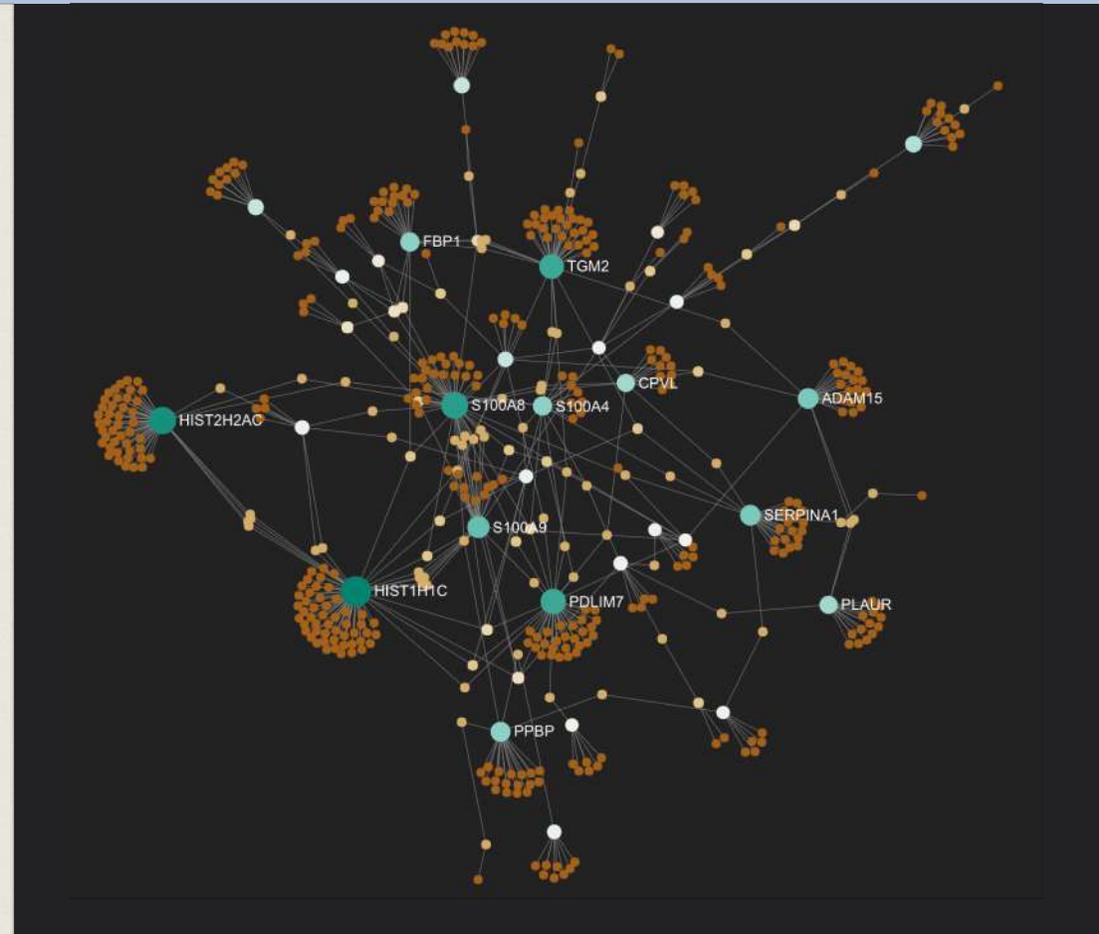
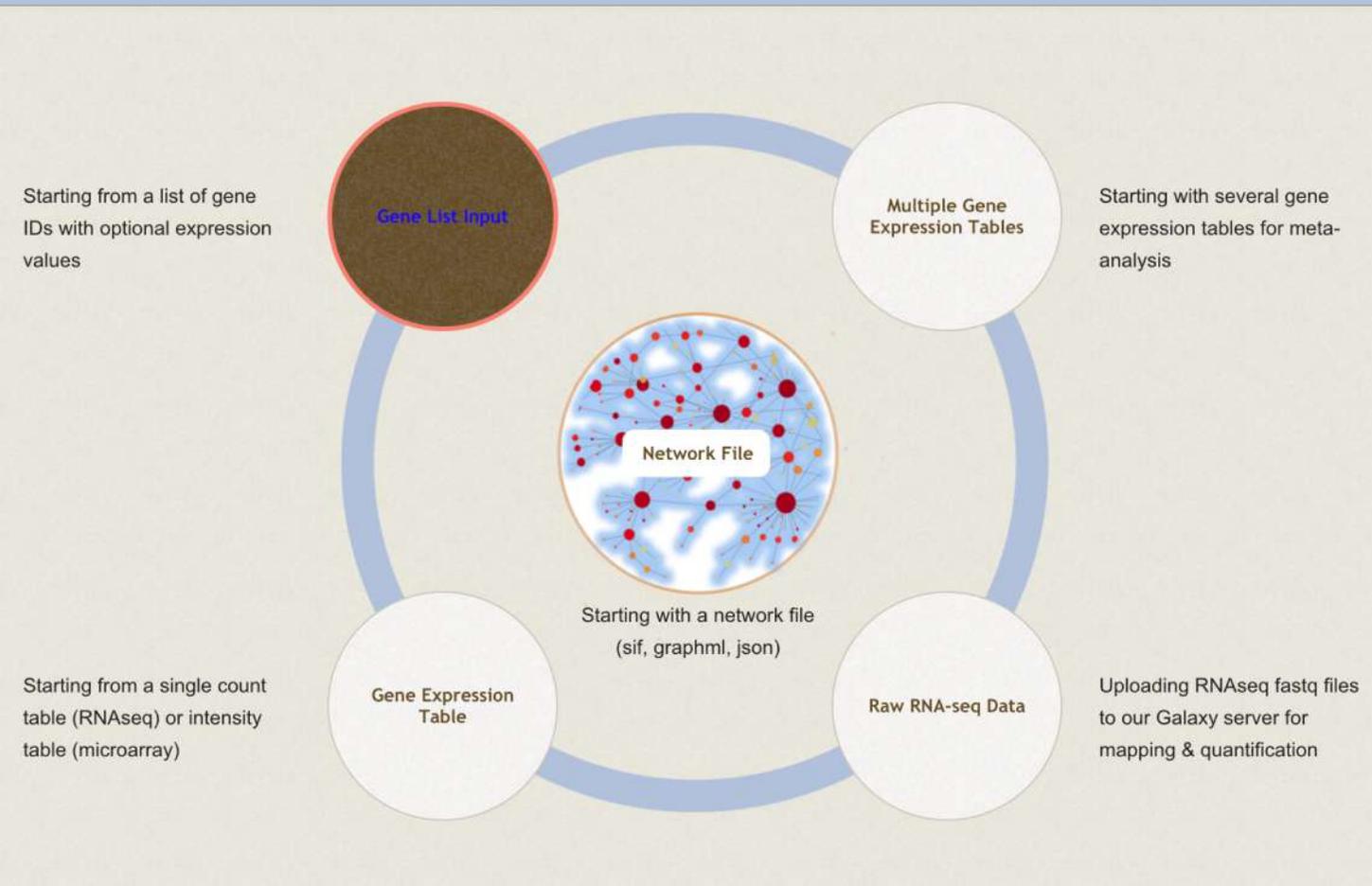


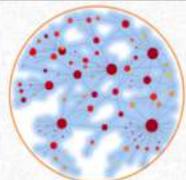
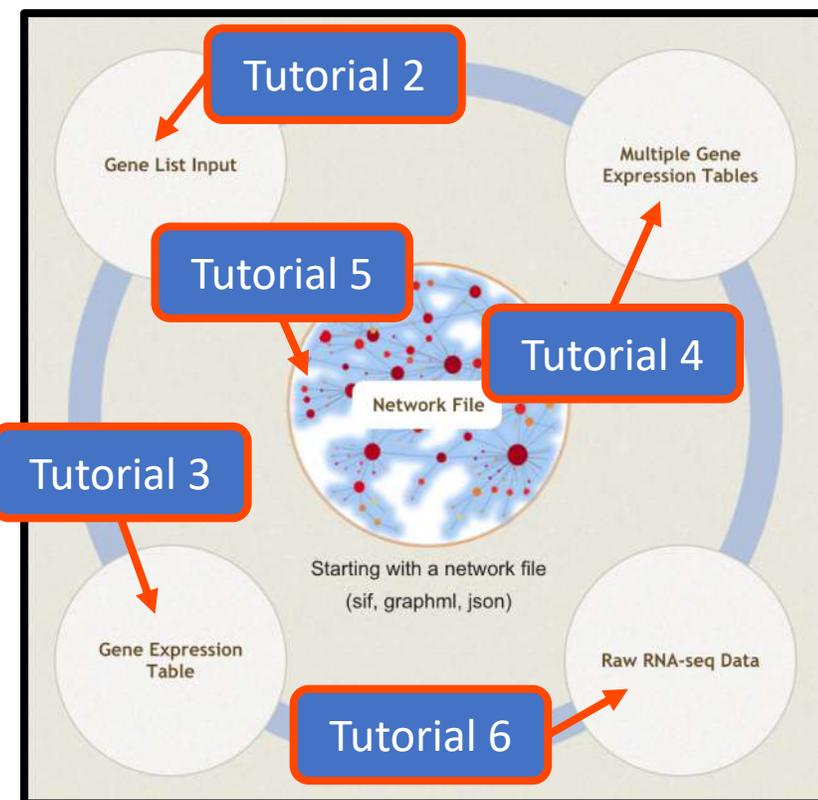
Tutorial 2a: single gene list input



Intro to NetworkAnalyst

- Web application that enables complex meta-analysis and visualization
- Designed to be accessible to biologists rather than specialized bioinformaticians
- Integrates advanced statistical methods and innovative data visualization to support:
 - Efficient data comparisons
 - Biological interpretation
 - Hypothesis generation

Tutorial 1: Overview



NetworkAnalyst -- a web-based platform for gene expression profiling & biological network analysis

Computer and browser requirements

- A modern web browser with Java Script enabled
 - Supported browsers include Chrome, Safari, Firefox, and Internet Explorer 9+
- For best performance and visualization, use:
 - Latest version of Google Chrome
 - A computer with at least 4GB of physical RAM
 - A 15-inch screen or bigger (larger is better)
- Browser must be WebGL enabled for 3D network visualization
- 50MB limit for data upload
 - ~300 samples for gene expression data with 20 000 genes

Goals for this tutorial

- A challenge for 'omics analysis is to find ways to benefit from big data without being overwhelmed by it
- Visualizing key functions and relationships within large datasets is an integral part of this
- Networks capture the topology of statistical and functional relationships in graphical form – thus they are a powerful tool for 'omics visual analytics
- The goal of this tutorial is to:
 - introduce NetworkAnalyst's tools for visualizing 2D and 3D networks from a list of uploaded gene IDs
 - Project and visualize these genes within their biological networks to explore their relationships and collective functions

Select example data

Select "Gene list 1"

2

Name	Parameter	Description
<input checked="" type="radio"/> Gene list 1	Organism: Human ID Type: Entrez Gene ID	Endotoxin tolerance gene signature (99 genes)
<input type="radio"/> Gene list 2	Organism: Mouse ID Type: Entrez Gene ID	Trem-1 activation gene signature (96 genes)
<input type="radio"/> Gene list 3	Organism: Fruit fly ID Type: Uniprot ID	Example gene list for testing purpose (41 genes)
<input type="radio"/> Gene list 4	Organism: C. elegans ID Type: Uniprot ID	Example gene list for testing purpose (74 genes)
<input type="radio"/> Gene list 5	Organism: S. cerevisiae ID Type: Uniprot ID	Example gene list for testing purpose (47 genes)
<input type="radio"/> Multiple Lists	Organism: Mouse ID Type: Symbol	Three lists of genes for testing purposes

NetworkAnalyst supports the upload of multiple lists.

NetworkAnalyst currently supports 17 organisms and 5 ID types. For this tutorial, select example gene list 1.

Select "Try Example"

1

Upload example data

NetworkAnalyst -- a web-based platform for gene expression profiling & biological network analysis

Please upload a list of genes/proteins

Upload a list of genes/proteins for network-based analysis and visualization, with built-in support for various functional enrichment analysis. For first time users, you are advised to explore the features using one of our example data sets.

Specify organism: H. sapiens (human)

Set ID type: Entrez ID

```
#Entrez logFC
4495 61.12
4496 51.06
4499 23.79
6354 21.04
6369 19.76
4494 16.24
4501 14.76
11026 14.04
199675 12.65
4316 12.04
771 8.19
6346 7.07
6367 6.97
5473 6.76
2357 5.71
5265 5.65
1462 5.27
2358 4.92
22918 4.58
```

1 Click "Upload"

2 Click "Proceed"

Previous Proceed

Xia Lab @ McGill (last updated 2018-12-08)

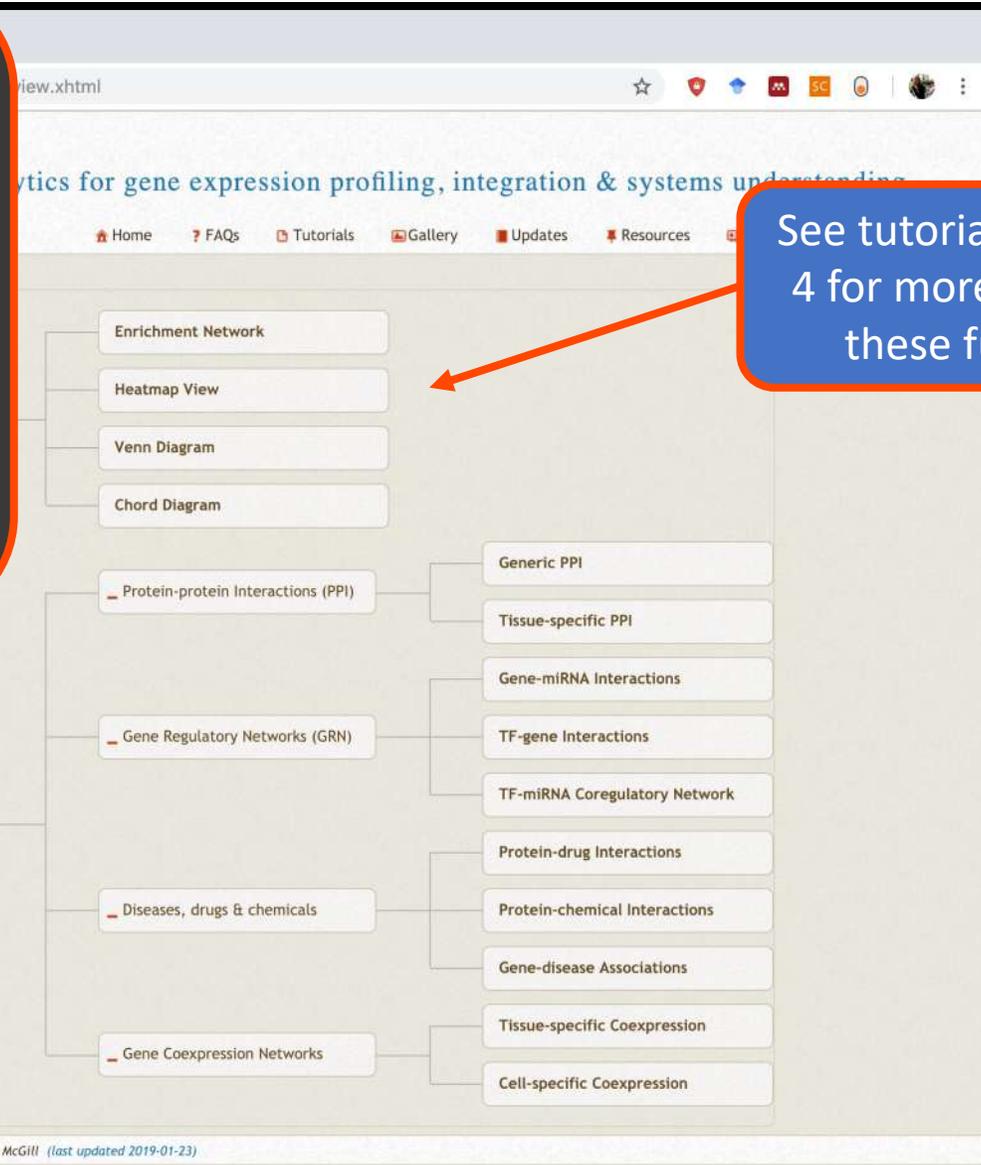
```
#Entrez logFC
4495 61.12
4496 51.06
4499 23.79
6354 21.04
6369 19.76
4494 16.24
4501 14.76
11026 14.04
199675 12.65
4316 12.04
771 8.19
6346 7.07
6367 6.97
5473 6.76
2357 5.71
5265 5.65
1462 5.27
2358 4.92
22918 4.58
```

Copy-paste a list of gene IDs from your file (optional: include fold changes to indicate the direction of change)

Analysis overview

NetworkAnalyst allows users to embed genes of interest within biological networks from 15 different databases.

These networks are mined to extract the genes/proteins, miRNA, drugs, chemicals, or diseases that have the strongest relationships to genes in the uploaded list (seed genes). The “Enrichment Network” function visualizes the results of enrichment analysis in network form. The “Venn Diagram” and “Chord Diagram” functions are for visualizing the overlap of more than one list.

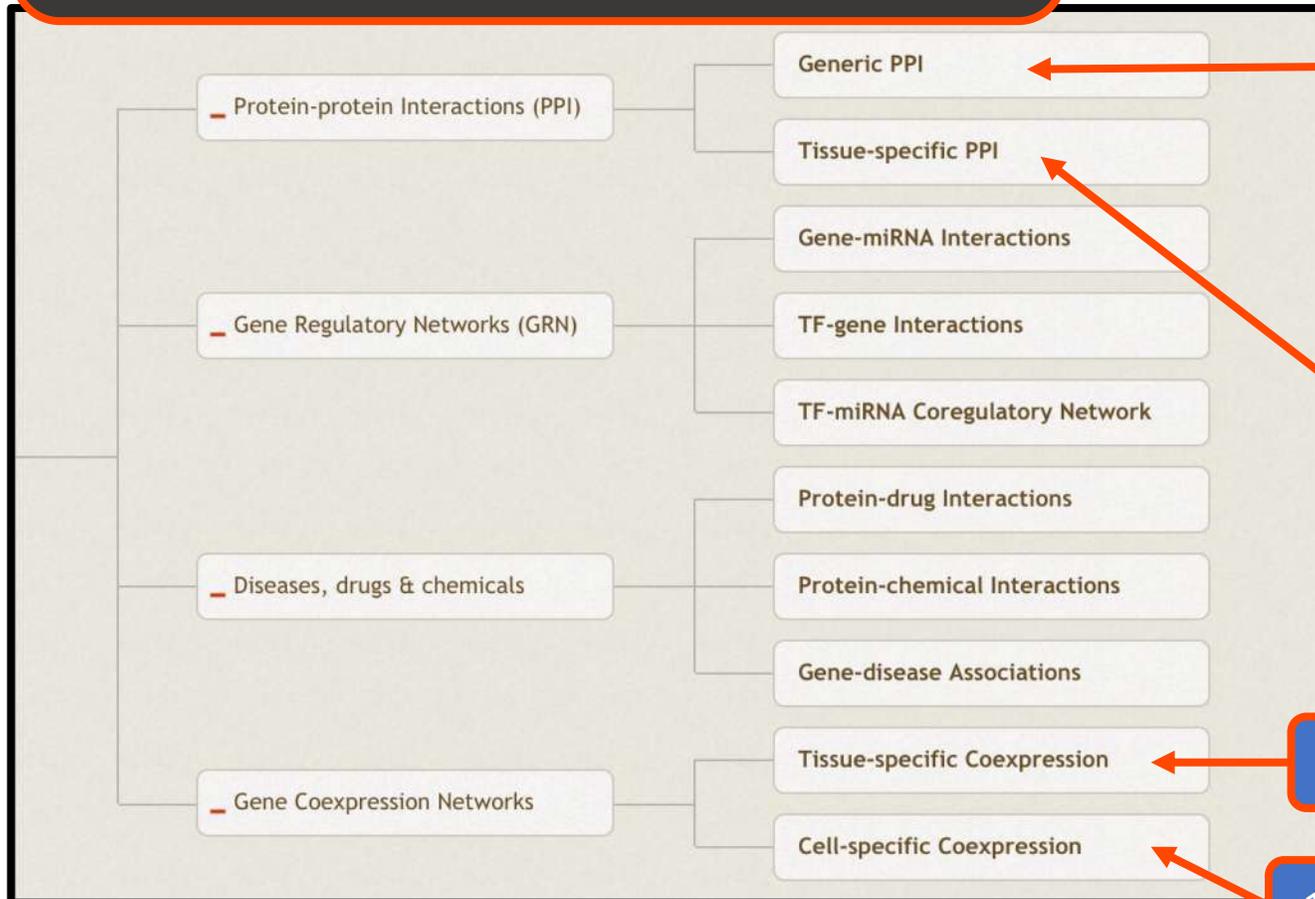


See tutorials 2b, 3 and 4 for more details on these functions

Network types available

The rest of this tutorial will focus on tissue-specific protein-protein interaction (PPI) networks, however the same visualization functions apply to multiple network types.

Confidence score based on multiple sources (experimental, computational, literature, orthology etc.)



Select a protein-protein interaction

	Name	Information	
<input type="radio"/>	IMEx Interactome	Literature-curated comprehensive Interactome from InnateDB (Breuer et al.)	
<input checked="" type="radio"/>	STRING Interactome	STRING interactome with medium (400) - high (1000) confidence score (Szklarczyk D et al.)	Confidence score cutoff: <input type="text" value="900"/> Require experimental evidence: <input checked="" type="checkbox"/>
<input type="radio"/>	Rolland Interactome	Experimentally validated binary human PPI data (Rolland T et al.)	None

42 tissues

Select a tissue type for PPI network:

--- Not Specified ---

Information	Filter
The tissue specific protein-protein interaction data were collected from the DifferentialNet database. It shows the differential protein-protein interactions across human tissues.	Filter: <input type="text" value="30"/> ?

46 tissues

Filter network based on PPI scores:
30 - PPIs unchanged across tissues
1 - PPIs unique to selected tissue

19 cell types

Select tissue-specific PPI network

PPI network databases contain relationships from many different sources, and so they are broader than some of the other types (i.e. TF-gene networks). Unlike chemical or disease nodes, many of the added PPI nodes will be connected to each other, greatly increasing the number of edges in the resulting network.

Select tissue of interest
(whole blood is closest
to PBMC)

1

Select a tissue type for PPI network:

Whole blood
Prostate
Skin (suprapubic)
Skin (lower leg)
Stomach
Testis
Thyroid
Uterus
Vagina
Whole blood

Filter

Filter: 15.0

OK

2

Click "Ok"

Network creation

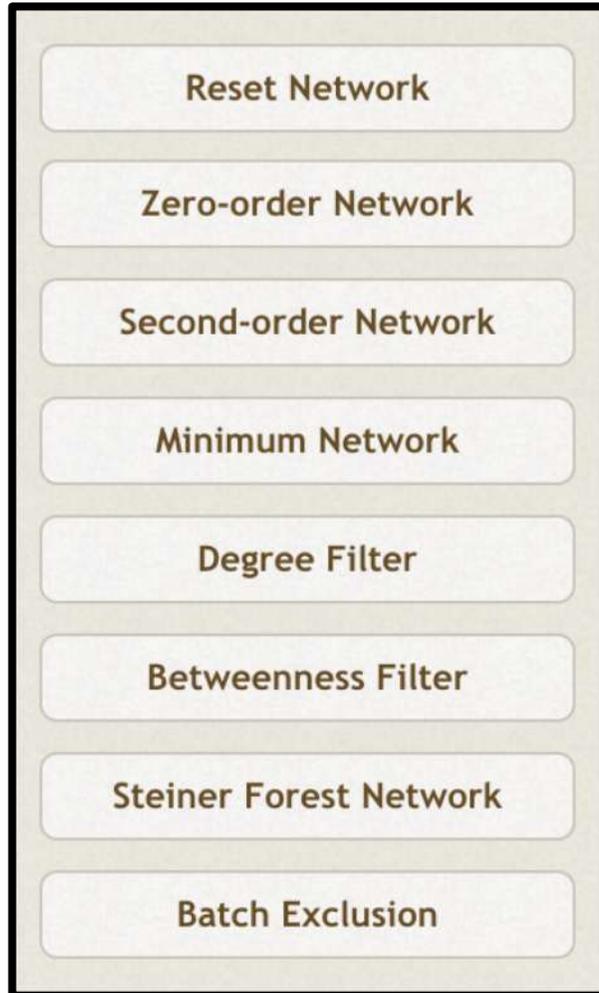
After selecting the network type, you will see a panel of **tools** for network construction, and a **mapping overview** summarizing the network statistics.

The screenshot displays the NetworkAnalyst web interface. At the top, the title reads "NetworkAnalyst -- a comprehensive platform for gene expression profiling, meta-analysis & systems understanding". A navigation bar includes links for Home, FAQs, Tutorials, Updates, Resources, Contact, and My Account. On the left, a sidebar menu lists options: Upload Data, Analysis Overview, Network Analysis (selected), Network Creation, Network Visualization, Download, and Exit. The main content area is titled "Mapping Overview" and features a table with columns: Networks, Nodes, Edges, Seeds, and Interactions (.SIF). The table lists subnetworks 1 through 9 with their respective statistics. To the right, a "Network Tools" panel contains buttons for: Reset Network, Zero-order Network, Second-order Network, Minimum Network, Degree Filter, Betweenness Filter, Steiner Forest Network, and Batch Exclusion. Several blue callout boxes with orange borders and arrows point to specific elements: "Sets of connected nodes and edges" points to the "Networks" column; "Genes/drugs/chemicals in network" points to the "Nodes" column; "Interactions between pairs of nodes" points to the "Edges" column; "Nodes in uploaded gene list" points to the "Seeds" column; and "Download to import network into other tools (i.e. Cytoscape)" points to the "Download" links in the "Interactions (.SIF)" column.

Networks	Nodes	Edges	Seeds	Interactions (.SIF)
subnetwork1	499	575	43	Download
subnetwork2	13	12	1	Download
subnetwork3	10	9	1	Download
subnetwork4	6	5	1	Download
subnetwork5	6	5	1	Download
subnetwork6	6	5	1	Download
subnetwork7	5	5	1	Download
subnetwork8	5	5	1	Download
subnetwork9	4	3	1	Download

Network creation

Network management tools



We recommend 200-2000 nodes for practical reasons: larger networks are difficult to interpret while smaller networks will not give systems-level understanding. NetworkAnalyst offers a range of tools for adjusting size:

- **Reset Network:** return to default network
- **Zero-order Network:** keep only seed genes that directly interact with each other (recommended when > 500 seeds)
- **Second-order Network:** add higher-order interactions from the query database (recommended when < 200 nodes)
- **Minimum Network:** keep seed genes as well as non-seed genes that maintain network connection (suitable for simplifying dense networks)
- **Degree Filter:** the degree of a node is its number of connections to other nodes, so this filter retains hub genes
- **Betweenness Filter:** the betweenness measures the number of shortest paths through a node, so this filter retains genes that connect clusters
- **Steiner Forest Network:** similar to minimum network, but based on a fast heuristic PCSF algorithm
- **Batch exclusion:** enter a list of genes to exclude from the network

Network creation

Navigation bar tracks progress through the analysis pipeline

After adding nodes from the whole blood PPI, subnetwork1 has 499 nodes and 575 edges – a reasonable size for effective visualization. It is not necessary to add or remove nodes from this network.

The screenshot displays the NetworkAnalyst web interface. The main content area is titled "Mapping Overview" and contains a table of subnetworks. The table has columns for Networks, Nodes, Edges, Seeds, and Interactions (.SIF). The first row, subnetwork1, has 499 nodes and 575 edges. Below the table are "Previous" and "Proceed" buttons. On the right side, there is a "Network Tools" panel with buttons for "Reset Network", "Zero-order Network", "Second-order Network", "Minimum Network", "Degree Filter", "Betweenness Filter", "Steiner Forest Network", and "Batch Exclusion". A blue callout box with an arrow points to the "Proceed" button, containing the text "Click 'Proceed'" and a circled number "1".

Networks	Nodes	Edges	Seeds	Interactions (.SIF)
subnetwork1	499	575	43	± Download
subnetwork2	13	12	1	± Download
subnetwork3	10	9	1	± Download
subnetwork4	6	5	1	± Download
subnetwork5	6	5	1	± Download
subnetwork6	6	5	1	± Download
subnetwork7	5	4	1	± Download
subnetwork8	5	4	1	± Download
subnetwork9	4	3	1	± Download
subnetwork10	4	3	1	± Download
subnetwork11	4	3	1	± Download
subnetwork12	3	2	1	± Download
subnetwork13	3	2	1	± Download
subnetwork14	3	2	1	± Download

Click "Proceed"

1

Network visualization tools

Here are some of the main network visualization functions. This tool bar is the same for all 2D networks on NetworkAnalyst.

- ✓ subnetwork1
- subnetwork2
- subnetwork3
- subnetwork4
- subnetwork5
- subnetwork6
- subnetwork7
- subnetwork8
- subnetwork9

The screenshot shows the NetworkAnalyst web application interface. At the top, there is a navigation bar with links for Home, FAQs, Tutorials, Updates, and Resources. Below this is a toolbar with various options: Network (subnetwork1), Background (Black), View (Topology), Layout (-- Specify --), Node (- Specify -), Edge (- Specify -), Scope (-- Specify --), and Download (-- Specify --). A 'More Options' button is also present. On the left, there is a 'Node Explorer' table with columns for ID, Name, Degree, Betweenness, and Exp. The table contains several rows of data, including ENSP000 HIST2H2, ENSP000 RAB13, ENSP000 HIST1H1, ENSP000 S100A9, ENSP000 TGM2, ENSP000 EGR2, and ENSP000 PDLIM7. The main area displays a network visualization with nodes and edges. A 'Function Explorer' panel is visible on the right. A 'Background Color' dialog box is open, showing a color selection interface with a color wheel and a 'choose' button. The background color is currently set to #222222.

ID	Name	Degree	Betweenness	Exp
ENSP000	HIST2H2	158	158840	3.61
ENSP000	RAB13	77	895	
ENSP000	HIST1H1	56	431	
ENSP000	S100A9	55	426	
ENSP000	TGM2	50	654	
ENSP000	EGR2	46	41628	2.03
ENSP000	PDLIM7	44	57357	3.1

- Black
- White
- ✓ Custom

- ✓ Topology
- Expression
- Plain

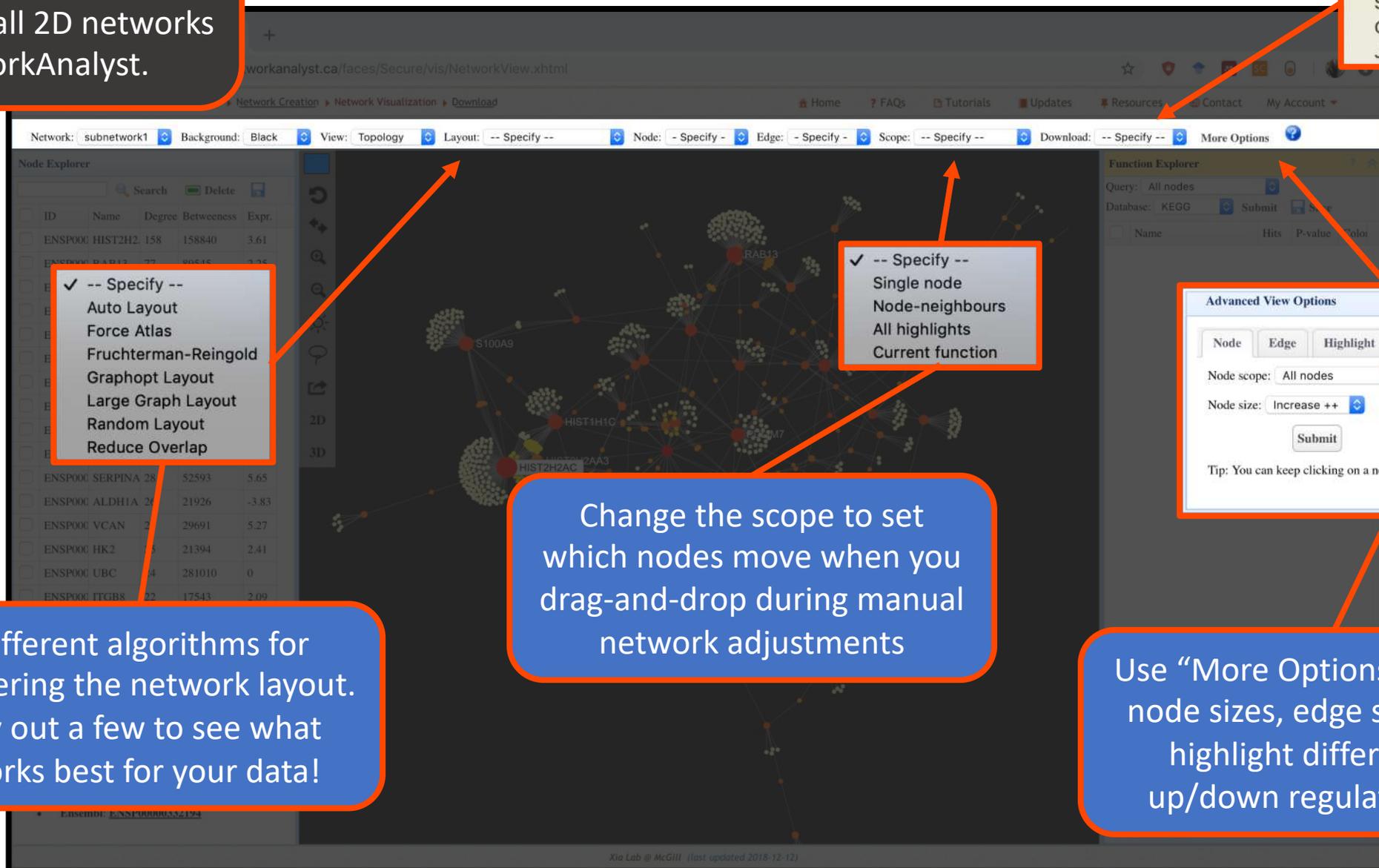
Select to view a different subnetwork of interconnected nodes. Typically subnetwork1 is significantly larger (“continent”) than all others (“islands”).

Topology colors nodes according to their degree, expression colors based on uploaded fold changes, and plain makes all nodes gray (good when you want to highlight specific network regions later on)

Set the background to any color

Network visualization tools

Here are some of the main network visualization functions. This tool bar is the same for all 2D networks on NetworkAnalyst.



- ✓ -- Specify --
PNG Image
SVG Format
GraphML
JSON

- ✓ -- Specify --
Auto Layout
Force Atlas
Fruchterman-Reingold
Graphopt Layout
Large Graph Layout
Random Layout
Reduce Overlap

- ✓ -- Specify --
Single node
Node-neighbours
All highlights
Current function

Advanced View Options

Node Edge Highlight

Node scope: All nodes

Node size: Increase ++

Submit

Tip: You can keep clicking on a node to increase its size

Different algorithms for rendering the network layout. Try out a few to see what works best for your data!

Change the scope to set which nodes move when you drag-and-drop during manual network adjustments

Use "More Options" to change node sizes, edge shapes, and highlight differently for up/down regulated nodes

Network visualization tools

Return network to default view

Autofit view to window, zoom in, and zoom out – more labels will appear as you zoom in

Manually select a batch of nodes to adjust their position (drag and drop)

Switch between 2D and 3D

Highlight all of the seed genes – genes that were in the uploaded list

Extract the highlighted nodes to visualize them separately from the rest of the network. This allows you to view regions of interest in greater detail.

Set the highlight to any color

Here are some of the main network visualization functions. This tool bar is the same for all 2D networks on NetworkAnalyst.

The screenshot displays the NetworkAnalyst web interface. At the top, a breadcrumb trail reads: Upload Data > Network Selection > Network Creation > Network Visualization > Download. Below this, a toolbar contains icons for home, autofit, zoom in, zoom out, highlight, and drag-and-drop. A data table on the left lists genes with columns for Name, Count, and other metrics. A network graph is visible in the center, with nodes and edges. A color selection dialog box is open, showing a color spectrum and the hex code #0080ff. The bottom of the interface includes a 'Current Selections' panel and a footer with the text 'Xia Lab @ McGill (last updated 2018-12-12)'.

Name	Count	Other
ENSP000 SHRM9	42683	2.5
ENSP000 TGM2	50	65494 2.07
ENSP000 EGR2	46	41628
ENSP000 PDLIM4	57357	3.1
ENSP000 UBC	24	281010 0
ENSP000 ITGB8	22	17543 2.09
ENSP000 CPVL	22	31956 -3.65
ENSP000 TLR7	22	21469 -2.23
ENSP000 ADAM15	21	
ENSP000 CD93	20	
ENSP000 MMP7	20	

Detect and select modules (PPI network)

The screenshot displays the NetworkAnalyst web interface. The central part shows a network graph with nodes of various sizes and colors (white, red, green) connected by edges. The interface includes several panels: a 'Node Explorer' on the left with a table of gene lists, a 'Function Explorer' and 'Module Explorer' on the right with a table of modules, and a 'Path Explorer' at the bottom right. The browser address bar shows the URL: https://www.networkanalyst.ca/NetworkAnalyst/faces/Secure/vis/NetworkView.xhtml. The top navigation bar includes links for Home, FAQs, Tutorials, Updates, Resources, and Contact. The main toolbar includes options for Network, Background, View, Layout, Node, Edge, Scope, and Download.

Module	Size	Query	P-value	Color
0	55	4	5.69e-11	Grey
1	34	3	0.00418	Grey
2	33	2	5.5e-11	Grey
3	14	2	2.62e-06	Grey
4	8	2	0.0277	Grey
5	6	2		Grey
6	61	1		Grey
7	44	1	1.24e-18	Grey
8	19	1	9.92e-08	Grey
15	10	1	0.000124	Grey
16	11	1	0.000273	Grey
17	12	1	0.000398	Grey
18	8	1	0.000967	Grey
19	8	1	0.00341	Grey
20	6	1	0.00753	Grey
21	9	1	0.0242	Grey
22	6	1	0.0309	Grey

1 Colour nodes by "Expression" to highlight input gene list

2 Select algorithm and "Submit"

3 Select largest module to highlight component genes

Sometimes networks are still too complex to understand after filtering and adjusting visual factors - breaking them into smaller units (modules) can reduce complexity while keeping the most interesting functions and connections

Perform gene set ORA on top module (PPI network)

Gene set overrepresentation analysis (ORA) is one method of exploring the biological meaning of computationally defined modules.

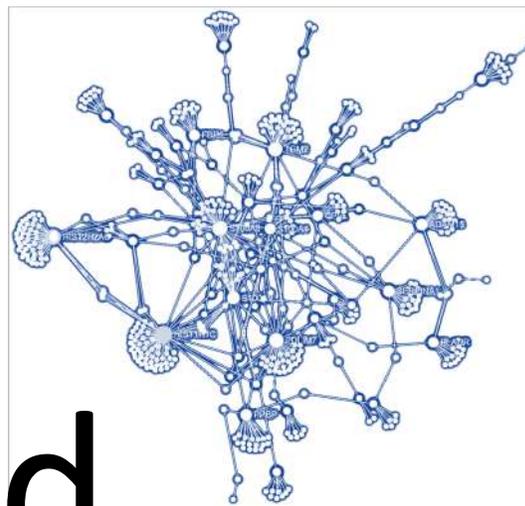
Change query to "Highlighted nodes"

Select database of interest and "Submit"

The largest module appears to be enriched in immune-related pathways

The screenshot shows the NetworkAnalyst web interface. On the left is a list of genes with columns for gene ID, symbol, count, and other metrics. In the center is a network graph with nodes and edges. On the right is the 'Function Explorer' panel, which includes a table of biological processes and their enrichment statistics.

Name	Hits	P-value	Color
Hemostasis	9	0.000038	
Platelet activation, signalin	6	0.000097	
Meiotic Synapsis	4	0.000139	
Immune System	12	0.000205	
Cell Cycle	8	0.000262	
Adaptive Immune System	9	0.000266	
Packaging Of Telomere En	3	0.000635	
Meiosis	4	0.000726	
CD28 dependent Vav1 patl	2	0.0008	
Chromosome Maintenance	4	0.000903	
SHC-mediated signalling	2	0.00126	
SHC1 events in EGFR sigr	2	0.00126	
Signal attenuation	2	0.00144	
Spry regulation of FGF sig	2	0.00144	
Costimulation by the CD2	3	0.00175	
Innate Immune System	7	0.00181	
SHC-related events	2	0.00183	
SHC-related events trigger	2	0.00183	
Signaling by ERBB4	4	0.00192	
Signaling by constitutively	2	0.00204	
Telomere Maintenance	3	0.00207	
Tie2 Signaling	2	0.00226	



The End

*For more information, visit the **FAQs, Tutorials, Resources,**
and **Contact** pages on www.networkanalyst.ca*