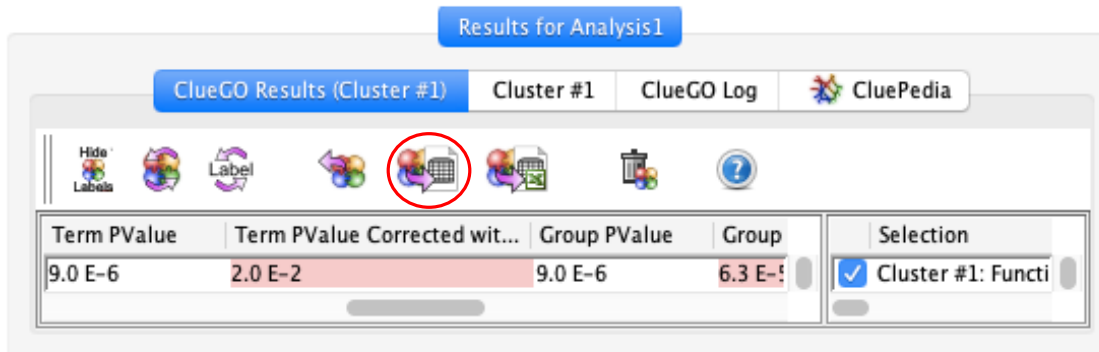
Advanced Statistical Options

Enrichment (Right-sided hypergeometric test)

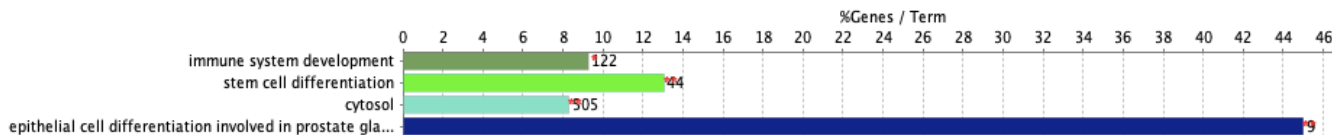
Bonferroni step down pV Correction

mid-P-values  Doubling ?

- Run the analysis! It will take several minutes. When it's finished, look to the bottom right of the screen. Select the "ClueGO Results (Cluster #1)" tab, and click the indicated icon to save ClueGO results tables to your machine.



- Open the new directory. Among the several files, we can find a .png with GO terms represented with a horizontal bar graph as they are shown in the "Cluster #1" tab of in the software. Adjacent bars of the same color are terms that belong to the same group.



- The most generally useful file is "NodeAttributeTables.txt". The fields here indicate annotation source, p-values, and gene composition and for each gene. Note the "GOLevels" column and the "Overview Term" column on the far right. Let's return to Cytoscape to clarify what these mean.
- Expand the options box we passed over before. Each GO term has one or more associated levels. Only those with a level within the values in the "GO Tree Interval" have the potential to appear in the results. The boundaries of this window can be changed directly, or by dragging the "Network Specificity" bar.

**Network Specificity** ?

---

Global Medium Detailed

Use GO Term Fusion ?

Show only Pathways with pV ≤  ↕

Advanced Term/Pathway Selection Options

GO Tree Interval

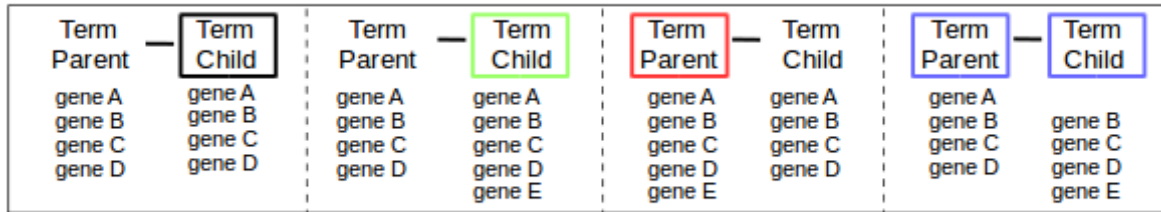
↕ Min Level  ↕ Max Level ?

GO Term/Pathway Selection (#/% Genes)

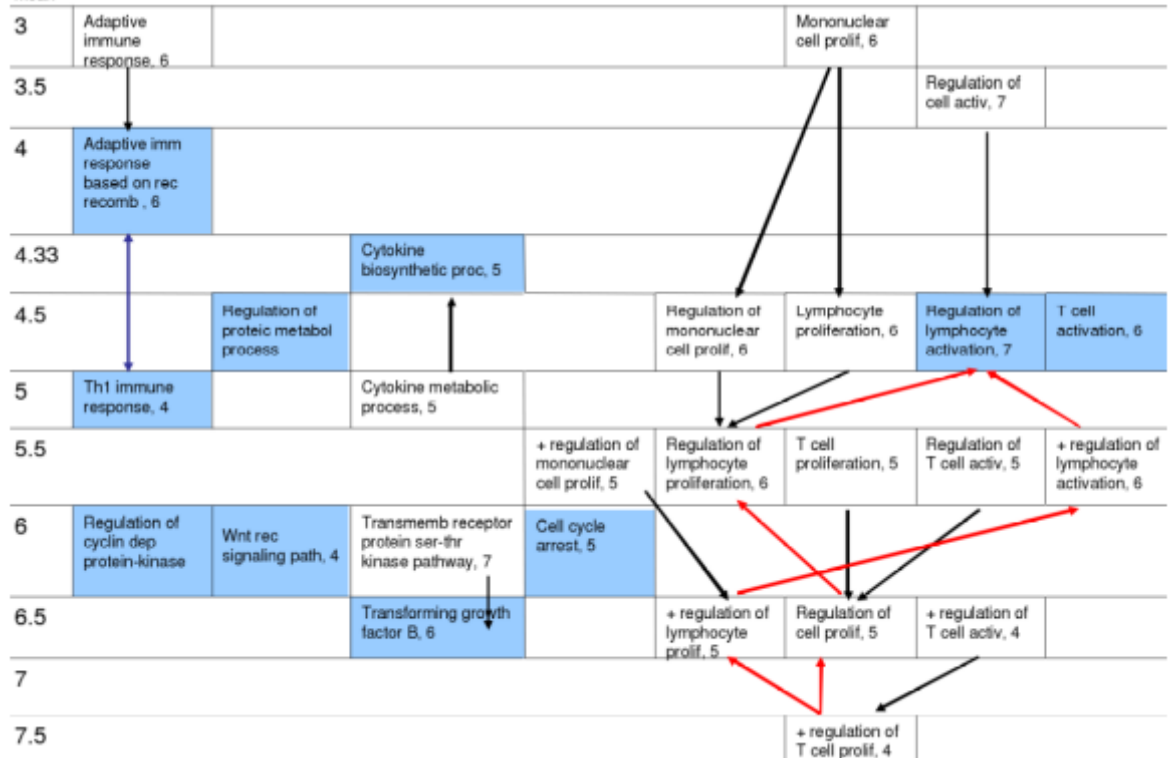
Cluster #1

↕ Min #Genes  ↕ %Genes

- It's typically a good idea to check the "Use GO Term Fusion" box. This will improve runtime and simplify the results by eliminating redundancy.



GO level mean *26 terms found (ClueGO standard settings, 30 test genes) are fused to 10 (shown in blue).*



- Below the “GO Tree Interval” box, there are two others. The “GO Term/Pathway Selection” box will set the criterion for what is a qualifying GO term.
- The “GO Term/Pathway Network Connectivity” scroll tab will set the kappa score threshold for what the edges that appear between nodes. Kappa score is a measure of relation any and all GO terms based on the number of associated genes they have in common. For more details regarding this calculation, click on the “?” icon.

GO Term/Pathway Selection (#/% Genes)

Cluster #1

3 Min #Genes 4.000 %Genes

GO Term/Pathway Network Connectivity (Kappa Score)

Low Medium High Score: 0.4

- Expand the “Grouping Options” box. Recall the variable (True/False) of the last column (Overview Term) in the node attribute table file we looked at. Those with a “True” value (as well as a check mark next to the GO id in the “Clue GO Results” tab in Cytoscape) are the leading term in the group, only meaning it is the group with the lowest p-value. The criterion to determine the leading term can be changed via the “Leading Group Term based on” selection. The threshold of commonality for what constitutes a group can also be adjusted, if so chosen.

Grouping Options

GO Term Grouping (Functional Grouping)

Use GO Term Grouping

Group coloring

Fix  Random

Leading Group Term based on


Highest Significance

Tree  Kappa Score

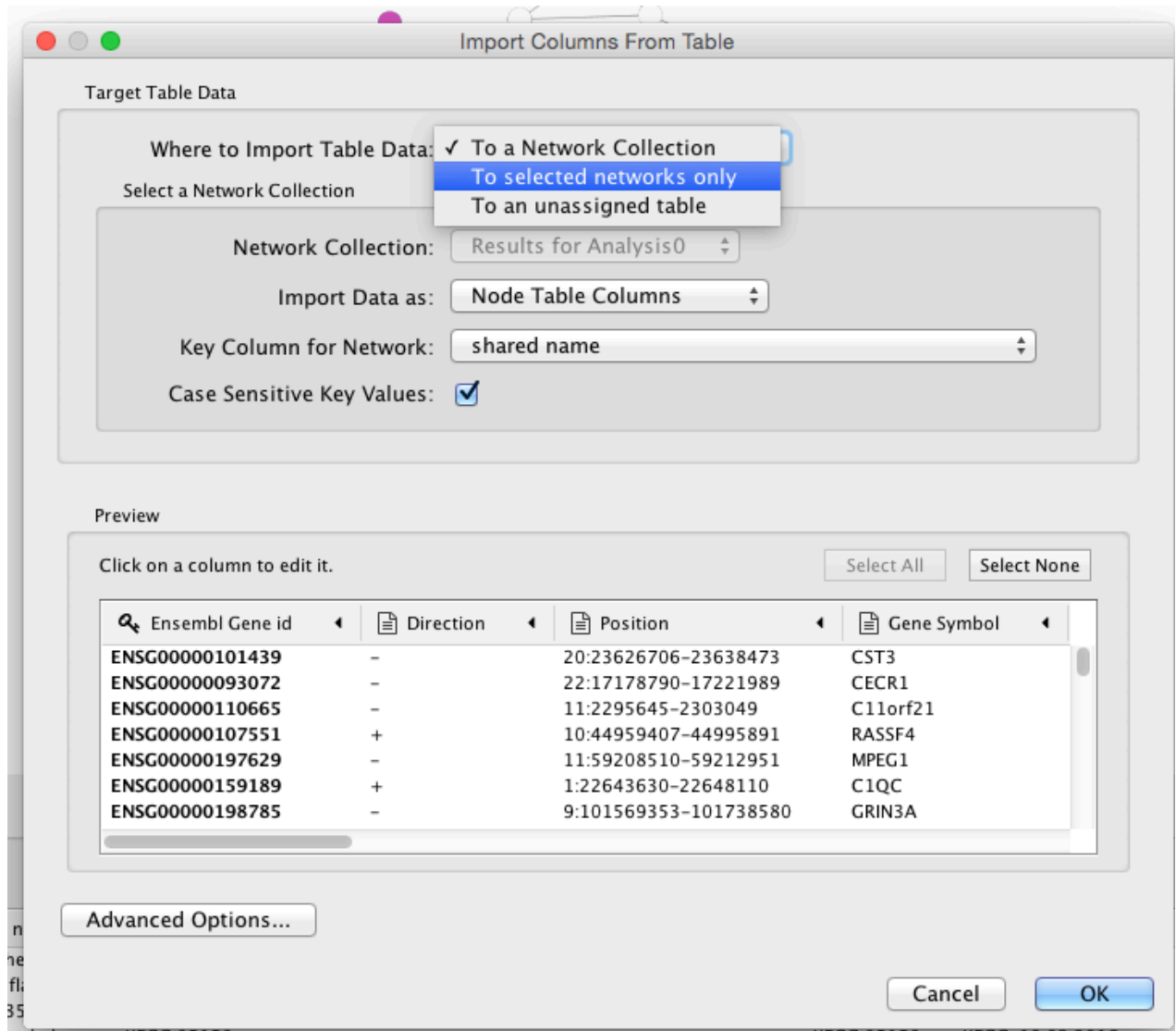
1 Initial Group Size

50 % Genes for Group Merge

50 % Terms for Group Merge

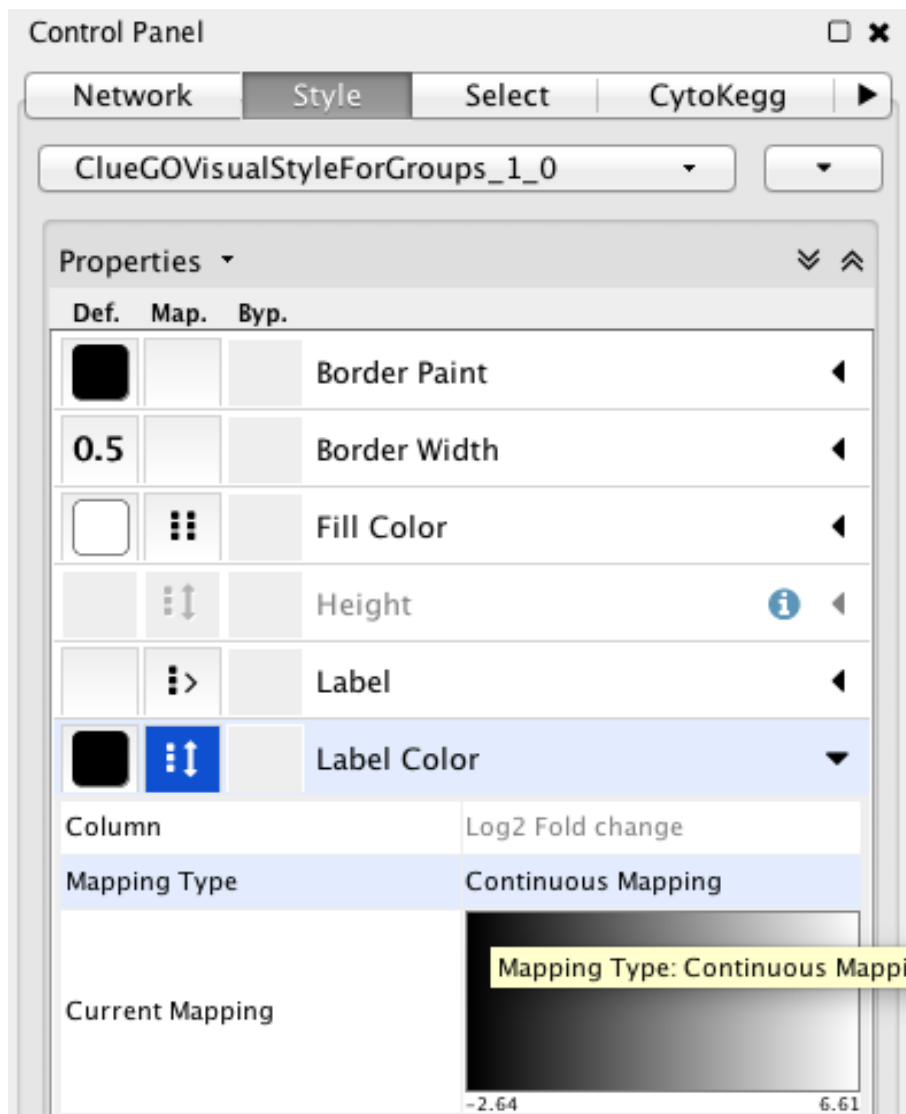
- Next, select the “CluePedia” tab and click this icon: 
- This will add miniature gene nodes within the edges between nodes. As of now, however, it is not apparent which of them are upregulated and which are downregulated.

- Navigate to File → Import → Table from File, and select the document from which the gene list was originally input.
- In the new window, chose “To selected networks only”. Under the “Key Column for Network” menu, pick “Input Gene ID”, then “OK”.

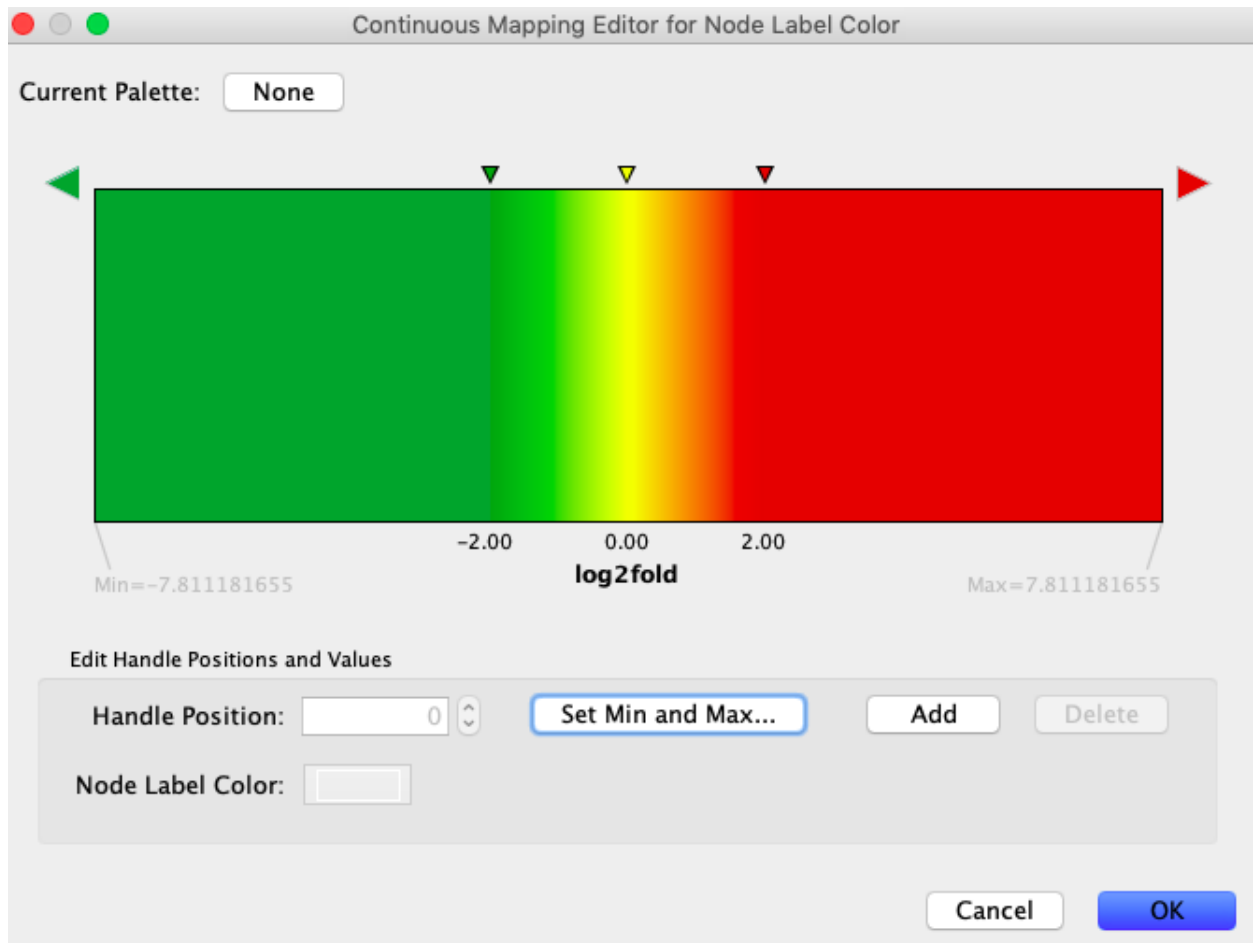


- In the control panel, on the left side of Cytoscape, select “Style”. For column, choose “Log2 Fold change” and for mapping type, select “Continuous Mapping” and click the spectrum map.





- In the window that appears, drag the handle position to 0 and change the color to grey or white. Then, drag the margins in to about a log2FoldChange of 2 and change their color. Change the arrows on the extremes to the same color.



- You can save the session by clicking the icon at the top right of the app, which will allow you to return to the visualization and settings of the analysis.