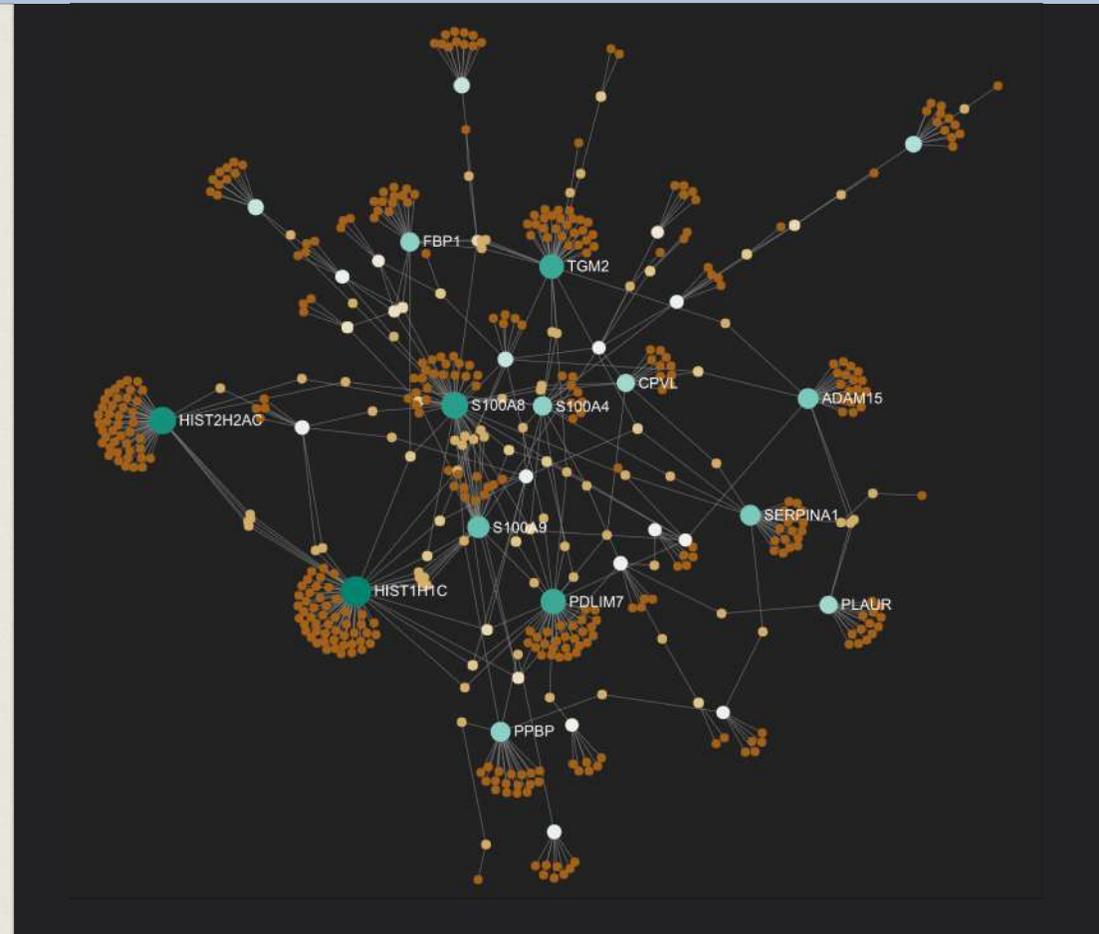
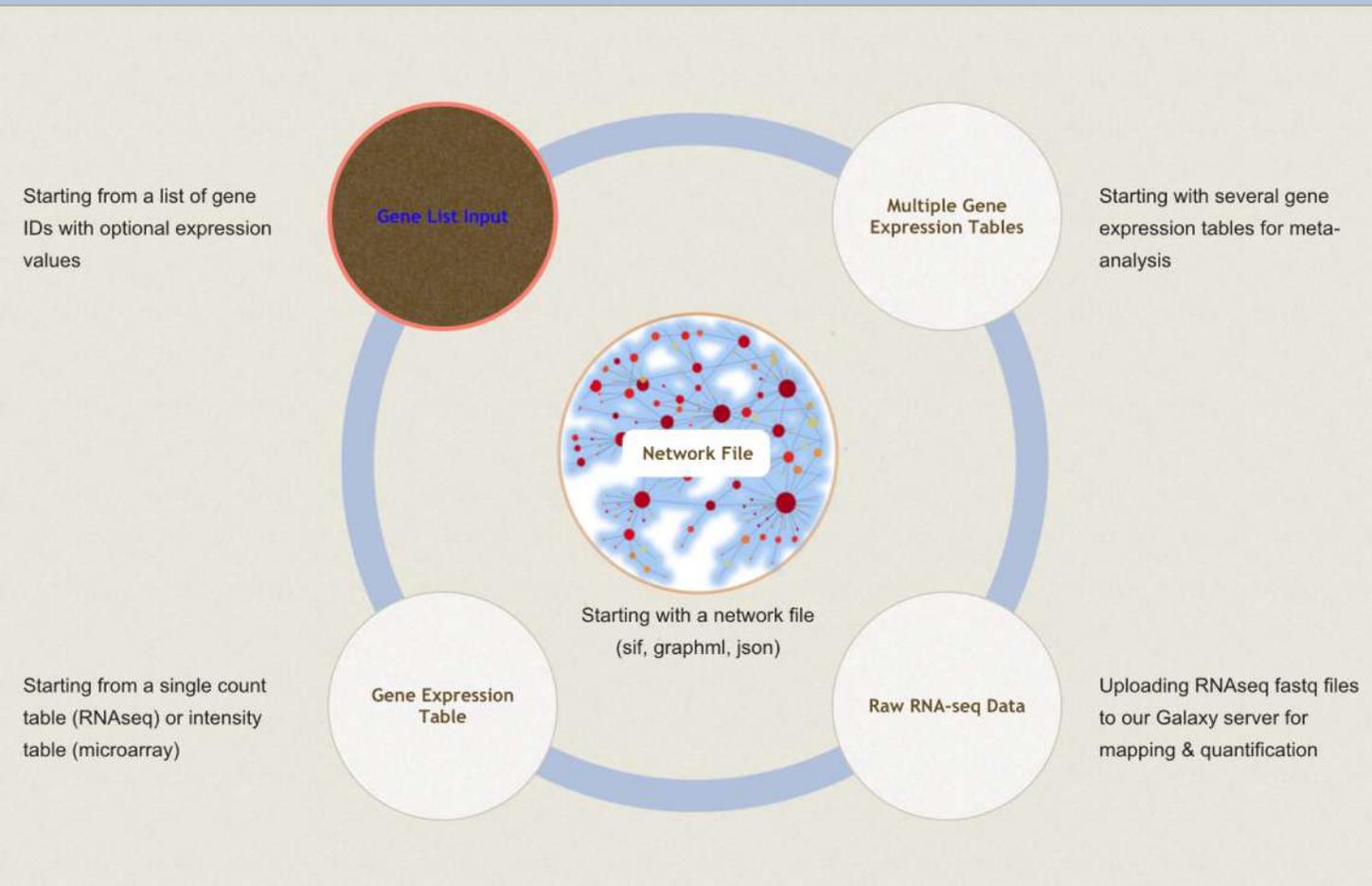


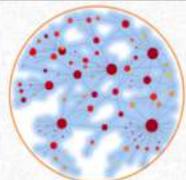
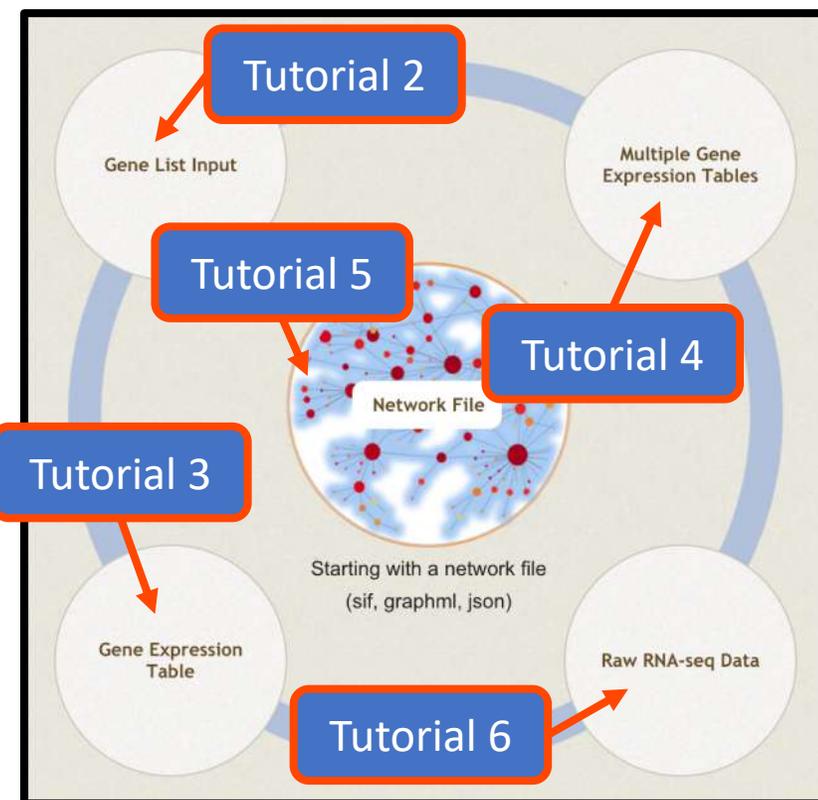
Tutorial 2a: single gene list input



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
```

Upload Try Example

Previous Proceed

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

Click "Upload"

1

Click "Proceed"

2

```
#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.

The screenshot displays the NetworkAnalyst web interface. At the top, there is a navigation bar with links for Home, FAQs, Tutorials, Gallery, Updates, and Resources. Below this, a main menu lists several analysis functions: Enrichment Network, Heatmap View, Venn Diagram, Chord Diagram, Protein-protein Interactions (PPI), Gene Regulatory Networks (GRN), Diseases, drugs & chemicals, and Gene Coexpression Networks. Each of these main functions has a sub-menu of more specific options. For example, PPI includes Generic PPI and Tissue-specific PPI; GRN includes Gene-miRNA Interactions, TF-gene Interactions, and TF-miRNA Coregulatory Network; Diseases, drugs & chemicals includes Protein-drug Interactions, Protein-chemical Interactions, and Gene-disease Associations; and Gene Coexpression Networks includes Tissue-specific Coexpression and Cell-specific Coexpression. An orange arrow points from a blue callout box to the 'Heatmap View' option in the main menu.

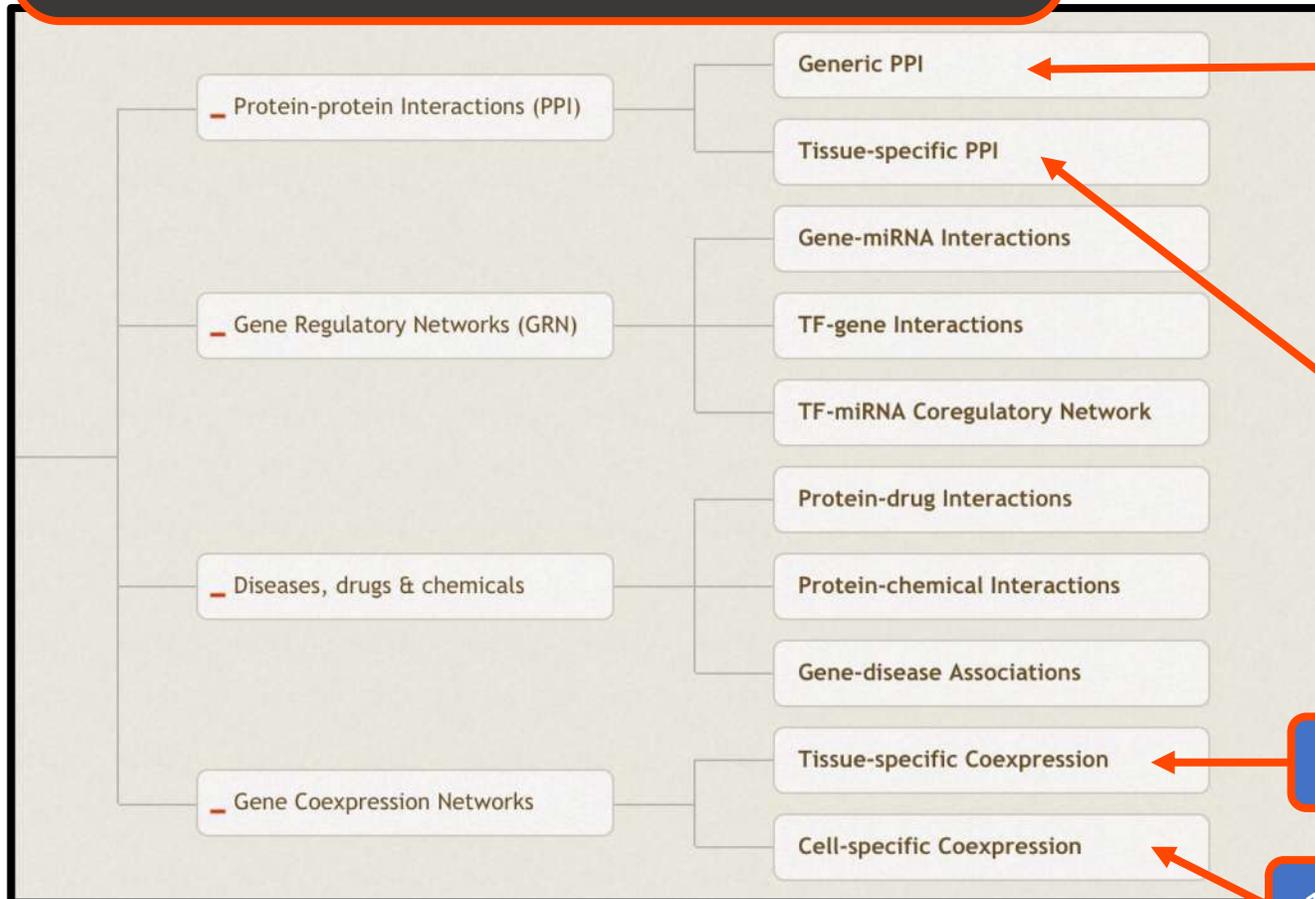
See tutorials 2b, 3 and 4 for more details on these functions

Xia Lab @ McGill (last updated 2019-01-23)

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: 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: 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

2

Click "Ok"

The screenshot shows the NetworkAnalyst web application interface. A dialog box titled "Select a tissue type for PPI network:" is open, displaying a list of tissue types. "Whole blood" is selected and highlighted in orange. Below the list is a "Filter" input field with the value "15.0" and a question mark icon. An "OK" button is visible at the bottom right of the dialog. The background shows the main application interface with a sidebar on the left containing buttons for "Upload Data", "Network Selection", "Network Creation", "Network Visualization", "Download", and "Exit". The main content area has a heading "Please choose a network type to proceed" and a paragraph of text about visual analytics technology. At the bottom of the page, there is a footer that reads "Xia Lab @ McGill (last updated 2018-12-08)".

Network creation

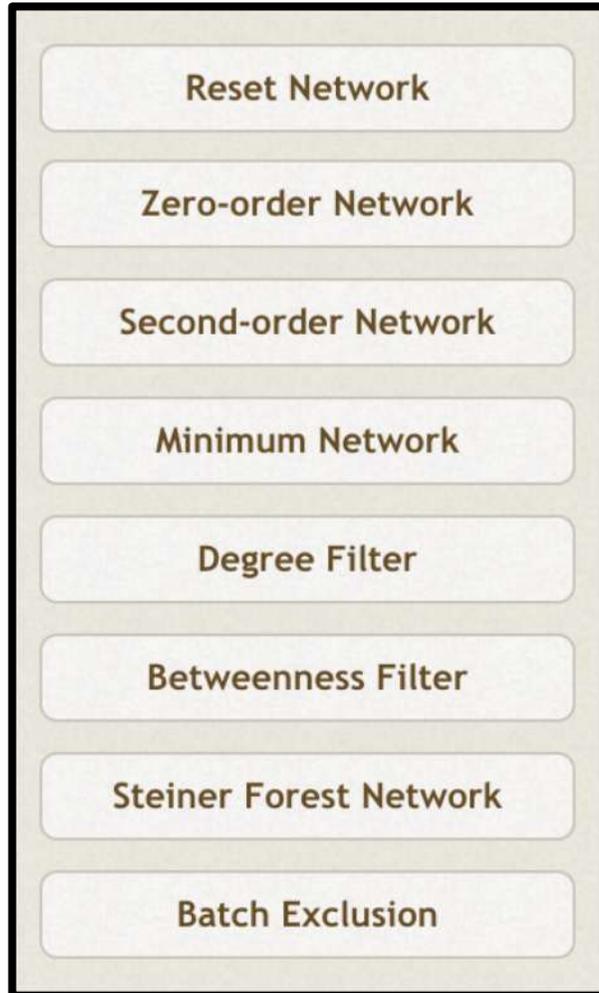
After selecting the network type, you will see a panel of **tools** for network construction, and a **mapping overview** summarizing the network statistics.

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 (expanded), 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 offers various functions: 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 application interface. The browser address bar shows the URL: `http://www.networkanalyst.ca/NetworkAnalyst/faces/Secure/network/NetworkBuilder.xhtml`. The page title is "NetworkAnalyst -- a comprehensive platform for gene".

The interface is divided into three main sections:

- Navigation Bar (Left):** A vertical menu with a home icon at the top. The menu items are: "Upload Data", "Analysis Overview", "Network Analysis" (expanded), "Network Creation" (highlighted in orange), "Network Visualization", "Download", and "Exit".
- Main Content Area (Center):**
 - Mapping Overview:** A text block explaining that significant genes (seeds) are mapped to a molecular interaction database, resulting in a large subnetwork ("continent") and smaller ones ("islands"). It notes that subnetworks with at least 3 nodes are listed.
 - Table:** A table listing 14 subnetworks with their respective node and edge counts.
 - Network Tools (Right):** A vertical list of buttons for network manipulation: "Reset Network", "Zero-order Network", "Second-order Network", "Minimum Network", "Degree Filter", "Betweenness Filter", "Steiner Forest Network", and "Batch Exclusion".
- Footer:** "Xia Lab @ McGill (last updated 2019-01-18)".

Annotations on the screenshot include:

- A blue callout box with an orange border pointing to the "Network Creation" menu item.
- A blue callout box with an orange border pointing to the "Proceed" button at the bottom right, with a circled "1" next to it.

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 options like 'Upload Data', 'Network Selection', 'Network Creation', 'Network Visualization', and 'Download'. Below this is a toolbar with various settings: 'Network: subnetwork1', 'Background: Black', 'View: Topology', 'Layout: -- Specify --', 'Node: - Specify -', 'Edge: - Specify -', 'Scope: -- Specify --', and 'Download: -- Specify --'. On the left, there is a 'Node Explorer' table with columns for ID, Name, Degree, Betweenness, and Exp. The main area displays a network graph with nodes and edges. On the right, there is a 'Function Explorer' panel. 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	65	
ENSP000	EGR2	46	41628	2.03
ENSP000	PDLIM7	44	57357	3.1
			3133.5	3.97
			1019	2.7
			27	-2.71
			93	5.65
			26	-3.83
			91	5.27
			94	2.41
			010	0
			43	2.09
			56	-3.65
			69	-2.23
			72	-2.09
			24	4.58
			44	17.04

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

Black
White
✓ Custom

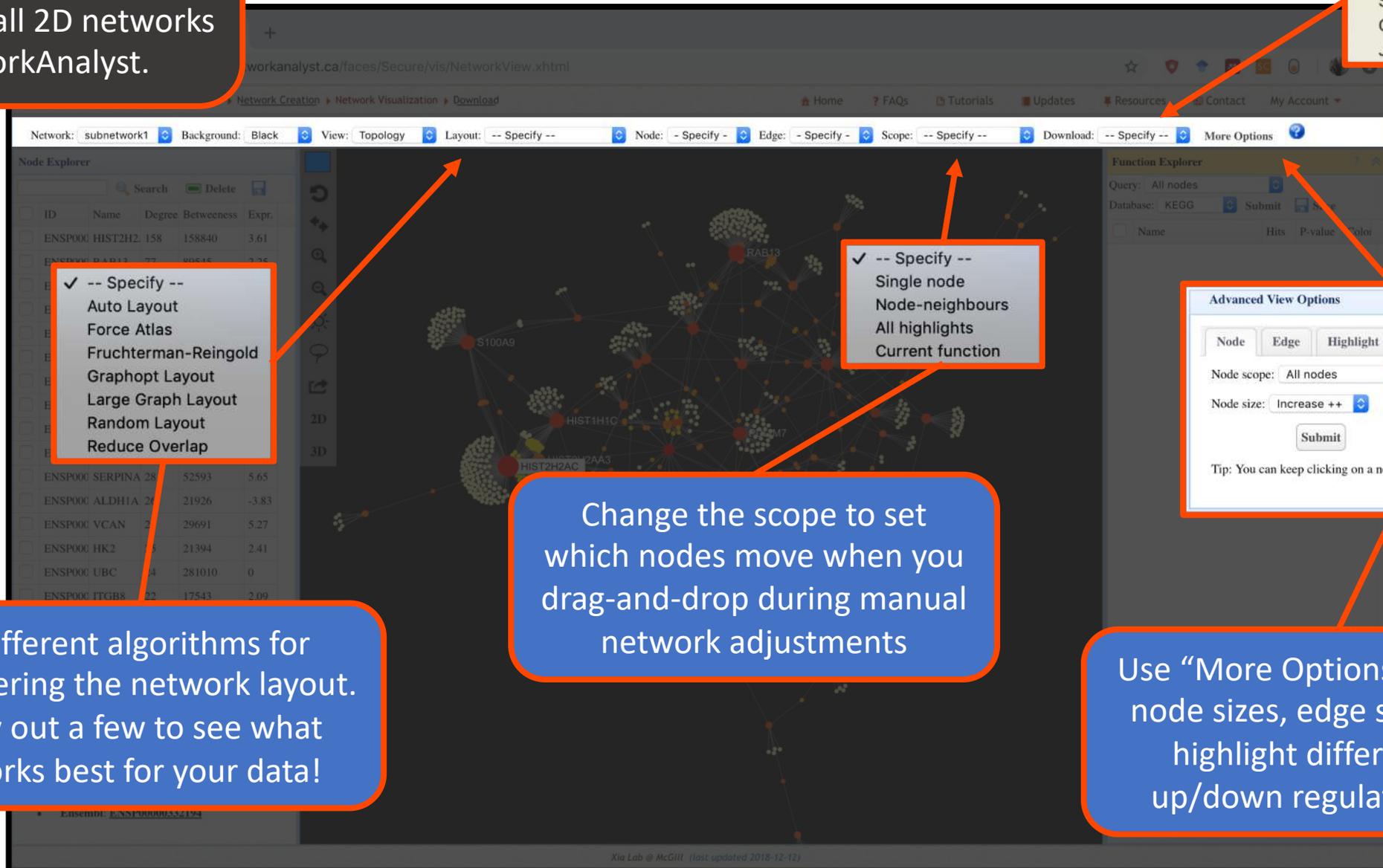
✓ Topology
Expression
Plain

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

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, and a 'Path Explorer' at the bottom. The 'Module Explorer' table is highlighted with a blue box and contains the following data:

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

Four numbered callouts (1, 2, 3, 4) point to specific UI elements: 1 points to the 'View: Expression' dropdown; 2 points to the 'Submit' button; 3 points to the 'WalkTrap' algorithm dropdown; 4 points to the first row of the 'Module Explorer' table.

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"

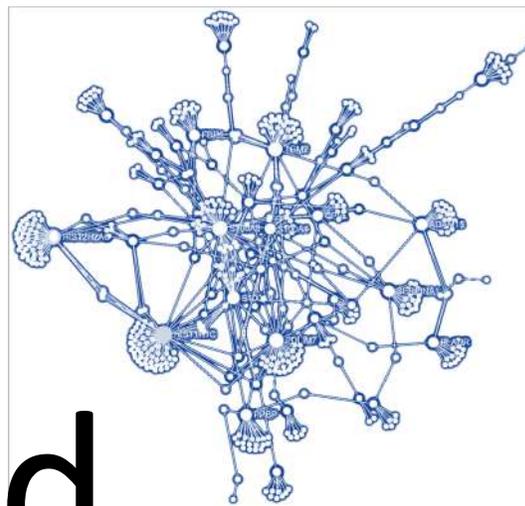
Select database of interest and "Submit"

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

The screenshot shows the NetworkAnalyst web interface. On the left, a table lists genes with their IDs, names, and various statistics. The central part of the interface displays a network graph with nodes and edges, where a cluster of nodes is highlighted in blue. On the right, the 'Function Explorer' panel shows a table of biological pathways with their respective hit counts and p-values. Three numbered callouts (1, 2, 3) point to the 'Highlighted nodes' query, the 'Reactome' database selection, and the 'Submit' button, respectively.

Gene ID	Gene Name	Count	Other Value 1	Other Value 2
9260	PDLIM7	41	20282	3.1
7052	TGM2	38	19804	2.07
6280	S100A9	28	13111	2.5
5265	SERPINA	23	12585	5.65
8751	ADAM15	23	13671	-2.09
5473	PPBP	21	15523	6.76
2203	FBP1	19	7687.5	3.18
6275	S100A4	19	16834	-2.71
54504	CPVL	16	6914.7	-3.65
5329	PLAUR	16	6516.9	2.7
1462	VCAN	13	6382	5.27
55755	CDK5RA	12	5030.6	2.22
4084	MXD1	12	5412	2.02
3099	HK2	11	5515.6	2.41
2710	GK	9	3948	2.3
54541	DDIT4	9	2473.3	2.1
22918	CD93	9	7730.8	2.1
1839	HREGF	8	17067	2.1

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*