NCIBI Workshop

Module 3

Visually Exploring Protein Interaction Networks and Pathway Associations

Exploring Protein Interaction Networks with MiMI Plugin to Cytoscape

This tutorial assumes that you have already:
- Installed Cytoscape to your desktop (from cytscape.org)
- Downloaded the following plug-ins – following Cytoscape directions:
  MiMI plug in (under Network and Attribute I/O)
  Cytoscape Tutorial (under Theme, includes BINGO and MCODE)
  Enhanced Search (under Analysis)
  Shortest Path

Getting Started

Accessing the Datasets to Explore with the MiMI Plug in for Cytoscape

Two datasets and one Excel spreadsheet are used in this tutorial. Please download them from:
http://portal.ncibi.org/gateway/ws_resources.html

Dataset 1. Seven differential genes for Part 1 associated with iron.
They include:
CP
HFE
SLC39A14
FTL
LCN2
TF
UGT1A1

These 7 genes (human orthologs) are significantly associated with iron homeostasis and heme metabolism (Fisher's exact score for term enrichment) – both of which have been implicated in prion disease. The genes are from Hwang et al's (2009) list of 333 candidate genes for susceptibility to prion replication and disease, derived from experiments with mice and microarray analysis to find differentially expressed genes. Five of the 7 (all but LCN2 and UGT1A1) are in the "short list" of 32 genes that Hwang et al found to be differentially expressed in all 5 strains of the prior-infected mouse brains.

Notably, none of the 333 candidate genes associated with cholesterol and sphingolipid metabolism was found to be differentially expressed in all 5 strains. Yet cholesterol and sphingolipid metabolism are critical to prion replication.

In this tutorial, you will explore if some connection can be found between iron homeostasis and heme metabolism, on the one hand, and cholesterol and sphingolipid metabolism, on the other.

Dataset 2. Fifteen differential genes for iron and sphingolipids

CP	TF	SPGL1
HFE	UGT1A1	EDG3
SLC39A14	NPC2	SOAT1
FTL	B4GALT1	ABCA1
LCN2	PLD4	APOD

Excel Spreadsheet. GO concepts distinguishing the gene in Dataset 2. You will import this file to Cytoscape and use it for color coding.

Background Problem, Purpose and Tutorial Objectives

Biology Problem and Purpose of Your Analysis

This tutorial shows how Cytoscape and the MiMI Plugin reveal interactions among gene products and how to group interactions to uncover fundamental processes relevant to prion replication and disease.

Background information about prion disease: Neurodegeneration proceeds from accumulation of PrPSc in cholesterol rich lipid rafts to early synapse degeneration to nerve cell death.

"Dynamic perturbations of gene expression associated with biological processes occur in a specific order, [reinforcing] that neuropathological changes proceed in a stereotypical set of steps (Bouzamondo-Bernstein *et al*, 2004). Among the first detectable expression changes (10 weeks) [in prion susceptibility are] increases in mRNA levels of *C1qa/b* and *C3ar1* reflecting activation of microglia and astrocytes, one of the earliest detectable responses to prion infection (Klein *et al*, 2001) (Fig. 4B, module 1). Around the same time, the increased activities of lysosomal proteases [are] suggested by elevation of the lysosomal protease gene *Ctsc* (Fig. 4B, module 2): lysosomal proteases may have a role in degradation as well as formation of PrPSc (Zhang *et al*, 2003). Changes within the lysosomal module [may become] more obvious at later time points.

"[Also present are] dysregulation associated with the synthesis, trafficking and processing of cholesterol and sphingomyelin, two of the major components of lipid rafts where the initial steps of PrPSc replication are thought to occur (Simons and Ehehalt, 2002). At 10 weeks changes [have been] detected in expression of genes controlling trafficking and conversion of cholesterol and sphingomyelin (Rodriguez-Agudo *et al*, 2008). At 14 weeks expression levels of cholesterol synthesis genes, such as *Mvk* and *Mvd* ... decrease, but expression of the cholesterol transporters or acceptors including *Abca1*, *A2m*, and *Apod* increase...Decreases in expression of genes involved in cholesterol biosynthesis are particularly pronounced in mice infected with the RML prion strain.

"Changes in sphingolipid metabolism, potentially leading to a decrease in galactosylceramide, which is found in PrPSc deposits (Klein *et al*, 1998), are evident by 18 weeks. An increase in GAG metabolism, indicated by increased *Gusb* and *Hexb* mRNA levels, [is] first detectable at 14 weeks (Mayer-Sonnenfeld *et al*, 2005)."

Summary points:
- Androgen metabolism may be involved in prion disease as it is tied to glial activation. Microglial activation and reactive astrocytic gliosis reflect responses to accumulation of PrPSc and are one of the earliest detectable responses to prion infection.
- Iron homeostatis and heme metabolism may play a complicated role. Activation of the heme degradation pathway in the brain has been associated with cellular response to oxidative stress. Regulation of protective functions against oxidative stress through iron homeostasis is essential for maintaining oligodendrocytes and myelin the brain. Dysregulation may relate to demyelination in prion disease
- Additionally, increased activities of lysomal proteins may play a role in degradation as well as in the formation of PrPSc (abnormal disease-produc8ing protein).
- The control and conversion of cholesterol and sphingomyelin are also implicated in prion replication. A decreased expression of genes involved in cholesterol biosynthesis has been pronounced in mice infected with prion disease. Dysregulated cholesterol metabolism may be coupled to the perturbation of neurosteroid metabolism. Furthermore, a close physical relationship among cholesterol, sphingolipid metabolism on lipid rafts in the plasma membrane, and galactosylceramide (GAG) found in PrPs deposits seems to be necessary for prior replication.

Reference:

Daehee Hwang1,Inyoul Y. Lee, Hyuntae Yoo, Nils Gehlenborg, Ji-Hoon Cho, rianne Petritis , David Baxter1, Rose Pitstick2, Rebecca Young2, Doug Spicer2, John Hohmann4, Stephen J. DeArmond3,
George A. Carlson2†, Leroy E. Hood (2009) Dynamical prion disease networks. Forthcoming

Objectives: In this Tutorial you will:
Part I
> ► Query on a list of genes
> ► Set up views to facilitate exploratory analysis, e.g. layouts, selection
> ► Explore details to conceptually characterize gene products and interactions
> ► Expand a network node (gene product) to see additional direct interactors
> ► Find literature associated with interactions to help verify the interactions

Part II
> ► Read in additional values related to the genes on your list
> ► Color code nodes by a selected attribute
> ► Link out to details, supporting literature, and subgraph matches with KEGG
> ► Find shortest paths between 2 genes and identify the mediating gene products
> ► Create network clusters and explore their traits

Steps

1. Open Cytoscape by clicking its icon on your desktop.

2. Go to the top menu: click **Plugins → MiMI Plugin → Query**.
A query dialogue box appears.

3. In the dialogue box, click **From File tab → Load gene file**. Select the first dataset that you downloaded to your desktop comprising iron-related genes. Click Open. A parameter setting dialogue box appears.

4. Leave all the defaults intact. Press Search. Cytoscape displays the 7 genes, their direct interactors, and interactions between the displayed genes.
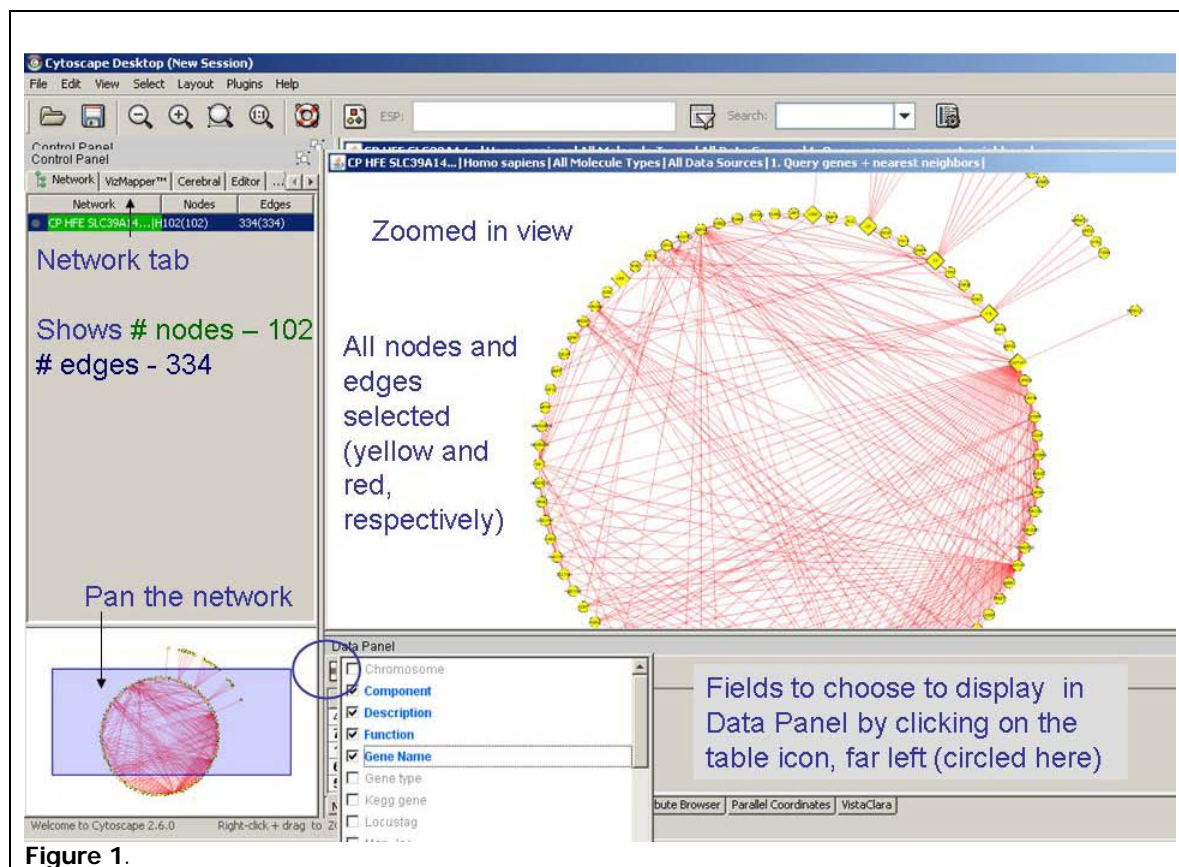
Figure 1.

4. Set up the display information about the connected gene products by doing the following (as highlighted in Figure 1):

    a. Click on the Network tab in the left hand panel to display the view shown in Figure 1. You see that a total of 102 gene products are displayed (the 7 genes on your list plus their interactors).

b. Size the network pane so that it fills the whole display window.

c. Select all nodes by making a box around them with your mouse (Alternatively, hold down Control+Alt+A). The nodes become highlighted as yellow to indicate selection. Edges are red.

d. Click on the Table icon in the lower Data Panel (directly under the word 'Data') to select data attribute fields to display. A pop up check list appears.

e. On the pop up, check: **Component, Description, Function, Gene name, Pathway**.

f. Click anywhere outside the pop up box to close it.

g. Information fills the data panel. Drag column headings to re-order the columns.

h. Zoom in by clicking the magnifying glass with the plus. The lower left blue translucent box in the Network tab (left) gets smaller. Move this box across the network context with your mouse to pan across the large network view.

## Exploring Results on Interactors with Iron-Related Genes

Steps

1. Look for pathways associated with the iron-related genes that may be related to sphingolipids or cholesterol. Click the column heading Pathway to sort alphabetically. Scroll through the list.

To scan more easily, undock the table from the Data Panel and expand it. Click the "docking icon." Maximize the table. To dock it again, click the icon.
See Figure 2.
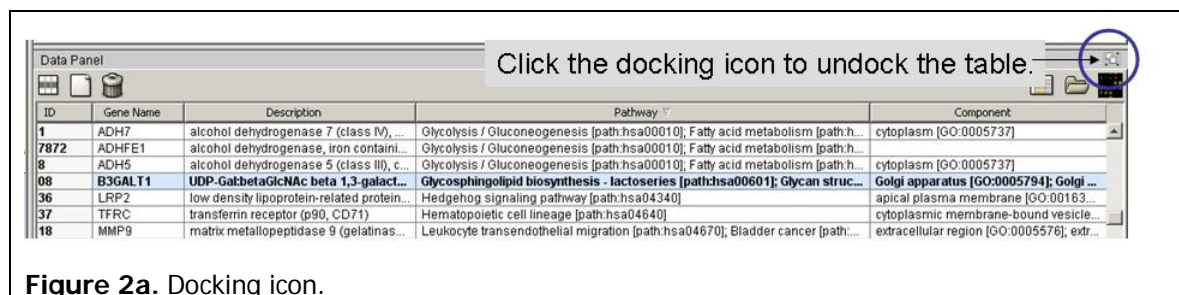
Figure 2a. Docking icon.
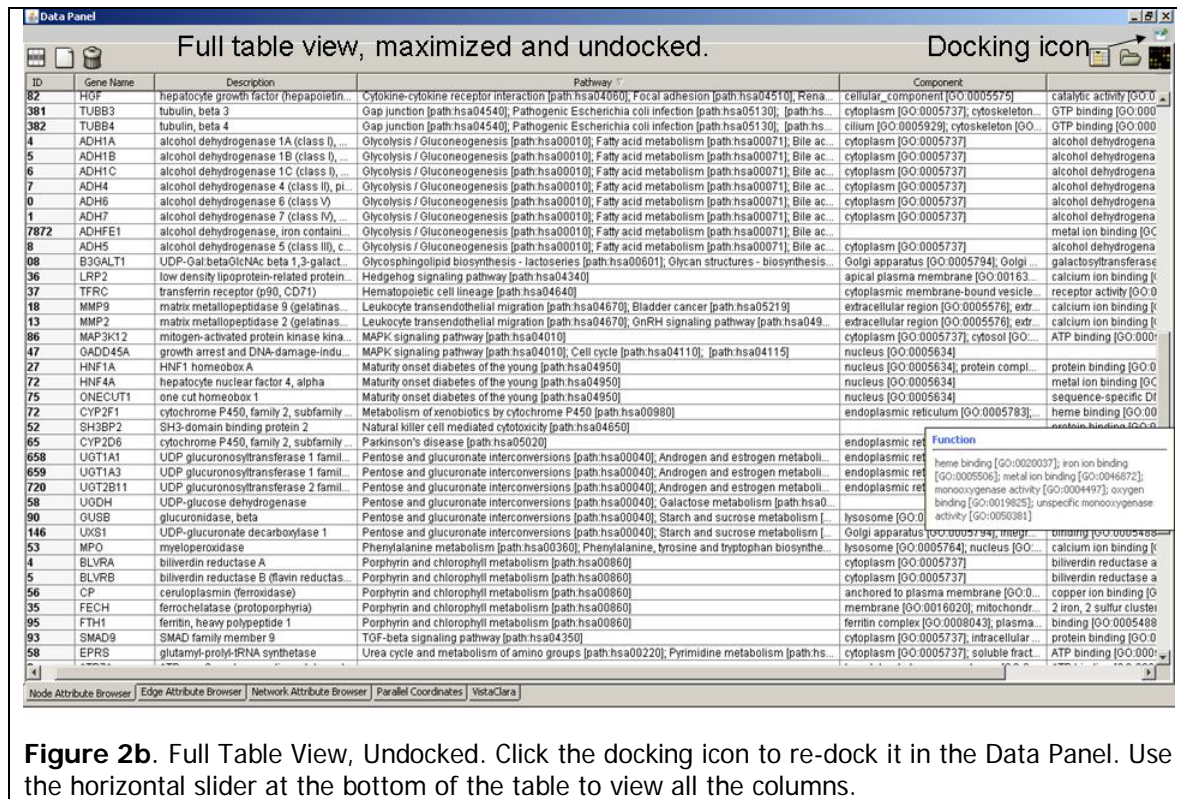
Figure 2b. Full Table View, Undocked. Click the docking icon to re-dock it in the Data Panel. Use the horizontal slider at the bottom of the table to view all the columns.

2. The pathway of one gene, ARSD, interests you: Androgen and estrogen metabolism [path:hsa00150]; Sphingolipid metabolism [path:hsa00600]. You now focus on this gene. Redock the table by clicking the docking icon.

3. Click on ARSD in the table to see where it is in the network. Its node turns green. Zoom in to see its neighbors. It is hard to see interactors because of the many crossed edges (links between nodes). Try a different layout.

4. To change the layout, go to the top menu and click: **Layout → Cytoscape Layout → Edge-weighted Spring Embedded →All nodes → (unweighted)**. A new layout appears (See Figure 3).
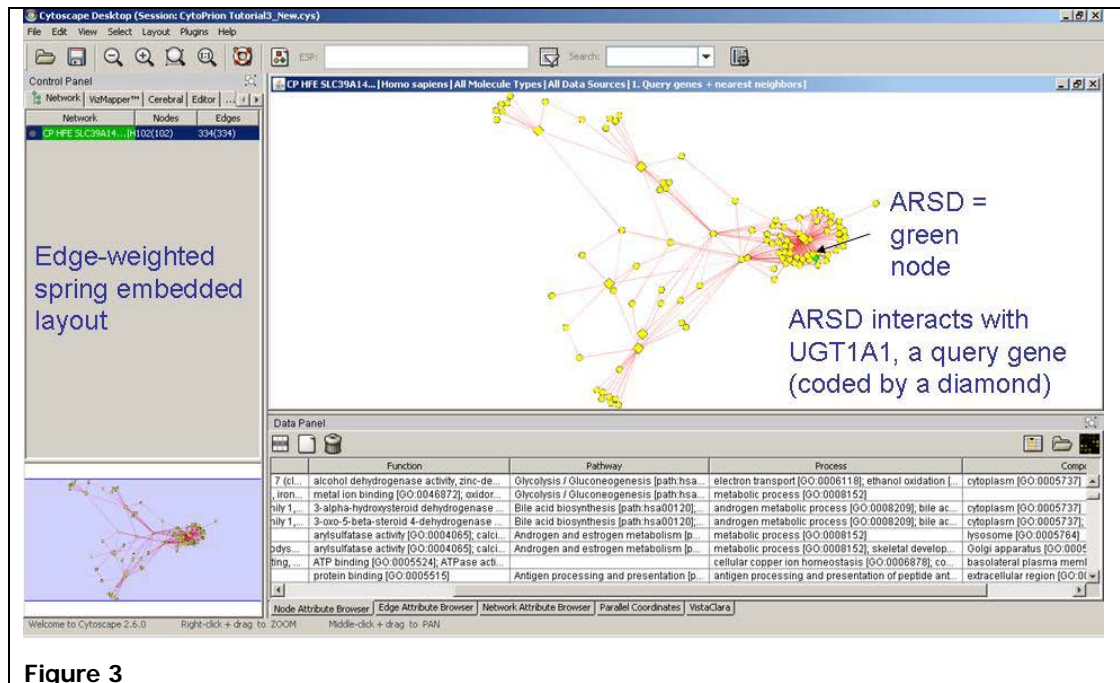
Figure 3

5. Zoom in on ARSD with the magnifying glass with the + . You also can zoom with Right-mouse-down + Slide-mouse-across-screen. Use the translucent blue box to pan the graphic.

You see that a diamond node is linked to the green ARSD node. Hover your mouse over the diamond. It is UGT1A1, one of the iron-related genes.

Verify Interactions by Finding Literature About Them

Steps

 1. Still in the same view as above, verify if interactions between ARSD - with sphingolipid metabolism implications - and UGT1A1 – an iron association – are covered in the research literature.

2. Follow the steps in Figure 4.

Steps:

1. Zoom and pan to see ASRD

2. Click on ARSD in the network. It turns yellow.

3. Pull ARSD out of the graph to see it better and to **pull UGT1A1 out with it**.

4. Add UGT1A1 and B3GALT1 to the selection by Shift+Clicking on each node.

5. Click **Edge Attribute Browser** in the bottom tabs of the Data Panel
6. Click the table icon. On the pop up check list, select : Gene, Name, Provenance, PubMed

Figure 4

3. The Edge Attribute information shows no literature for the ARSD- UGT1A1 interaction. There is one PubMedID for B3GALT1 - UGT1A1 from the curated GRID database [PMID:12925779].

You look up the article and find Sprong et al (2003) **Association of the Golgi UDP-galactose transporter with UDP-galactose:ceramide galactosyltransferase allows UDP-galactose import in the endoplasmic reticulum.** This article tells you that the proteins are co-immunoprecipitated. You are encouraged.

ARSD is more promising a lead since it has more interactors than B3GALT1. You will keep focusing on ARSD.

Verify Interactions by Finding Literature About Them

Steps
1. On the same view, you now want to see all of ARSD's direct interactors. Click on ARSD to select it. Then go to the upper menu and click: **Selec**t → **Nodes** → **First neighbors of selected nodes**. Its neighbors turn yellow. Drag ARSD out to see the interactors better. Scan the table, which now displays annotation data on only these selected genes. Nothing particularly interesting jumps out at you.
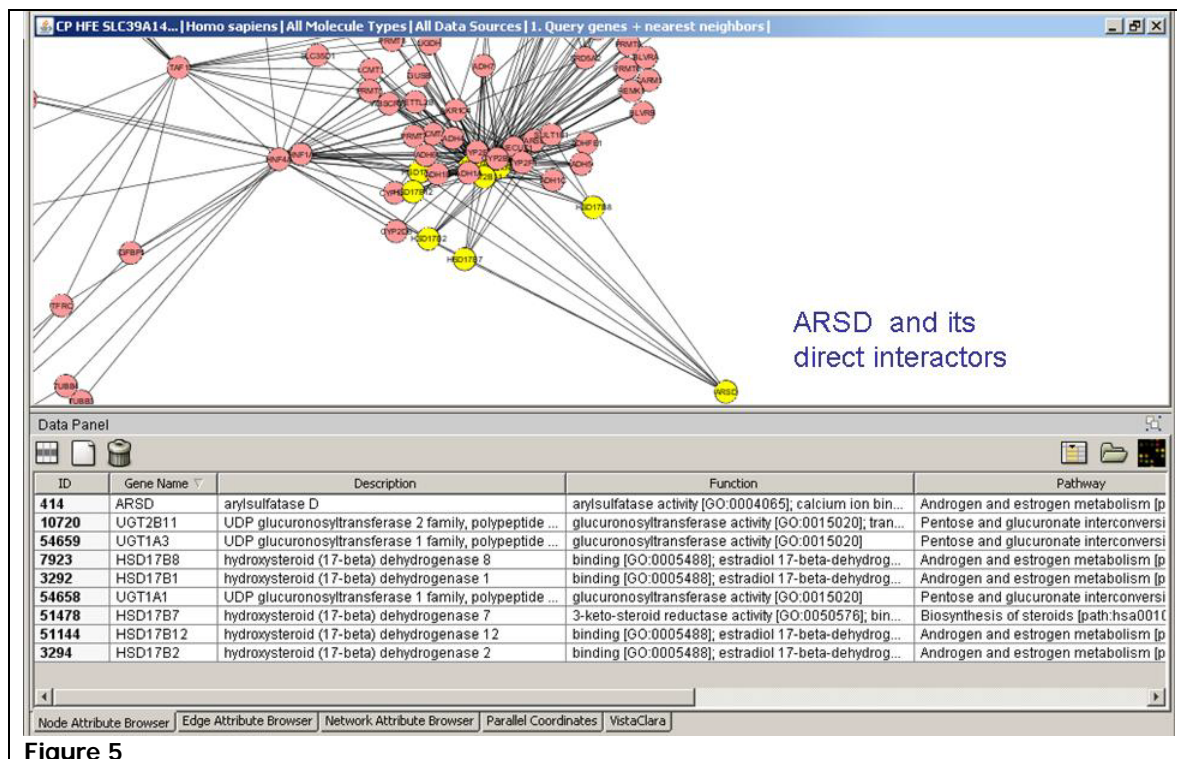
Figure 5

Try one more way of checking out potential connections between sphingolipids via ARSD and iron-related genes and interactions.

2. Double click ARSD. It expands to all nearest neighbors that are known about it in the MiMI database (See Figure 6). The set up has to be repeated.

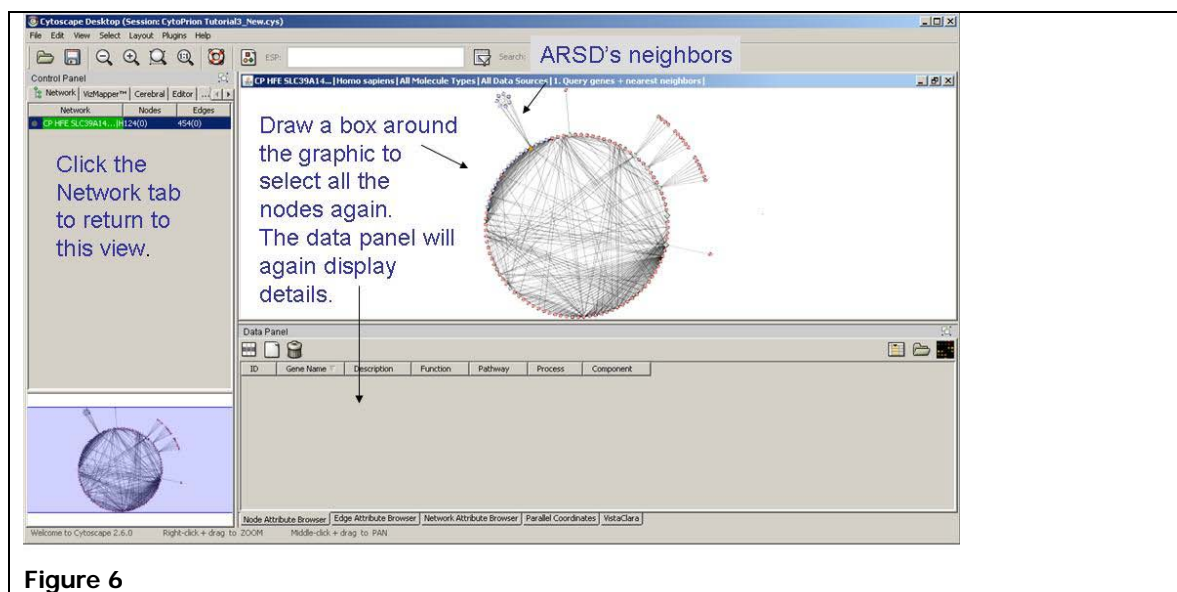3. To set up the graphic, follow the steps in Figure 6.

Figure 6

4. When the view is set up, click back to the Node Attribute Browser tab in the lower tabs beneath the lower Data Panel. Zoom in and use the translucent blue box (lower left) to move to ARSD. ARSD is designated by a triangle as shape-coding for nodes you expand to their neighbors.

5. Select ARSD and its newly expanded neighbors by drawing a box around them. They turn yellow. ARSD is now a triangle – the shape that indicates that this is a gene that you have expanded. You can collapse the expansion by clicking on ARSD again.

6. The details of the selected gene products appear in the Data Panel table below. For these neighbors, you now see a number of Sphingolipid-related connections. Inferring that a connection may be possible, you decide it is worth querying on both the iron- and sphingolipid-related genes from the Hwang et al study at once together to find connecting links.
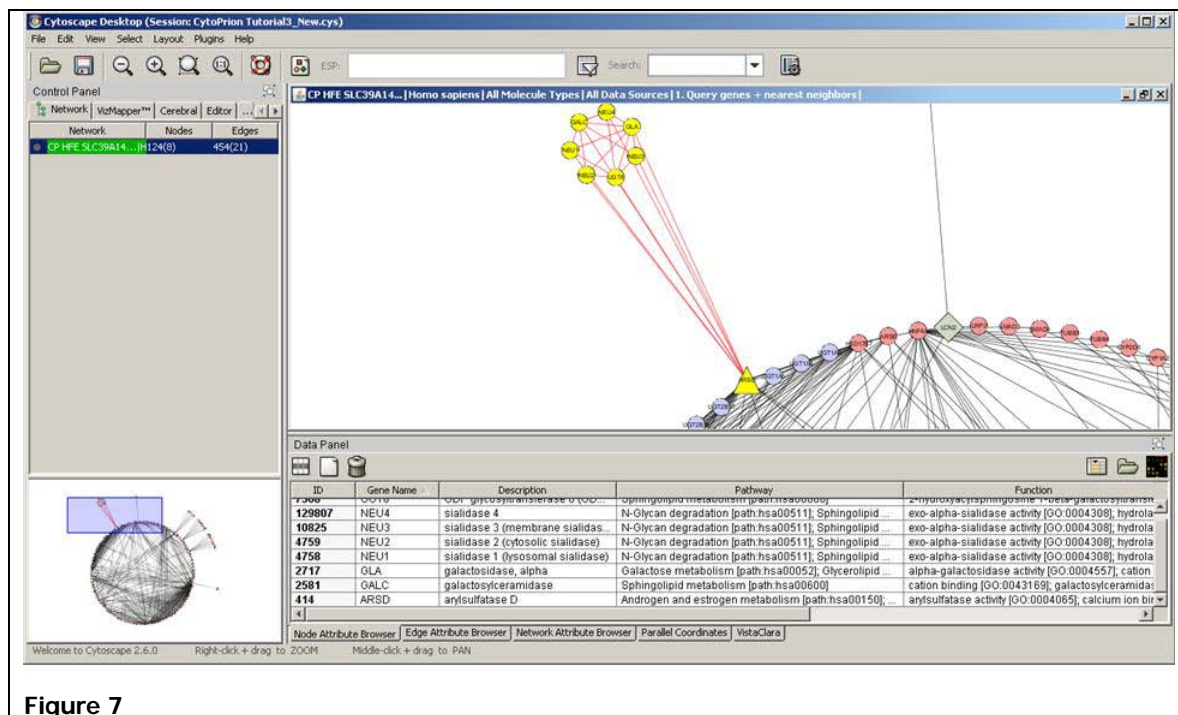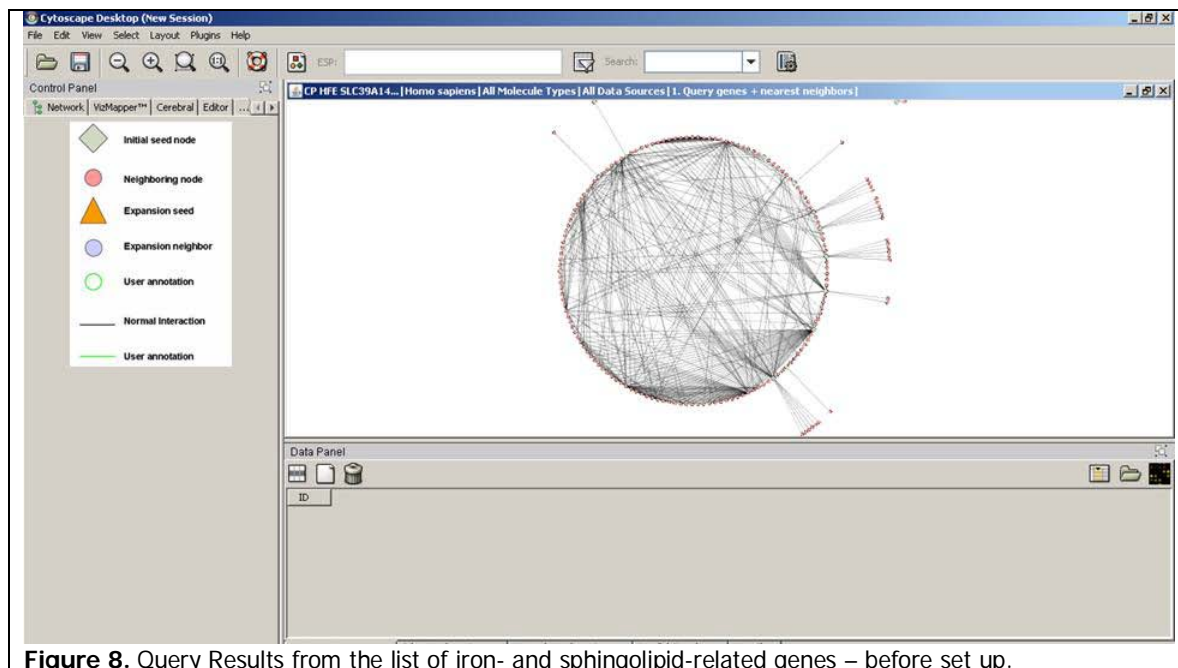

Figure 7

Querying on Iron- and Sphingolipid-related Genes Together

Steps

1. Run a new query. On the top menu, click: **Plugins →MiMI Plugin → Query**.

2. In the query dialogue box, click the **From File tab → Load gene file**.
Load the second dataset that you are using for the tutorial –"HemeANDSphingolipidGenes.txt"
Click Open. A parameter setting dialogue appears.

3. Keep all the defaults intact and pull down the Retrieve options. Select "Query genes and nearest neighbors." Click Search. The network appears (Figure 8).

Figure 8. Query Results from the list of iron- and sphingolipid-related genes – before set up.

4. Set up the view as you did above in Part 1 (repeated below).

   a. Click on the Network tab in the left hand panel to display the view shown in Figure 1. You
       see that a total of 175 gene products are displayed (genes on your list plus their
       interactors).

   b. Size the network pane so that it fills the whole display window.

   c. Select all nodes by making a box around them with your mouse (Alternately use
       Control+Alt+A). The nodes become highlighted (yellow) to indicate selection. Edges become
       red.

   d. Click on the Table icon in the Data Panel under the network view to select data fields to
       display.

   e. Choose these fields: Component, Description, Function, Gene name, Pathway.

   f. Click anywhere outside the list to close it.

   g. Information now fills the data panel. Drag column headings to put the columns in an order
       you like.

   h. Zoom in by clicking the magnifying glass with the +. The blue translucent box in the
       Network tab (left) gets smaller. Move this box with your mouse to pan across the network.

   i. Lay out the graph in Weighted Spring Embedded form: : **Layout → Cytoscape Layout →
       Edge-weighted Spring Embedded →All nodes → (unweighted)**

Steps

1. Locate ARSD in the graph that you just set up. Type ARSD in the ESP search box and press ENTER. ARSD turns yellow but it is hard to find it.
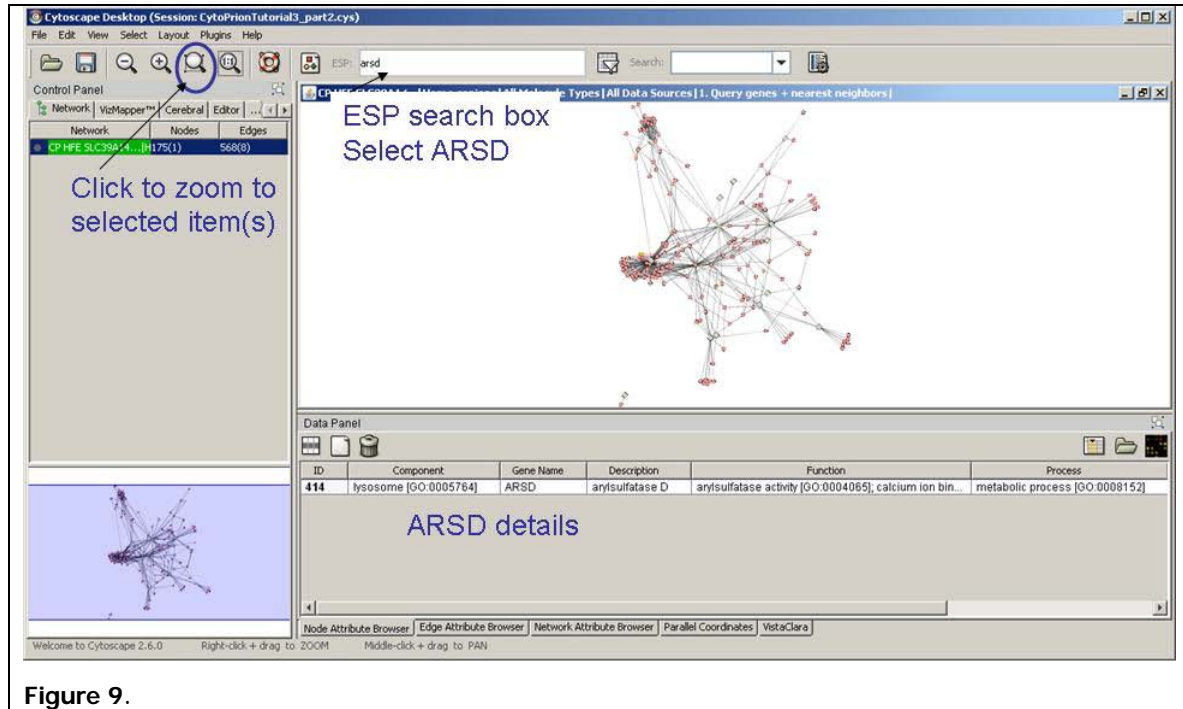

Figure 9.

2. Under the menu bar is the toolbar with shortcut icons. Click the blank magnifying glass (circled in Figure 9). This magnifying glass lets you zoom directly into the item you selected – ARSD. See Figure 10. The zoom is so close that you lose your context in the large network and cannot find the translucent blue box in the context view in the left hand Network tab.
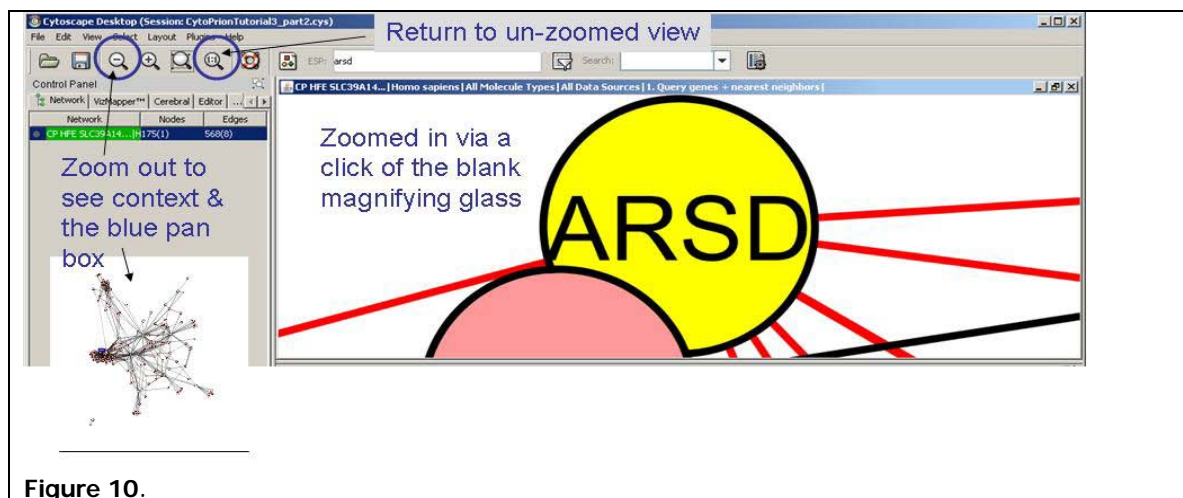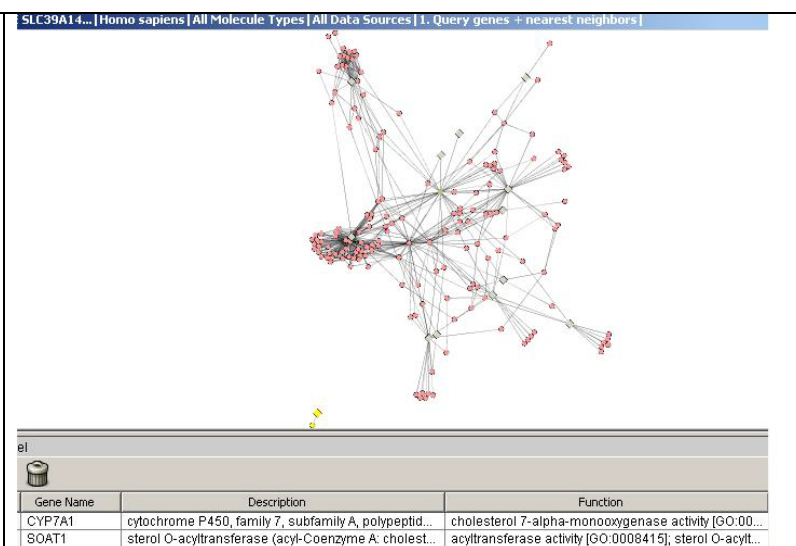

Figure 10.

3. To re-orient, zoom out a little. Click the magnifying glass with the minus (-) repeatedly. You now zoom out enough to see context and to have the blue translucent box reappear.

4. Move the blue translucent box to pan the graphic and see what gene products interact. Any time you want to see data on a given node, click it: The node turns yellow and information about it appears in the Data Panel. Shift+click items to see information about all of them at once.

5. You see that the thicker strip of genes (pointed to with an arrow in Figure 11) seems to relate roughly to lipid metabolism in some way. The lower genes (the zoomed in view in Figure 11) pertain to iron.
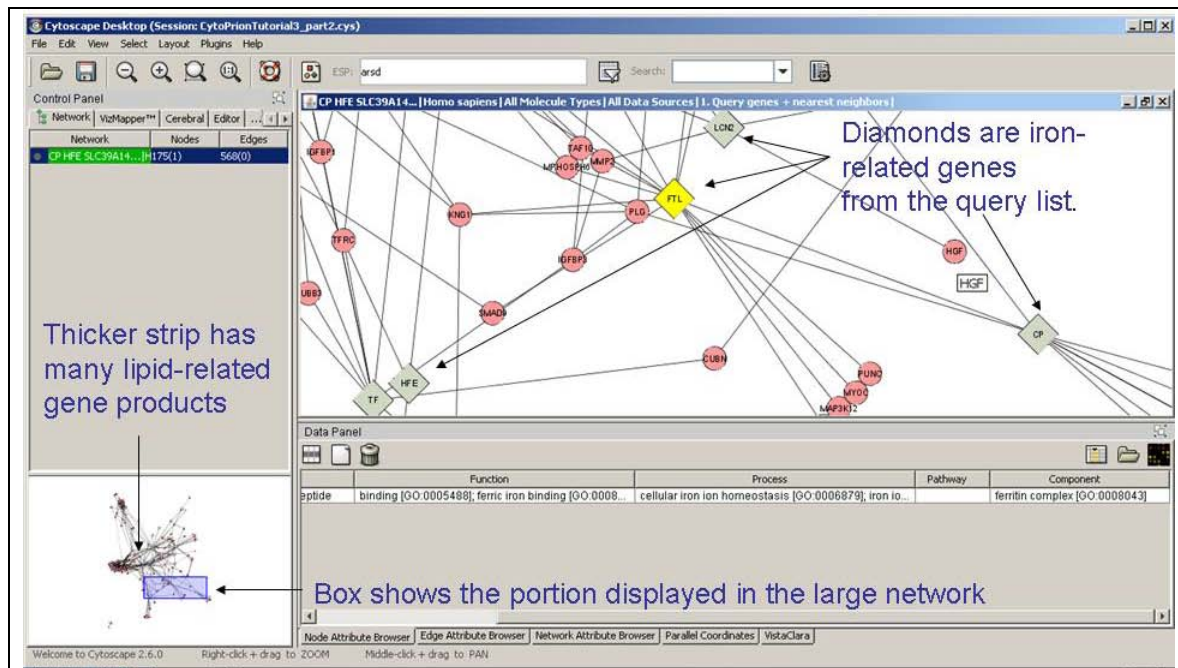
Figure 11.

6. Click the 1:1 (one-to-one) magnifying glass under the menu (circled in Figure 10) to return to the full network view. The view in Figure 12 appears.

You see two gene products disconnected from the network (bottom middle of the view). All other gene products are connected, suggesting the possibility of some link between iron- and sphingolipid genes.

Figure 12. 2 lipid/cholesterol gene products are disconnected.

Steps

Looking for possible links between iron- and sphingolipid related genes, you now explore clusters. To spot the gene products easily from your original list and their GO associations (iron or sphingolipid), you will first color code the nodes by these GO distinctions.

1. Color-code the two types of genes with the following steps.

      A. Bring in the Excel spreadsheet values for the genes –
"CytoiHemeAndSphingNoHdgReadInList.xls"
- On the top menu, click **File → Import → Attribute from Table (Text/MS Excel)**.
- In the resulting dialogue box follow the steps in Figures 13a-d.

Figure 13a. Step 1 of setting up Import: Keep the default attribute (radio button) on Node; select the xls file to import.
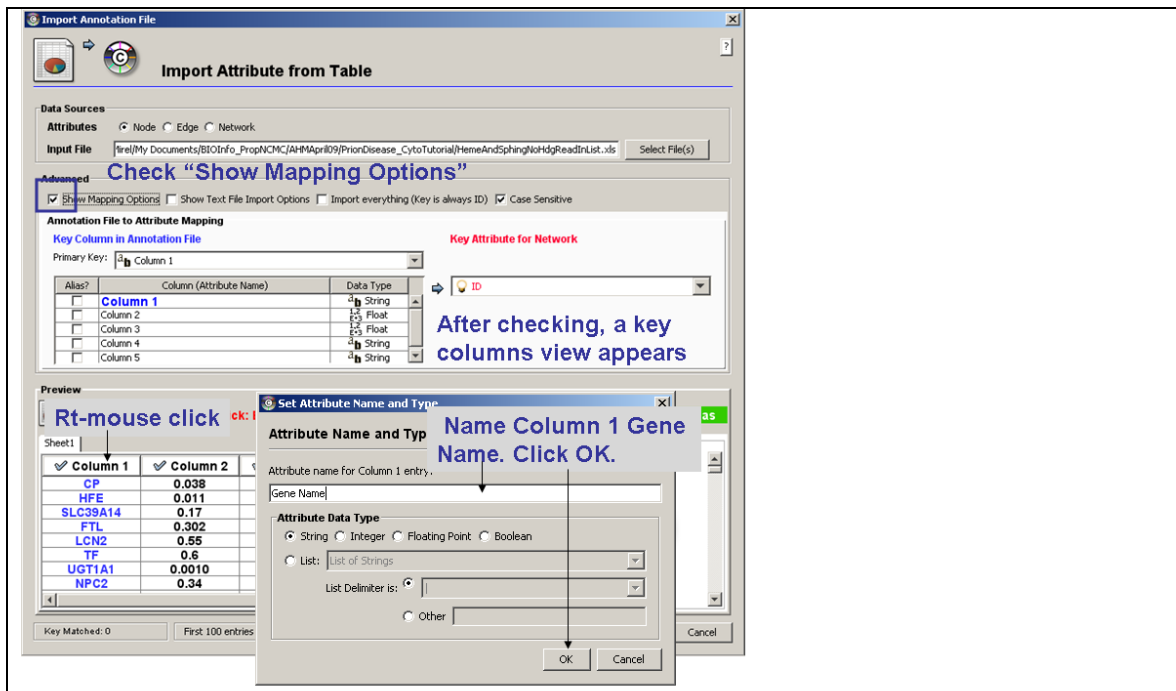
Figure 13b. Set up the data fields for a mapping of your file to the data already in Cytoscape.
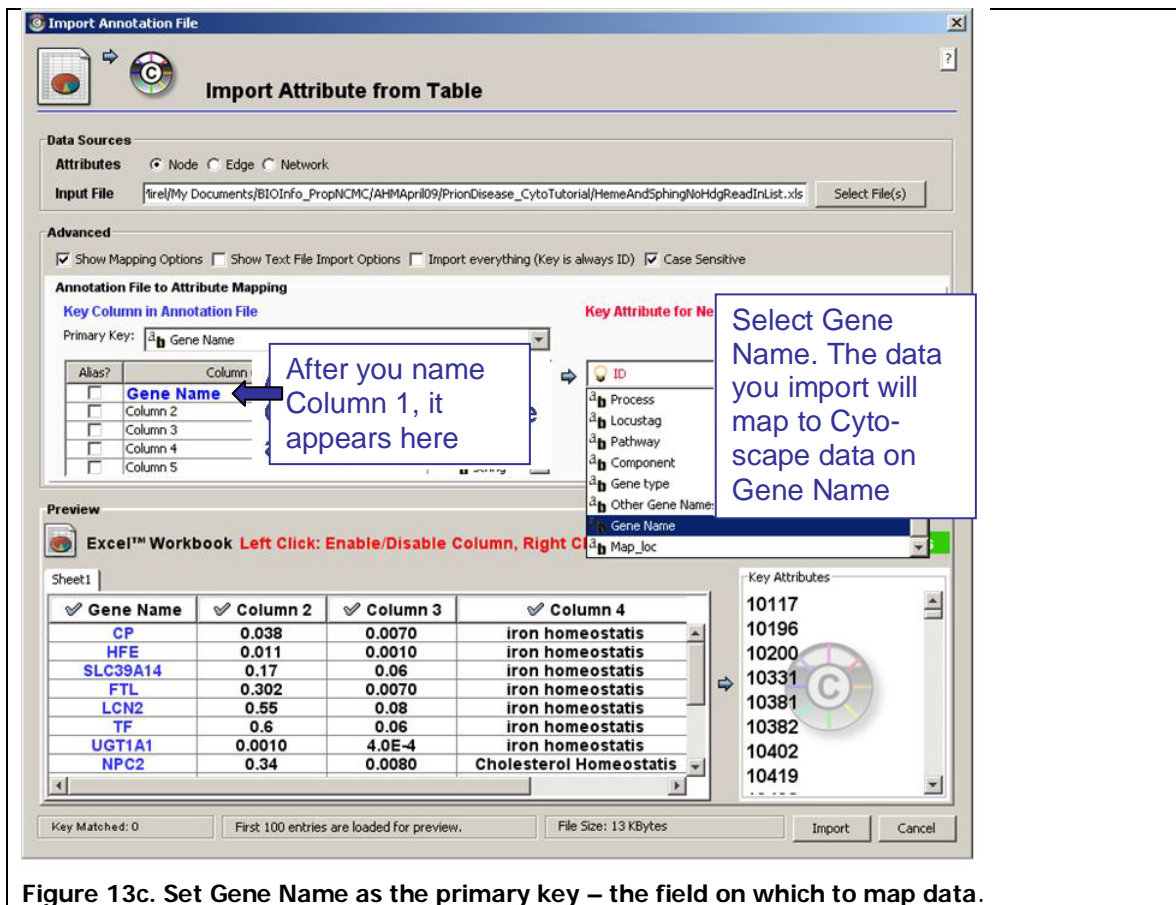

Figure 13c. Set Gene Name as the primary key – the field on which to map data.
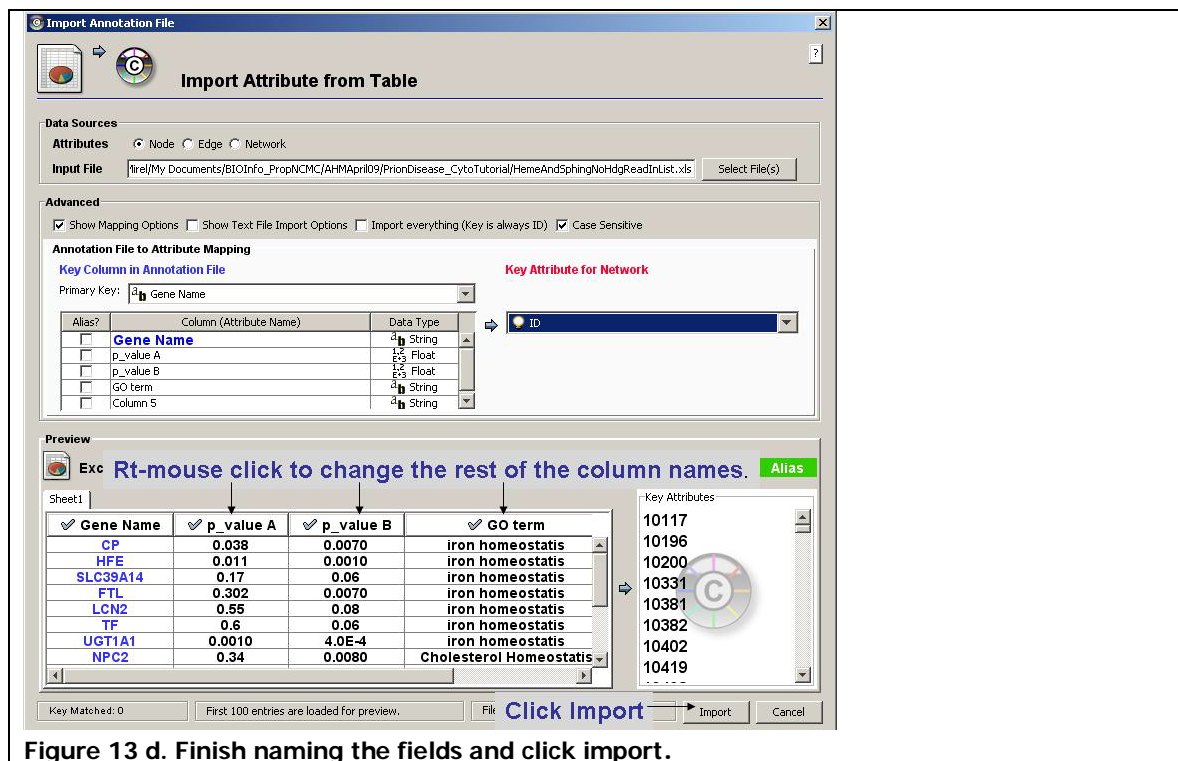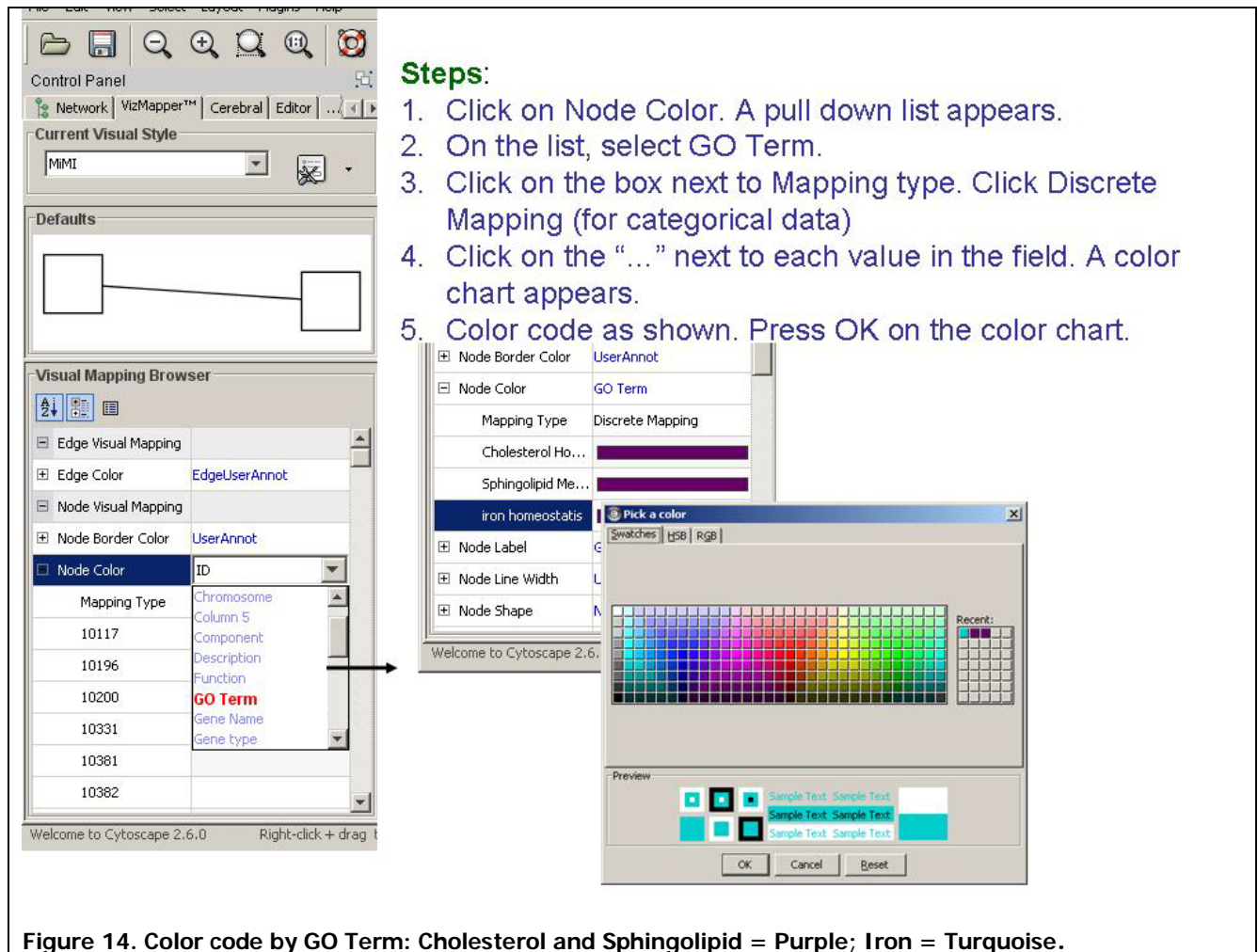
Figure 13 d. Finish naming the fields and click import.

B. Color Code by the field named GO Term, which simply differentiates between iron- and sphingolipid/cholesterol related genes.
- Click the VizMapper tab in the Control Panel.
- Follow the steps detailed in Figure 14.

Figure 14. Color code by GO Term: Cholesterol and Sphingolipid = Purple; Iron = Turquoise.

Finding Networks of Interest:

2. From the initial display it is clear that there are genes from both of the major metabolic areas: Cholesterol / Sphingolipid metabolism and Iron homeostasis, however none of them is directly linked. There are some highly clustered regions around specific pathways, but in addition there are some other loose clusters that bear examination.

3. One method for looking for pathways is to 'browse' genes and highlight small groups of them and then use the SAGA tool to find approximate subgraph matches to the input group in the KEGG pathway database.

a. Use Shift Click to select the following genes which are linked to both EDG3 and ABCA1.
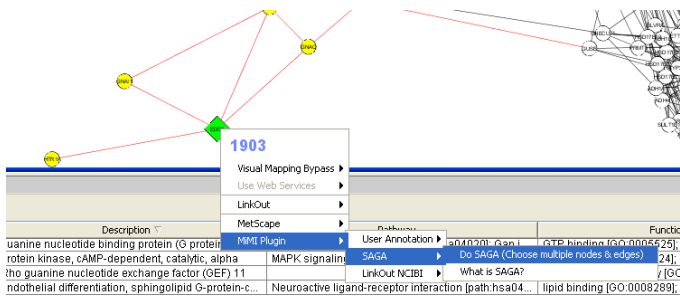>   HTR1A
>   GNA13
>   ARHGEF12
>   GNAI1
>   GNAQ
>   PRKACA
>   ARHGEF11

EDG3
ABCA1

b. Select all the adjacent edges to these nodes (Click **Select → Edges → Select Adjacent Edges**)
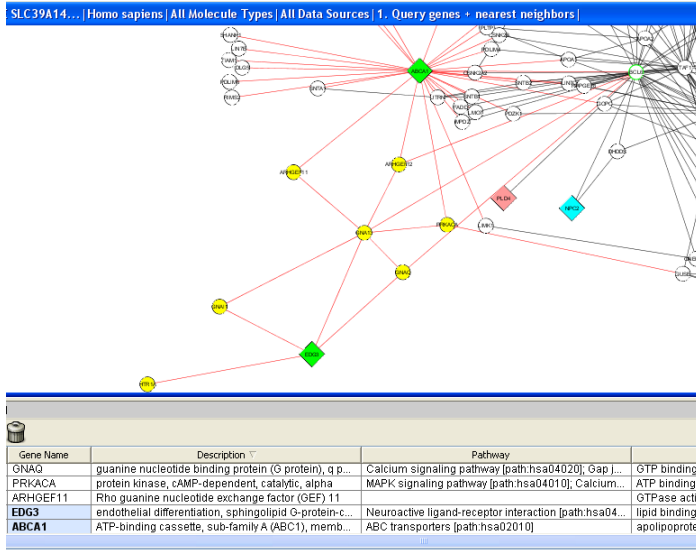
c. Right click on EDG3 and launch SAGA (**Right Click EDG3 → MiMI Plugin → SAGA → Do SAGA**)

d. In the popup window – leave the default settings and click "Query"

e. The SAGA results include a table of the possible matching pathways, and the score and number of nodes from your query pathway against the KEGG database.

The Query Result:

Result for the Query: MiMIPlugin2SAGA

Total Execution Time (s): 0.039039

Number of Matches: 6 (excluding self matches)

Matches Overview:

Match No.	Match Graph Name (#Nodes, #Edges)	Graph Distance	Matching Nodes
Match #1	path:hsa04540 (48,56) [Gap junction]	24.00	2 out of 8
Match #2	path:hsa04730 (35,39) [Long-term depression]	24.00	2 out of 8
Match #3	path:hsa04730 (35,39) [Long-term depression]	24.00	2 out of 8
Match #4	path:hsa04916 (35,31) [Melanogenesis]	25.00	2 out of 8
Match #5	path:hsa04720 (26,38) [Long-term potentiation]	26.00	2 out of 8
Match #6	path:hsa04912 (36,38) [GnRH signaling pathway]	26.00	2 out of 8

f. Click on Match #2 to explore the possible matches to the genes you queried.
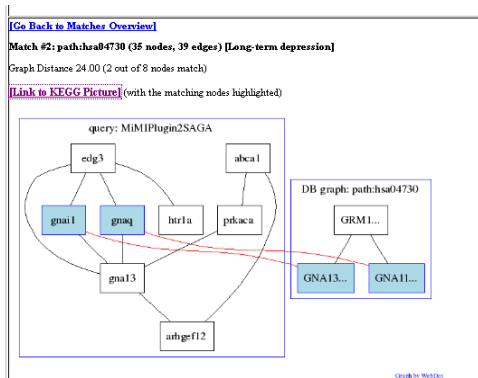
g. The cartoon representation of the network shows where your query matched KEGG. To visualize the KEGG pathway, click on Link to KEGG Picture.
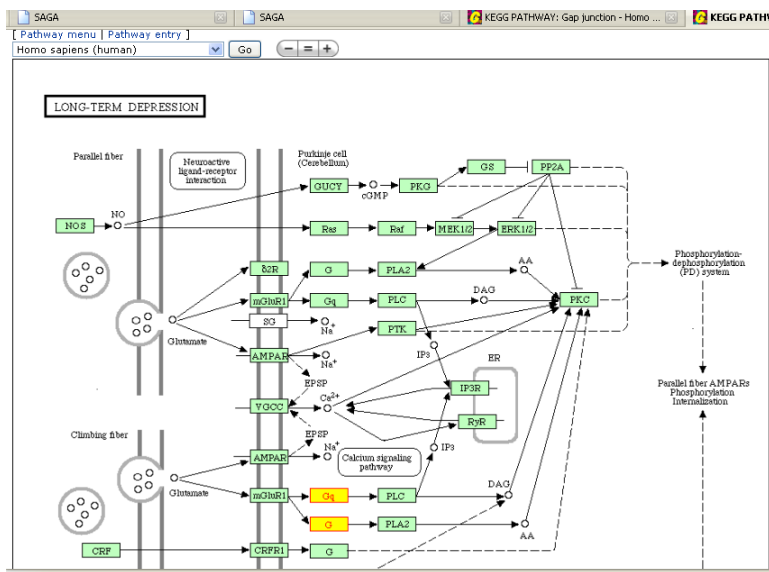

It is interesting and perhaps useful to discover that the short number of genes involved between EDG3 (sphingolipids) and ABCA1 (cholesterol synthesis) are also related to genes involved in Long-Term depression / potentiation of neuronal signaling.

4. Using a Network Based Cluster analysis tool: MCODE

a. Now examine clusters across the whole network without bias. Click the menu item **Plugins →** **MCODE → Start MCODE**. The MCODE panel appears on the left Cytoscape Control panel. MCODE algorithms analyze your network and pull out structural patterns (clusters) that stand out. These are network structural clusters, not necessarily biological ones. They often, however, represent shared pathways among the members of the sub-network.

b. On the MCODE panel, leave the defaults intact and click Analyze. Processing may take a moment. The MCODE Results panel appears on the right. It presents the most salient (high scoring) cluster first and the rest in descending order (See Figure 15).
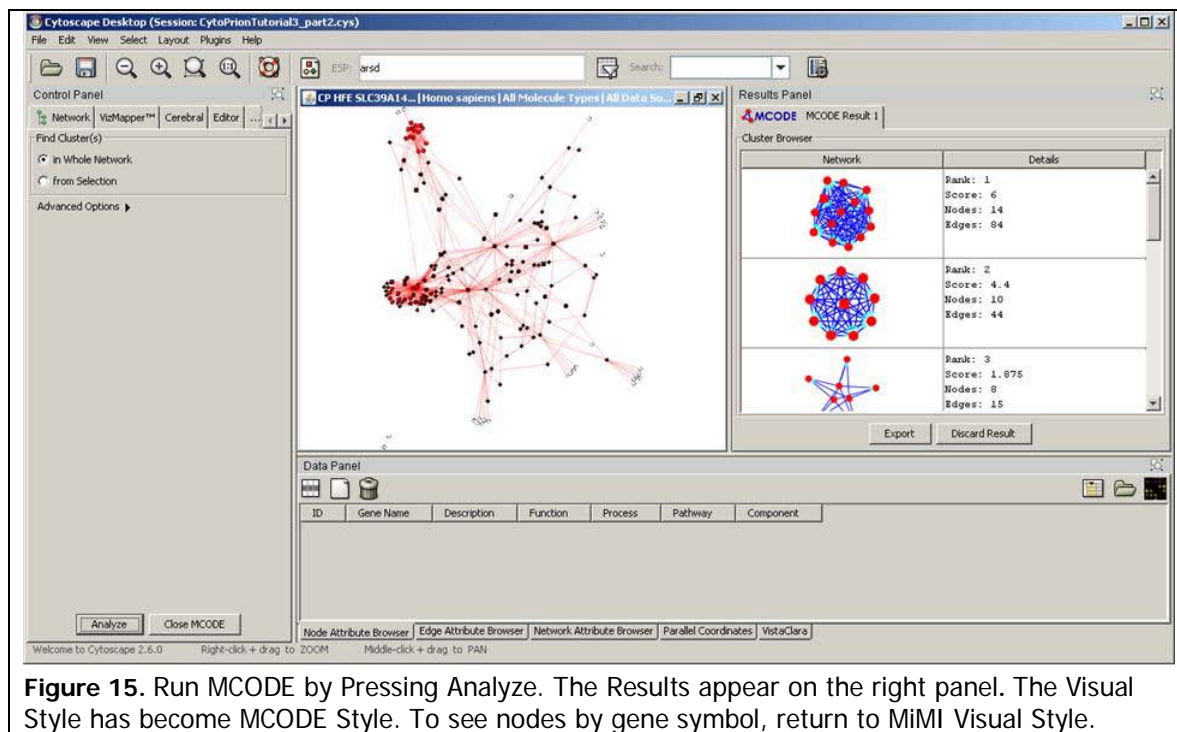
Figure 15. Run MCODE by Pressing Analyze. The Results appear on the right panel. The Visual Style has become MCODE Style. To see nodes by gene symbol, return to MiMI Visual Style.

c. Set up the graphics using the following steps as indicated in Figure 16:

Resize the Results panel and network window.

To see Results and the network better, click the Float Window (docking) icon el. The icon is circled in Figure 16. To re-dock, click the icon again.

Running MCODE changes out of the MiMI Visual Style. Return to the MiMI Style by clicking the VizMapper tab in the left Control Panel. In VizMapper under Current Visual Style, pull down the list and click MiMI. MiMI style returns, e.g. gene symbols on the nodes and the color and shape coding that you are used to.

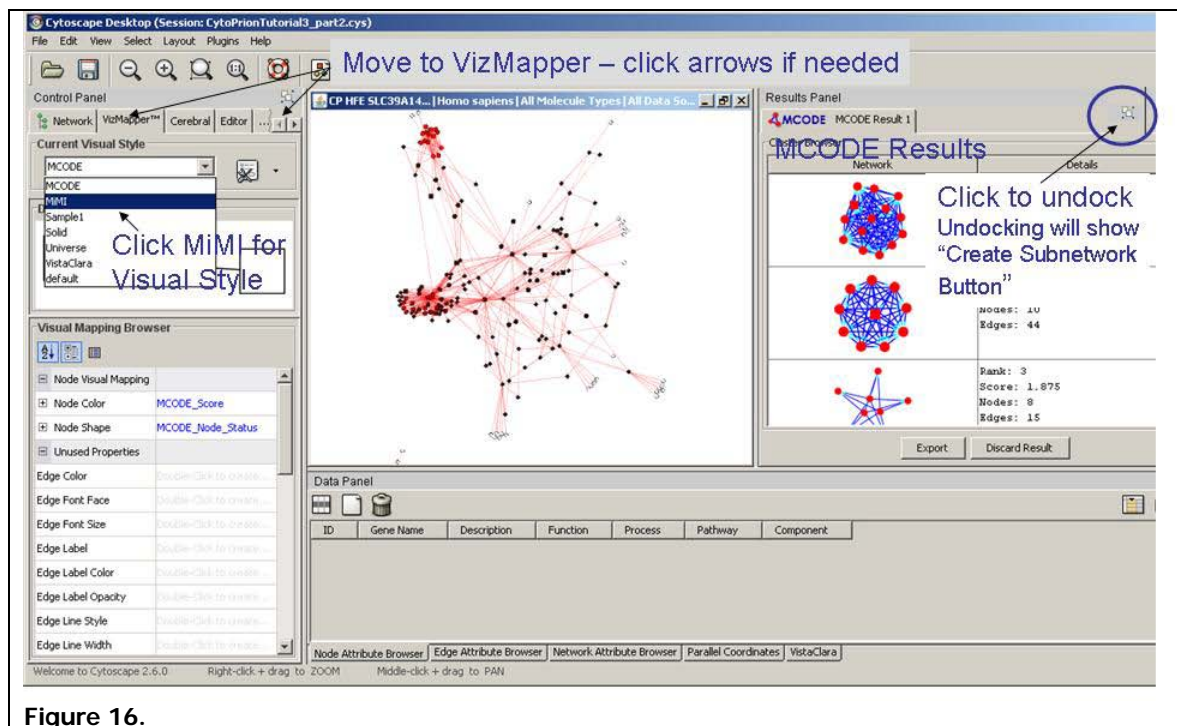Return to the Network tab from VizMapper in the left hand Control Panel.

Figure 16.

d. Examine each of the clusters extracted by MCODE in turn.

For each cluster, examine the details of its gene products displayed in the Data Panel when you select the gene products. Try to gain insight into genes that may link iron and sphingolipid. Use these steps for each successive cluster:

1. Click on the first cluster image on the Results panel. The Data Panel displays the details of only these gene products.

2. Click **Create Subnetwork** on the MCODE panel to see the network view on the large graph of just these molecules. The visual style changes.

3. Consequently, after creating each subnetwork, you will have to reset the view to the MiMI Visual Style as detailed in Step 4 above and as shown in Figure 16.

4. When none of the nodes is selected, you can see the color-coding to distinguish iron- and sphingolipid -related gene products.

5. Select all the nodes and edges in the subgraph by drawing a box around them or pressing Control+ Alt+A. The yellow highlighting replaces your color coding.

6. View the interactions and study the nodes' Function, Pathway, Component and Process details. You may see a biological story beginning to take shape as you move through the different clusters.

When finished with one cluster, move to the next and repeat these steps. Each cluster is automatically saved in this Cytoscape session. You can find it listed in the Network tab view on the left Control Panel (See Figure 17).
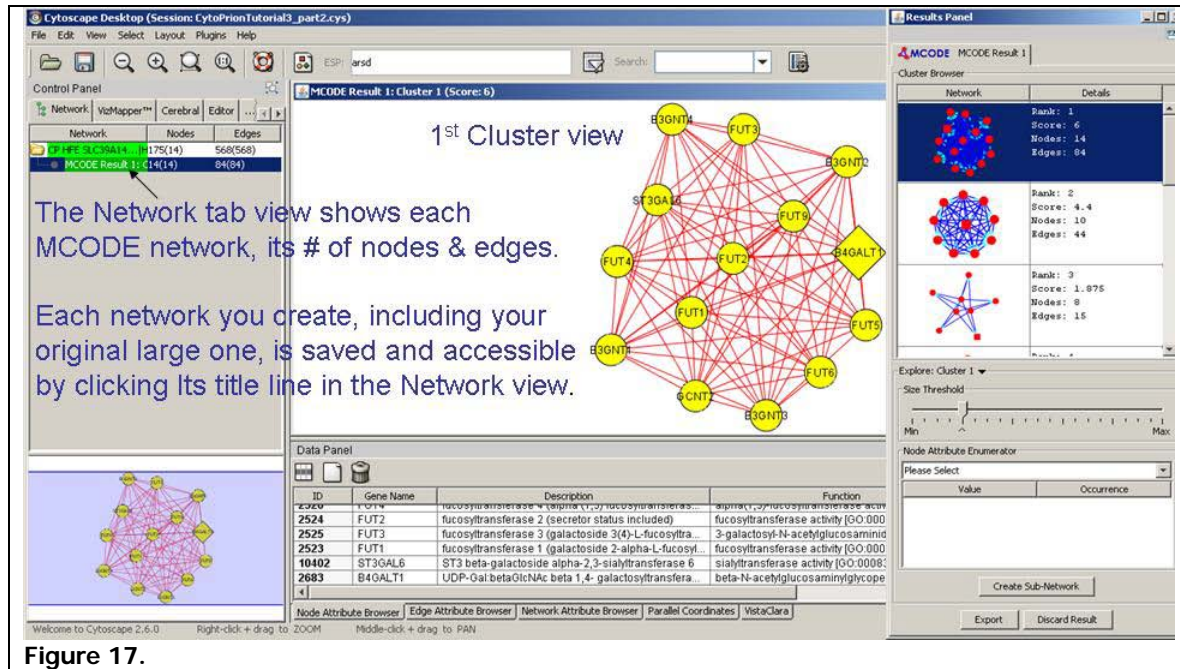

Figure 17.

Some insights from this exploration of clusters are highlighted in Figure 18:
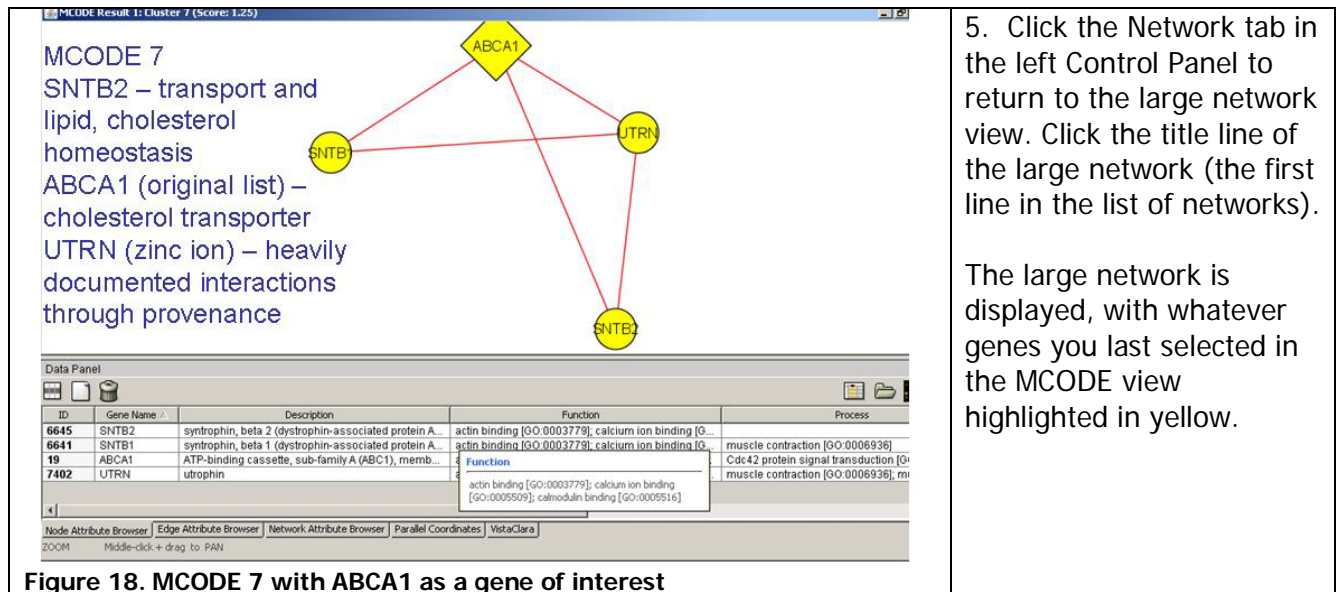

**Figure 18. MCODE 7 with ABCA1 as a gene of interest**

5. Click the Network tab in the left Control Panel to return to the large network view. Click the title line of the large network (the first line in the list of networks).

The large network is displayed, with whatever genes you last selected in the MCODE view highlighted in yellow.

Based on interesting interactions you uncovered from the cluster exploration, see if you can find indirect links between iron metabolism and sphingolipids or cholesterol.

e.   Use the ESP search box to find and select two genes. Then use the "Shortest Path" plug in see how many hops separate the two genes and view the connectors of the hop.

f. In the ESP box: Type ABCA1 OR UGT1A1. These two nodes are selected and highlighted.

g. On the top menu, click: **Plugins → Shortest Path… → Shortest path undirected**.
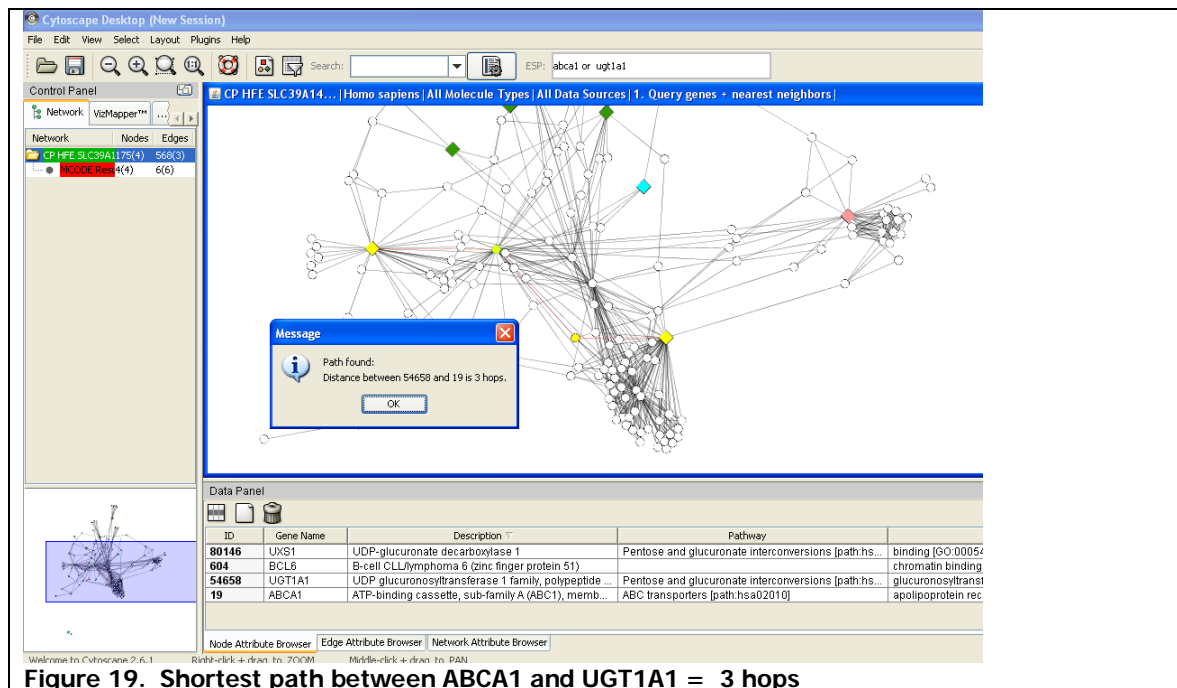
Figure 19. Shortest path between ABCA1 and UGT1A1 = 3 hops

Three hops in this network include ABCA1 → BCL6 → UXS1 → UGT1A1. This is a rapid way to identify the path through a complex network.

<u>SUMMARY</u>:

In this module we have shown you how to:

Part I
> ► Query on a list of genes
> ► Set up views to facilitate exploratory analysis, e.g. layouts, selection
> ► Explore details to conceptually characterize gene products and interactions
> ► Expand a network node (gene product) to see additional direct interactors
> ► Find literature associated with interactions to help verify the interactions

Part II
> ► Read in additional values related to the genes on your list
> ► Color code nodes by a selected attribute
> ► Link out to details, supporting literature, and subgraph matches with KEGG
> ► Find shortest paths between 2 genes and identify the mediating gene products
> ► Create network clusters and explore their traits