

CYTOSCAPEを使った データの可視化

統合データベース講習会: AJACSみちのく2

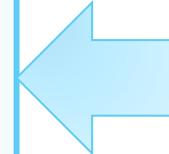
2012年8月23日(木)

(独)科学技術振興機構 バイオサイエンスデータベースセンター
櫛田達矢

ライフサイエンス大規模データの可視化

- ゲノムの位置情報(ゲノムブラウザ)
- ヒートマップ
- 発現部位表示
- 系統樹
- パスウェイ、ネットワーク
 - 代謝マップ
 - シグナル伝達マップ
 - 遺伝学的相互作用
 - タンパク質-タンパク質相互作用
 - 転写制御ネットワーク
 - ...
- ...

Cytoscapeが
取り扱う領域



Cytoscapeとは？

- Cytoscape: An Open Source Platform for Complex Network Analysis and Visualization
- 開発者
 - http://www.cytoscape.org/development_team.html
- マニュアル
 - http://cytoscape.org/manual/Cytoscape2_8Manual.html
- 最新版(2012年7月10日現在)
 - 2.83
 - <http://www.cytoscape.org/download.html>

Cytoscapeの特徴と機能

- 様々な標準化データ(フォーマット)に対応
- ウェブサービスへの技術提供
- 独自のファイル形式
- データの相互運用
- 柔軟なデータ可視化機能(VizMapper™)
- 画像データ出力
- 豊富なグラフの自動レイアウト
- パスウェイ検索機能
- ブラウジング機能
- フィルタリング機能
- 部分パスウェイ、モジュール構造の発見
- プラグインによる機能追加(データ分析機能など)
- 多言語対応

様々な標準化データ(フォーマット)に対応

- SIF, XGMML, GML, SBML, PSI-MI, BioPAX, Excel, OBO, etc.

グラフ表記のフォーマット

Systems Biology Markup Language

Biological Pathway Exchange

Proteomics Standard initiative
Molecular Interaction



各種データの再利用を容易にする

Open Biological
Ontology

ウェブサービスへの技術提供



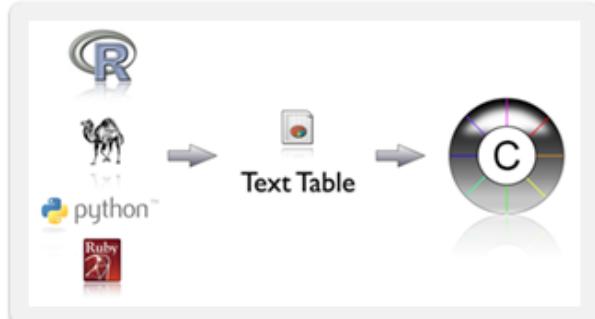
独自のファイル形式

グラフ(パスウェイ、ネットワーク)のノード、エッジの属性、画面サイズ、解析結果を一括保存

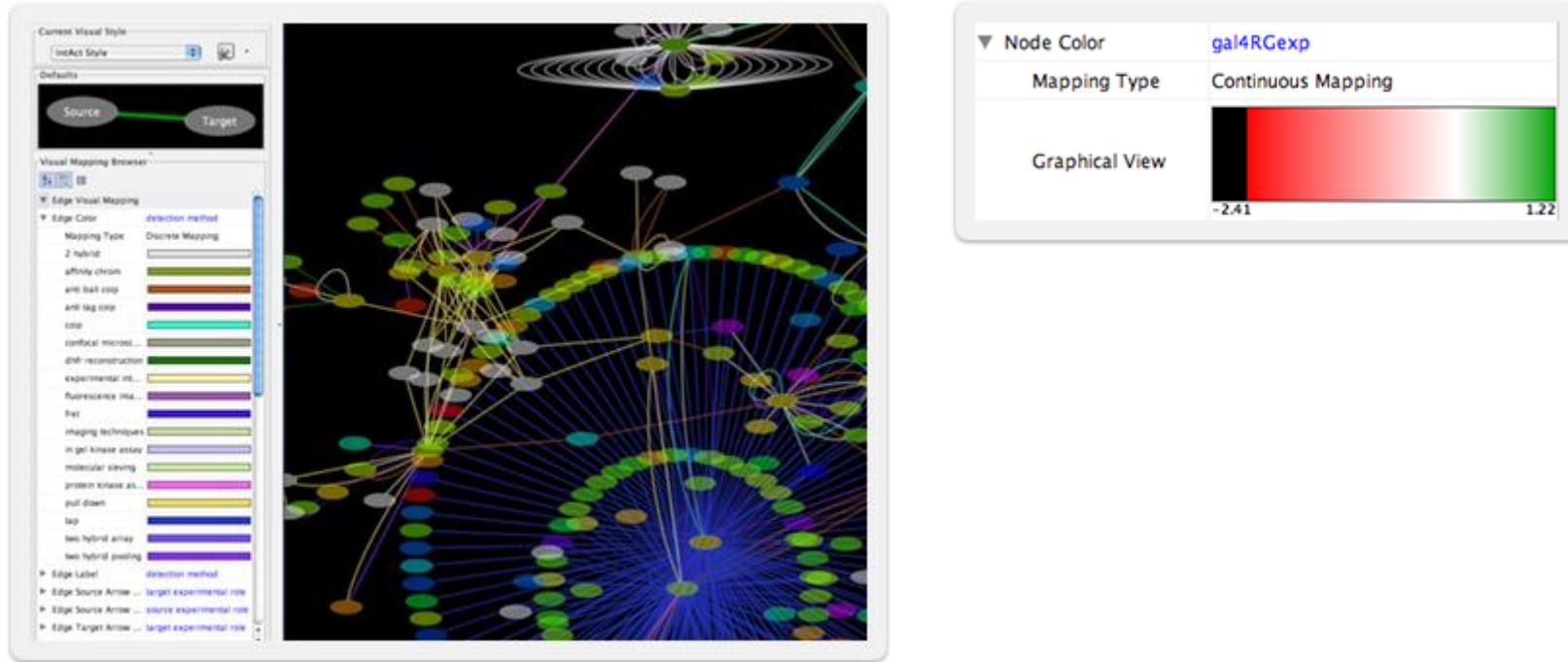


データの相互運用

- 使用例(Rのigraphパッケージを利用した複雑ネットワーク解析の紹介)
 - <http://cytoscape.seesaa.net/article/47154734.html>



柔軟なデータ可視化機能(VizMapper™)

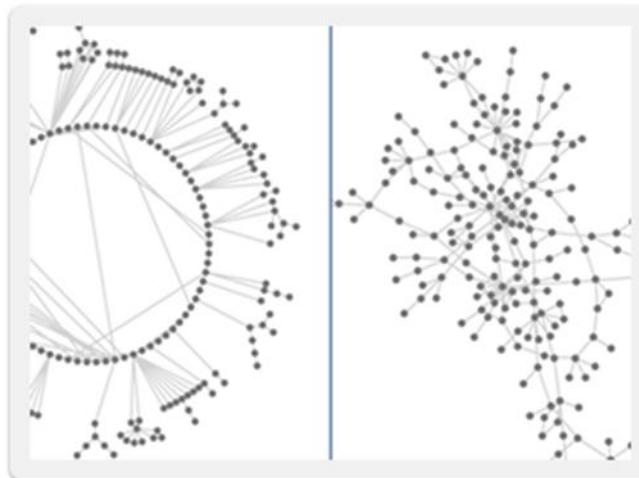


- Visual Style: 名前、タイプ、度数、頻度、発現量などの属性
データを、ノードやエッジの色、大きさ、形、フォントタイプで表現。
- VizMapper™はそのインターフェイス。

画像データ出力

- PDF, EPS, SVG, PNG, JPEG, BMP の各種画像フォーマットで出力可能

豊富なグラフの自動レイアウト



Circular

Organic

- Cytoscapeオリジナル、yfilesなどのレイアウトを実装

パスウェイ検索機能

The screenshot shows a search interface for a pathway database. At the top, there is a search bar with the query "cell wall (sensu the fungi re...)" and an "ESP:" dropdown menu. Below the search bar is a list of search results:

- carbamoyl-phosphate synt... 2 hits
- ccaaat-binding factor complex 3 hits
- cellular_component 9 hits
- cell wall (sensu the fungi re... 2 hits
- central plaque of spindle pol... 1 hit

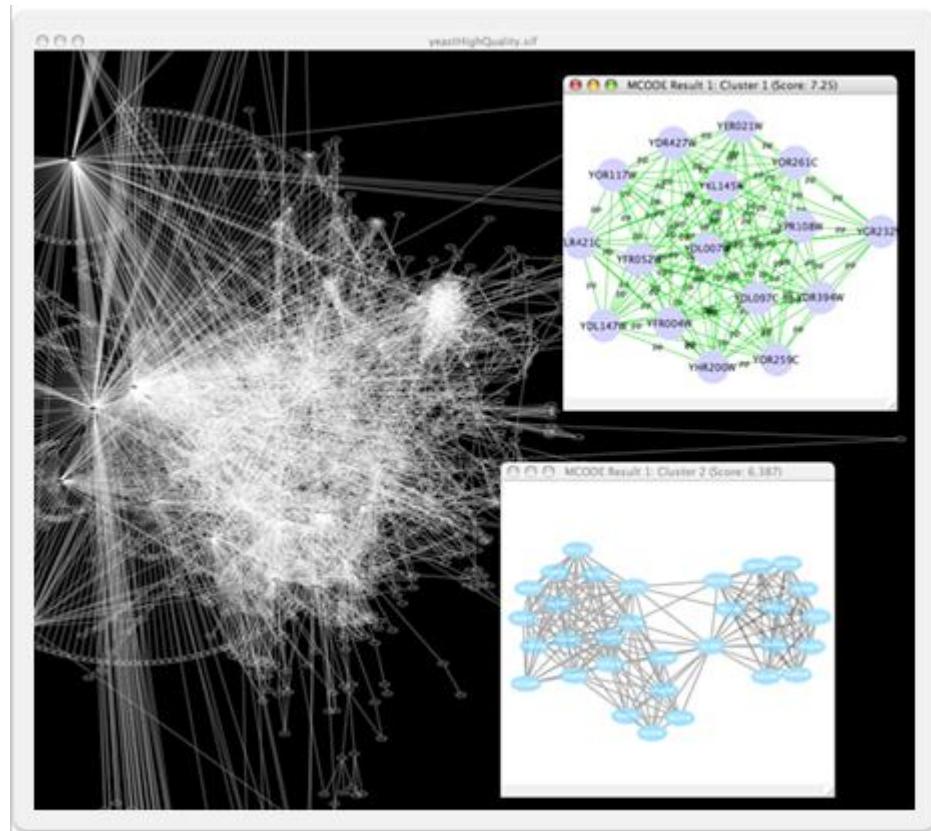
Below the results, there is a search bar containing the query "(KEGG AND mapk*) AND nucleus". The interface includes a "Data Panel" tab and several icons at the bottom left.

ID	annotation.GO_CELLULAR_COMPONENT	Pathway
YHR030C	[cellular bud tip, cytoplasm, nucleus]	[KEGG pathway: MAPK signaling pathway]
YHR084W	[nucleus]	[KEGG pathway: MAPK signaling pathway]
YPL089C	[nucleus]	[KEGG pathway: MAPK signaling pathway]
YMR043W	[nuclear chromatin, nucleus]	[KEGG pathway: MAPK signaling pathway]
YDR103W	[cytoplasm, mating projection tip, ...]	[KEGG pathway: MAPK signaling pathway]
YJL157C	[cytoplasm, mating projection tip, ...]	[KEGG pathway: Cell cycle - yeast, KEGG pathw...]
YER111C	[nucleus]	[KEGG pathway: Cell cycle - yeast, KEGG pathw...]

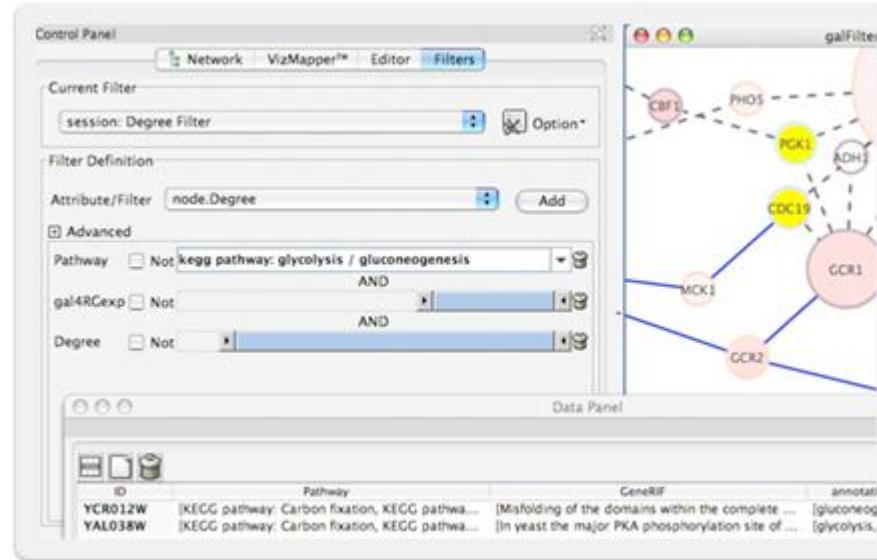
- ノードやエッジ(の属性)に対するキーワード検索を実装
- And/or検索、前方一致、後方一致などにも対応

ブラウジング機能

- ・パスウェイ上の任意の箇所のズームイン/アウト、ピックアップ。
- ・パスウェイの統合。
- ・100,000以上のノードとエッジからなるパスウェイに対するスムーズなナビゲート。

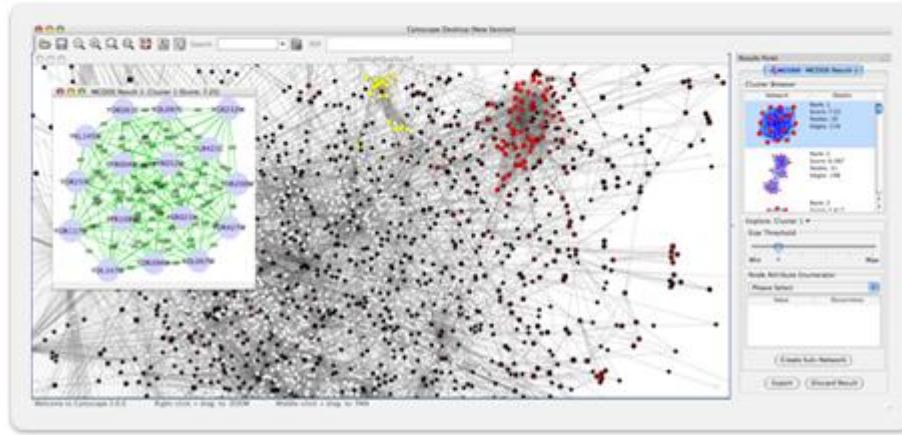
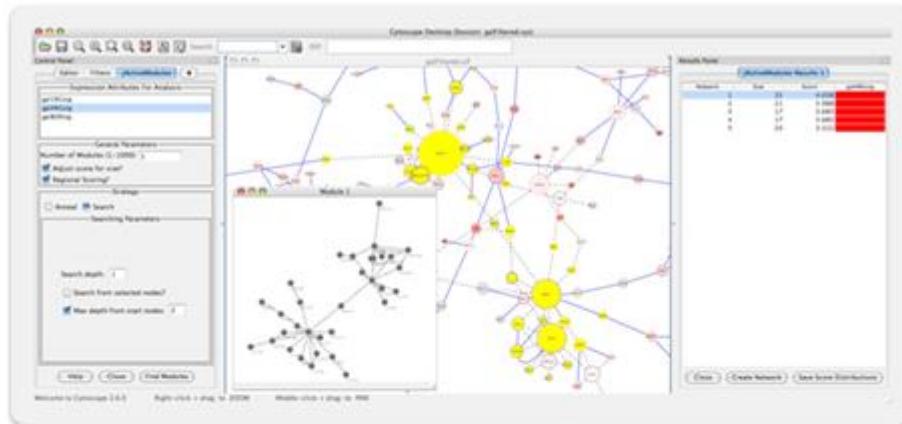


フィルタリング機能



- ノードやエッジの属性情報に対して、データの閾値(発現量、p値など)に基づくノードやエッジの抜出し(新規ネットワークの作成)が可能

部分パスウェイ、モジュール構造の発見

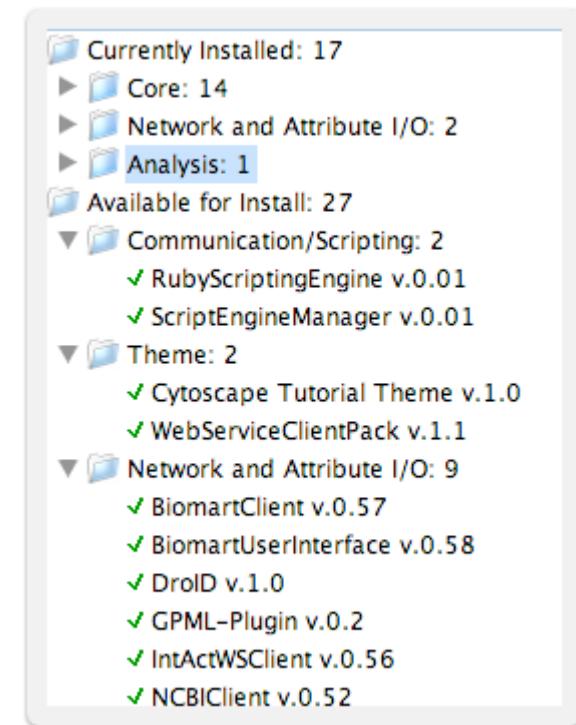
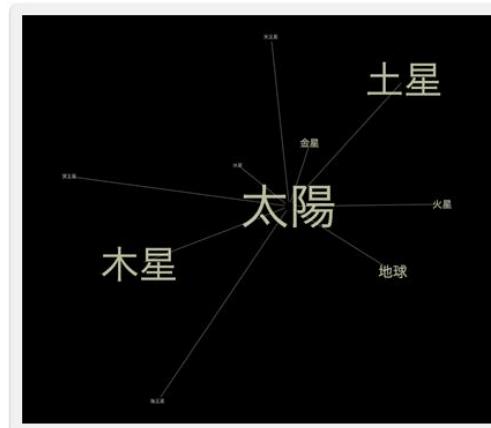


- (特定のプラグインを用いることで、)遺伝子ネットワーク内で特徴的に発現しているパスウェイの部分構造(サブパスウェイ)や、PPIにおける複合体、およびProtein similarity networkにおけるプロテインファミリーのクラスター発見を可能にする。

プラグインによる機能追加(データ分析機能など)

- ・多数のデータ解析、インポート、可視化のプラグインが利用可能。
- ・プラグインマネージャーにより簡単に導入可能。
- ・最新の解析アルゴリズムがプラグインとして活用できることも！

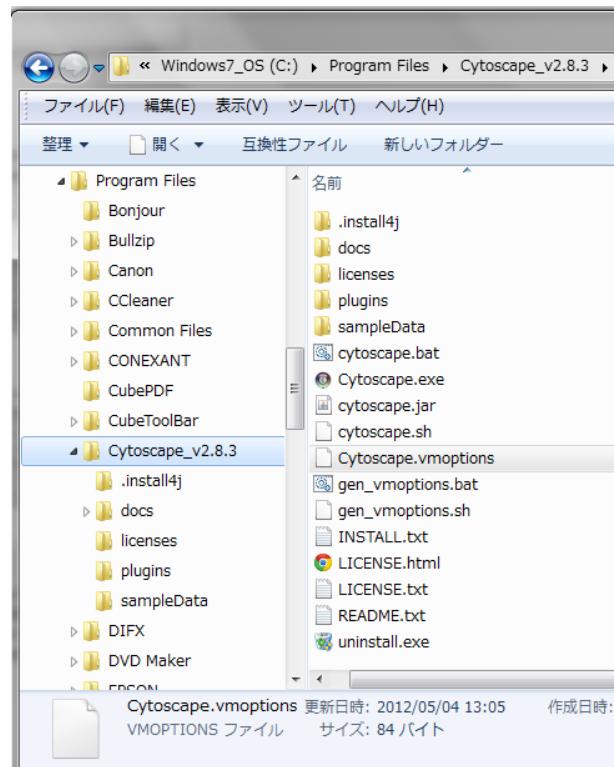
多言語対応



基本操作

使用メモリー量の設定 1 of 2

- 取り扱うネットワークの大きさ(ノード数+エッジ数)によってメモリーの設定を調整したほうがよい。
- ファイルCytoscape.vmoptions(例、C:¥Program Files¥Cytoscape_v2.8.3にある)をテキストエディタで開き、「Xmx***」を「Xmx1G」に修正。



→

```

-Xms10m
-Xmx1G
-Xss10m
-Dswing.aatext=true
-Dawt.useSystemAAFontSettings=lcd

```

実習1. Cytoscape.vmoptionsの中身を確認してみましょう。

使用メモリー量の設定 2 of 2

ネットワークの大きさと推奨されるメモリーサイズ(Xmx)の目安

オブジェクト数 (ノード数+エッジ数)	推奨される メモリーサイズ(Xmx)
0 - 20,000	512M (default)
20,000 - 70,000	800M
70,000 - 150,000	1G

レイアウト機能を使った場合に「メモリーエラー」が起こる場合は、Xssを変更する(Xssはヒープサイズを指定する。例、「Xss***」→「Xss10m」)。

詳細な情報およびMacの場合の対応は、以下を参照

- http://cytoscape.org/manual/Cytoscape_2_8Manual.html#Getting Started
- http://wiki.cytoscape.org/How_to_increase_memory_for_Cytoscape#

起動

実習2. Cytoscape.exe(例、C:\ProgramFiles\Cytoscape_v2.8.3)を選択(ダブルクリック)して起動してみましょう。

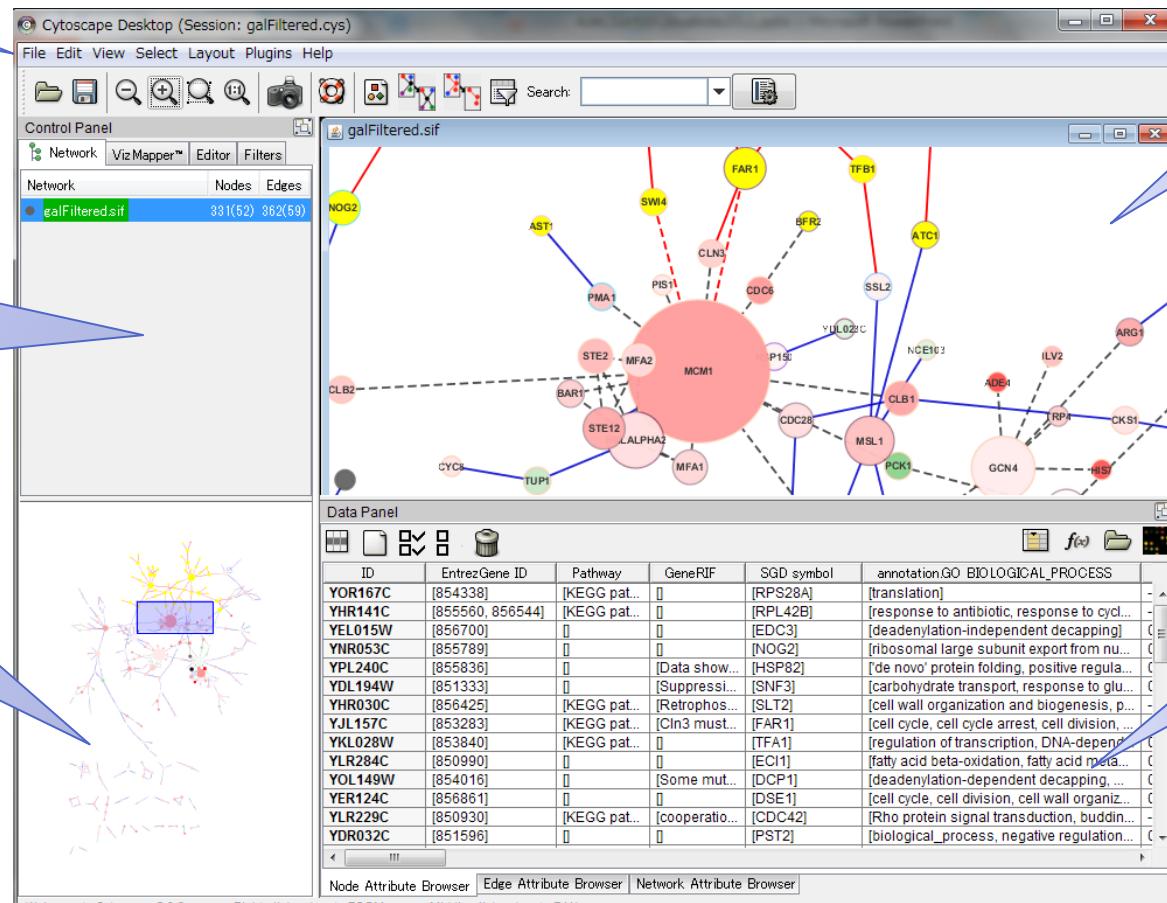
メニュー

メインネットワークビュー

コントロールパネル(ノードやエッジのグラフィック編集など)

ネットワークの全体表示

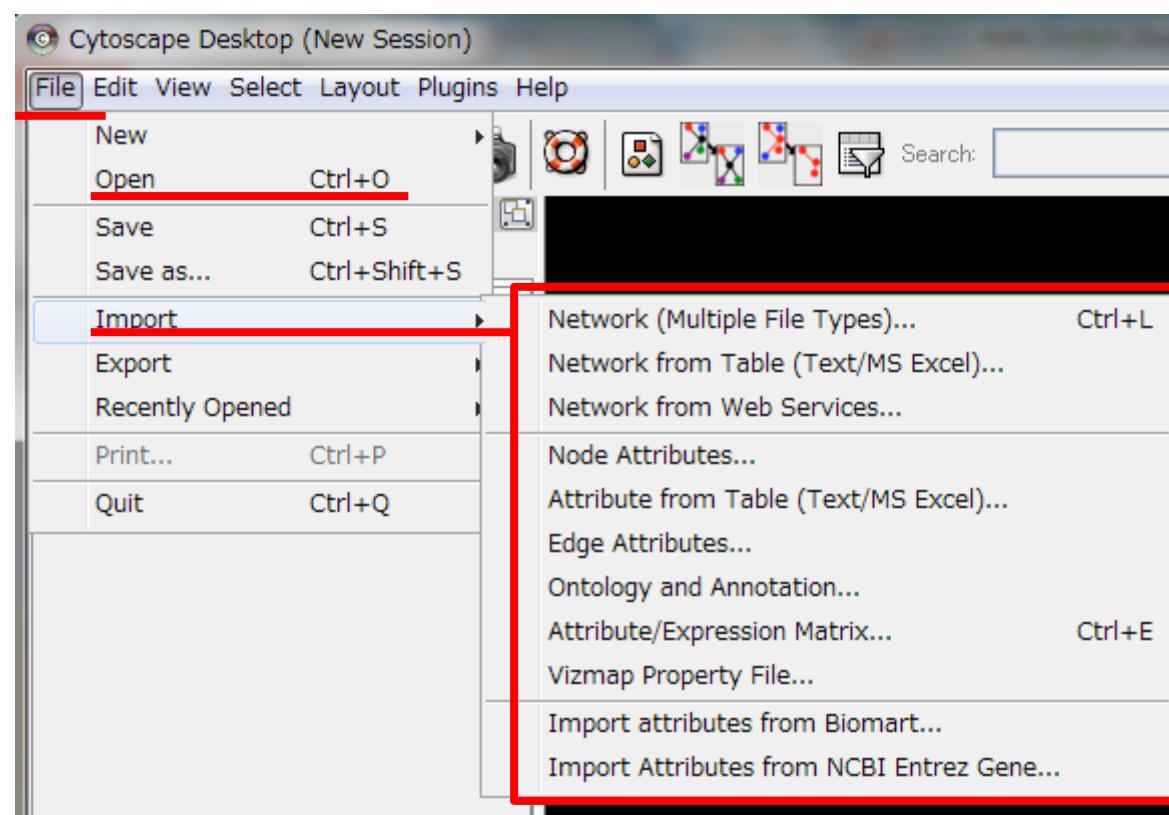
データパネル(属性値表示、編集)



* 図はファイルを開いた後の表示。

ファイルを開く

- .cysファイル
 - メニュー「File」の「Open」から
- .sif, .xgmml, .gmlファイル
 - メニュー「File」の「Import」、
「Network (Multiple File Types)」から
- .txt, .xlsファイル
 - メニュー「File」の「Import」、
「Network from Table
(Text/MS Excel)」から



実習3. Cytoscapeフォルダにあるサンプルデータのフォルダ(例、C:\Program Files\Cytoscape_v2.8.3\sampleData)の「galFiltered.csv」、「galFiltered.cys」、「galFiltered.sif」、「galFiltered.txt」、「galFiltered.xgmml」、「galFiltered.xls」をテキストエディタで開いて中身を確認してみましょう。

パスウェイの描き方

1. 既存のパスウェイデータを活用

- Pathway Commons (<http://www.pathwaycommons.org>)からデータをダウンロードする。
- Pathguide (<http://www.pathguide.org/>)で探す。
 - メモ: BioPAX, SBML(L2V1), PSI-MI(2.5.3)は可。
- WikiPathway (<http://www.wikipathways.org>)のgpmI形式のデータを利用する(プラグインが必要)。

2. テキストエディタやExcelを使ってパスウェイデータを作成する。

3. メインネットワークビューにお絵描きする。

Pathway Commonsとは？

- <http://www.pathwaycommons.org>
- 9種類のパスウェイ、相互作用DBのデータが利用可能。
- パスウェイ数: 1,668、生物種: 414

メニュー「Data Source」で、収録されているパスウェイを確認することができる。



The screenshot shows the Pathway Commons homepage. A red box highlights the 'Data Sources' link in the top navigation bar. To the left, a blue callout box contains the text: 'メニュー「Data Source」で、収録されているパスウェイを確認することができる。' (You can check the pathways listed in the 'Data Source' menu). The main content area includes a search bar, a 'What's New' section with a list of datasets added recently, and a grid of logos for various data sources.

Pathway Commons

Search and visualize public biological pathway information. Single point of access. [more...]

Home | Data Sources | Download | FAQ | Web Service | About |

Send us your feedback. Sign up for Pathway Commons announcements. RSS Feed

Using Pathway Commons:

Biologists: Browse and search pathways across multiple valuable public pathway databases.

Computational biologists: Download an integrated set of pathways in BioPAX format for global analysis.

Software developers: Build software on top of Pathway Commons using our [web service API](#). Download and install the [cPath software](#) to create a local mirror.

Current Data Sources:

Pathway Commons currently contains the following data sources ([batch download](#)):

What's New:

- **NEW!** Oct 27, 2011:
 - BioGRID data set (September 25, 2011 Version 3.1.81).
 - IntAct data set (September 29, 2011 Version 3.1.17288).
 - Nature Pathway Interaction data set (October 12, 2011).
 - Reactome data set (September 20, 2011 Version 38).
- June 24, 2011:
 - BioGRID data set (May 1, 2011 Version 3.1.76).
 - HumanCyc data set (June 8, 2011 Version 15.1).
 - Nature Pathway Interaction data set (June 14, 2011).

Pathway CommonsからCytoscapeへ

実習4. メニュー「Data Source」で、「Reactome」の「[Browse](#)」を選択。パスウェイのリストから、「[Pathway: The role of Nef in HIV-1 replication and disease pathogenesis](#)」を指定し、画面左のCytoscape:「[View in Cytoscape](#)」をクリック。次に、Download:「BioPAX」をクリック。

Cytoscapeで表示へ

Pathway: The role of Nef in HIV-1 replication and disease pathogenesis

Authored: Gillespie, ME, 2007-07-25 19:42:36 more...

[Sub-Pathways \(2\)](#) [Molecules \(49\)](#)

Showing 1-2 of 2

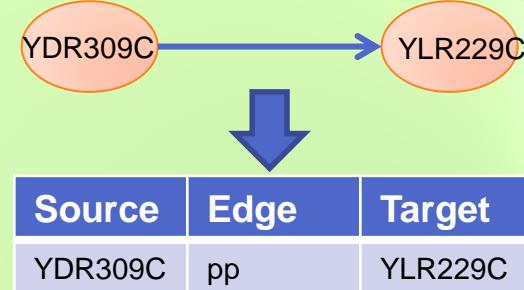
- [Nef mediates down modulation of cell surface receptors by recruiting them to clathrin adapters](#) [Reactome](#) [\[+\]](#)
- [Nef and signal transduction](#) [Reactome](#) [\[+\]](#)

Cytoscapeの「File」の「Import」、「Network (Multiple File Types)」でファイル選択。表示へ

テキストエディタ、Excelを使ってパスウェイデータを作成する

- ステップ1

- ノードとエッジのつながりを三項関係で記述する。
- エッジの属性値を記述する。
 - 例、エッジの種類(例、pp, pd, phosphorylate)、PubmedID
- 例、galFiltered.csv



- ステップ2

- 別ファイルに、ノードの属性値を記述する。
 - 例、Symbol名, GeneID, 実験データ(例、発現値、統計値)
- 例、galExpData.csv

GeneID	Symbol	Expression
YDR309C	GIC2	0.427
YLR229C	CDC42	0.074

ステップ1：ノードとエッジの繋ぎりを読み込む

①メニュー「File」の「Import」、「Network from Table (Text/MS Excel)」

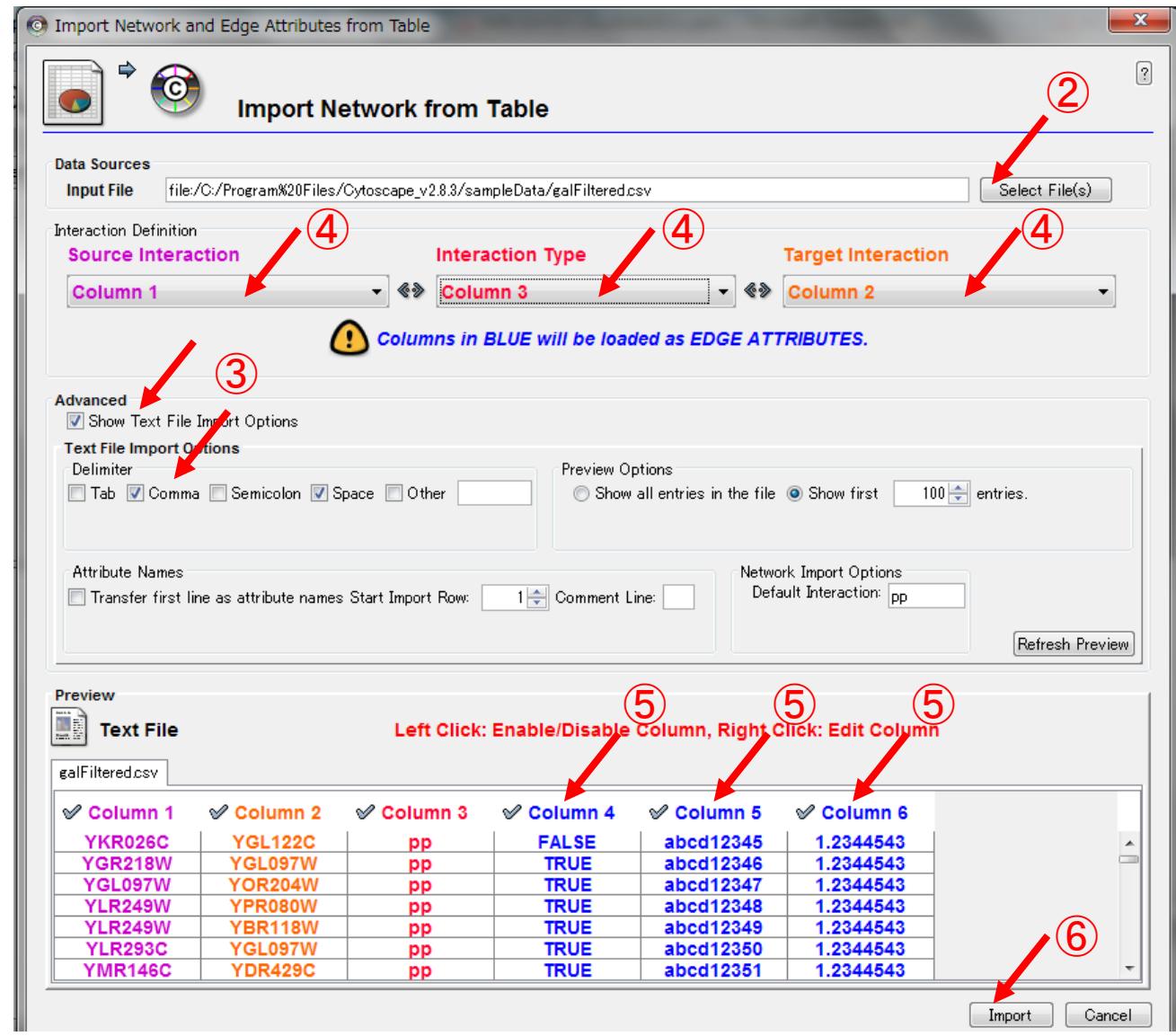
②ファイル「galFiltered.csv」を選択

③「Show Text File Import Options」に✓。「Comma」に✓

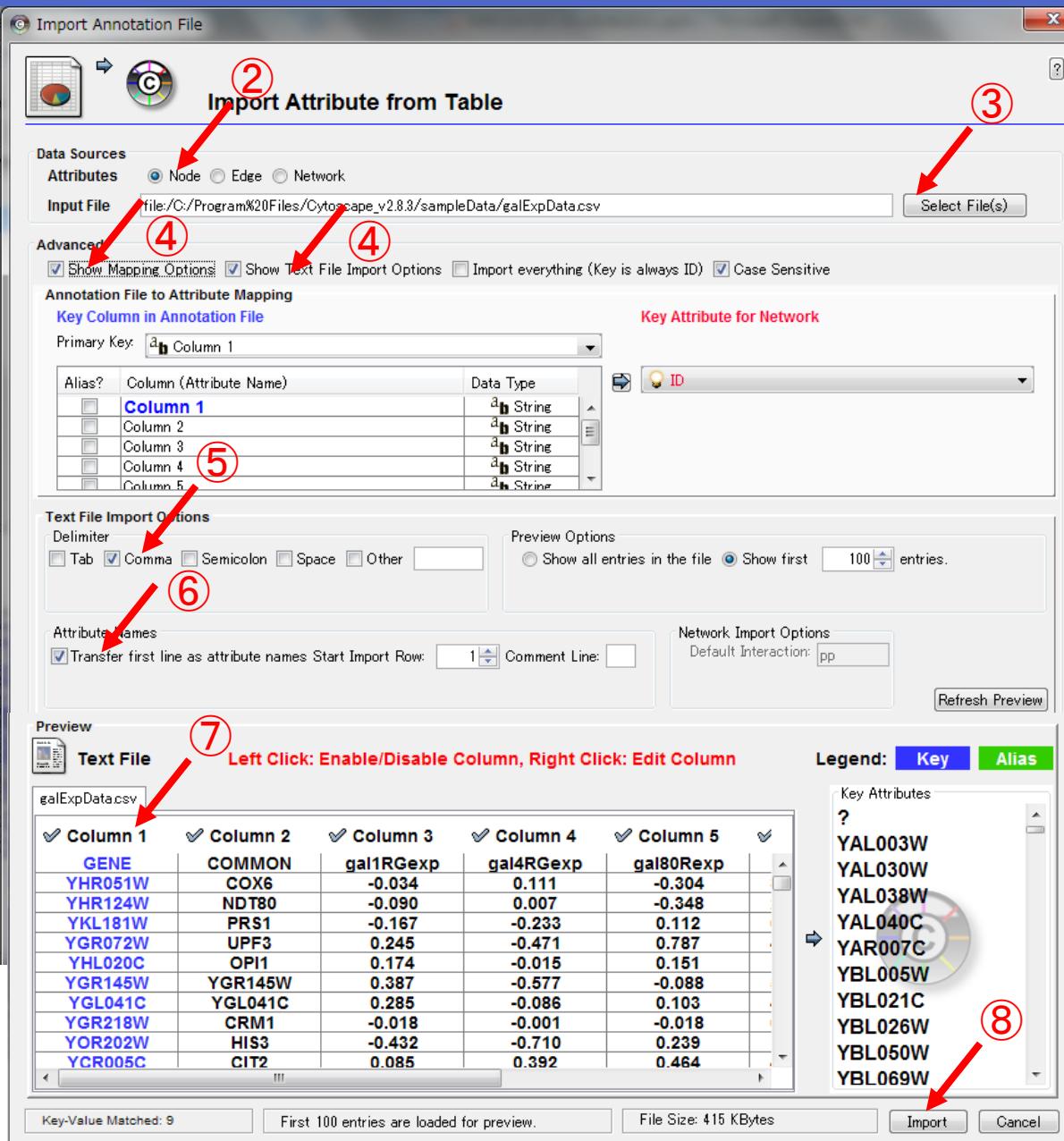
④Sourceを「Column1」、Targetを「Column3」、Typeを「Column2」

⑤「Column4,5,6」を選択

⑥「Import」をクリック

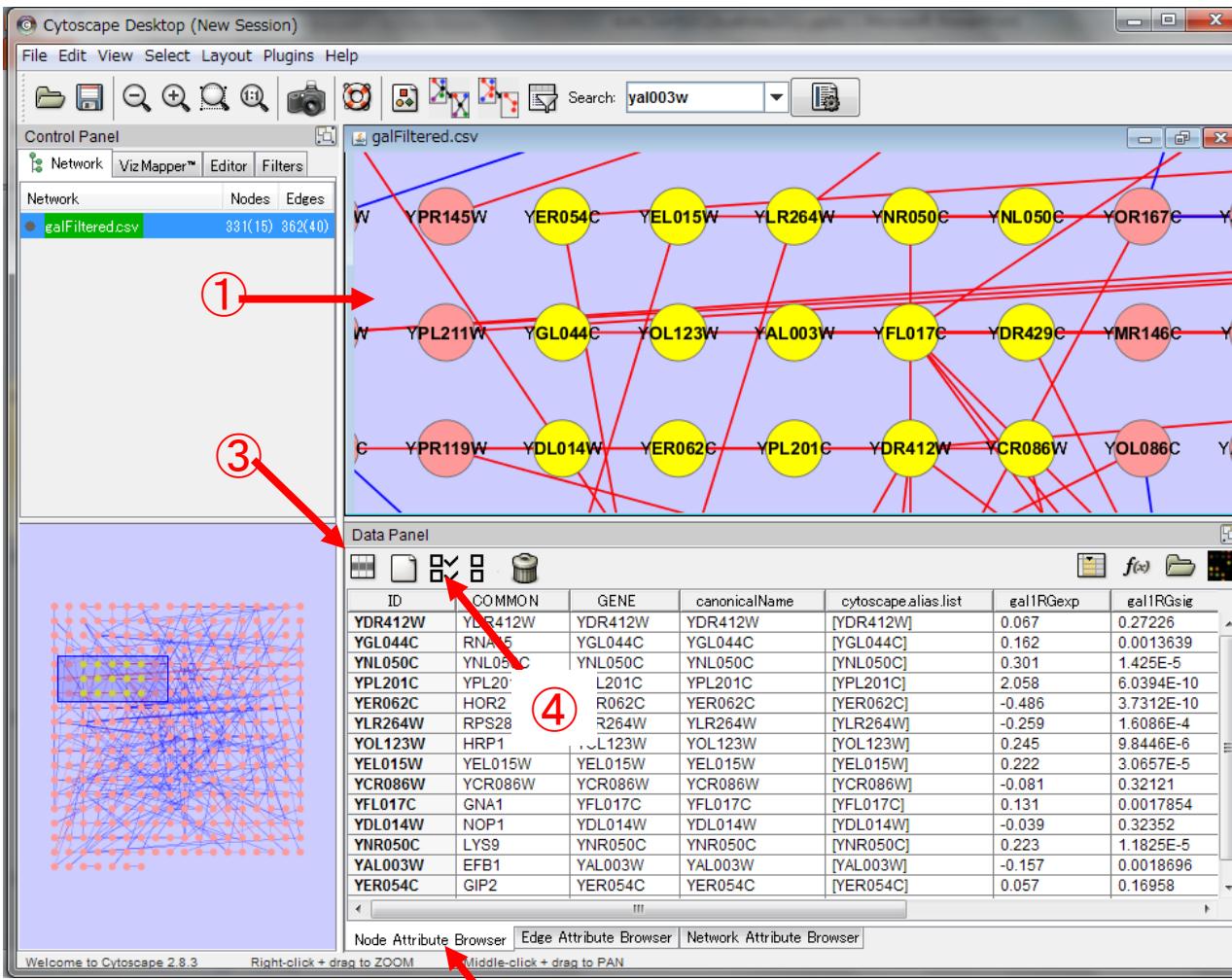


ステップ2:属性値を読み込む



- ①メニュー「File」の「Import」、「Attribute from Table (Text/MS Excel)」
- ②「Node」を選択
- ③ファイル「galExpData.csv」を選択
- ④「Show Mapping Options」、「Show Text File Import Options」に✓。
- ⑤「Comma」を選択
- ⑥「Transfer first line as attribute names Star Import Row」にチェック、「1」を入力
- ⑦「Column1」がキーであること(青色)を確認。
- ⑧「Import」をクリック

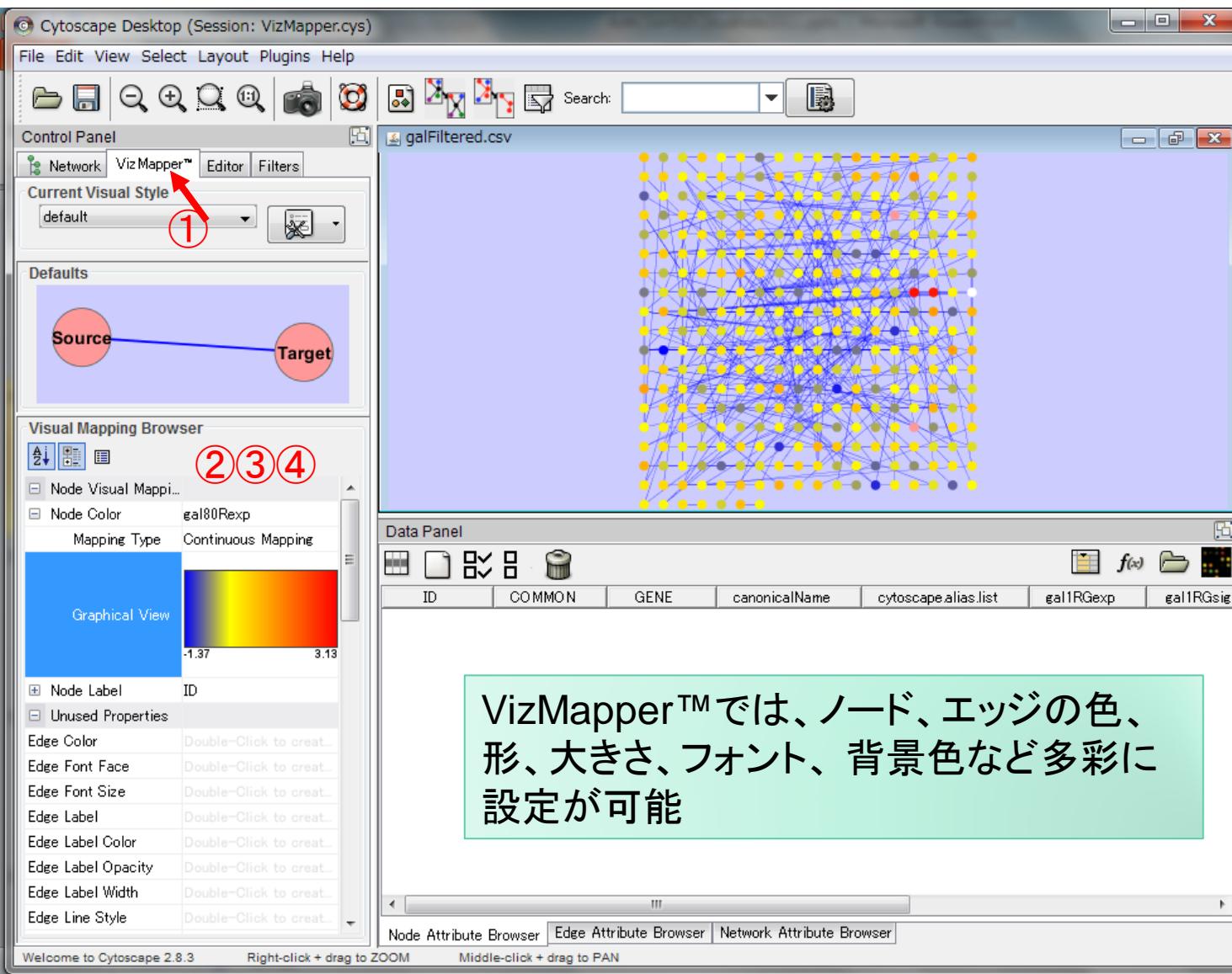
実習5. Cytoscapeフォルダにあるサンプルデータ 「galFiltered.***」や、自分で作成したテキスト、Excelファイル を使ってCytoscapeでパスウェイを表示、編集してみましょう。



ノード、エッジの属性の確認、編集

- ①属性の確認、編集したいノードやエッジを選択する
- ②ノード、エッジもしくは、ネットワーク全体を選択
- ③表示したい属性に✓を入れる
- ④このアイコンに押すとすべての属性が選択される

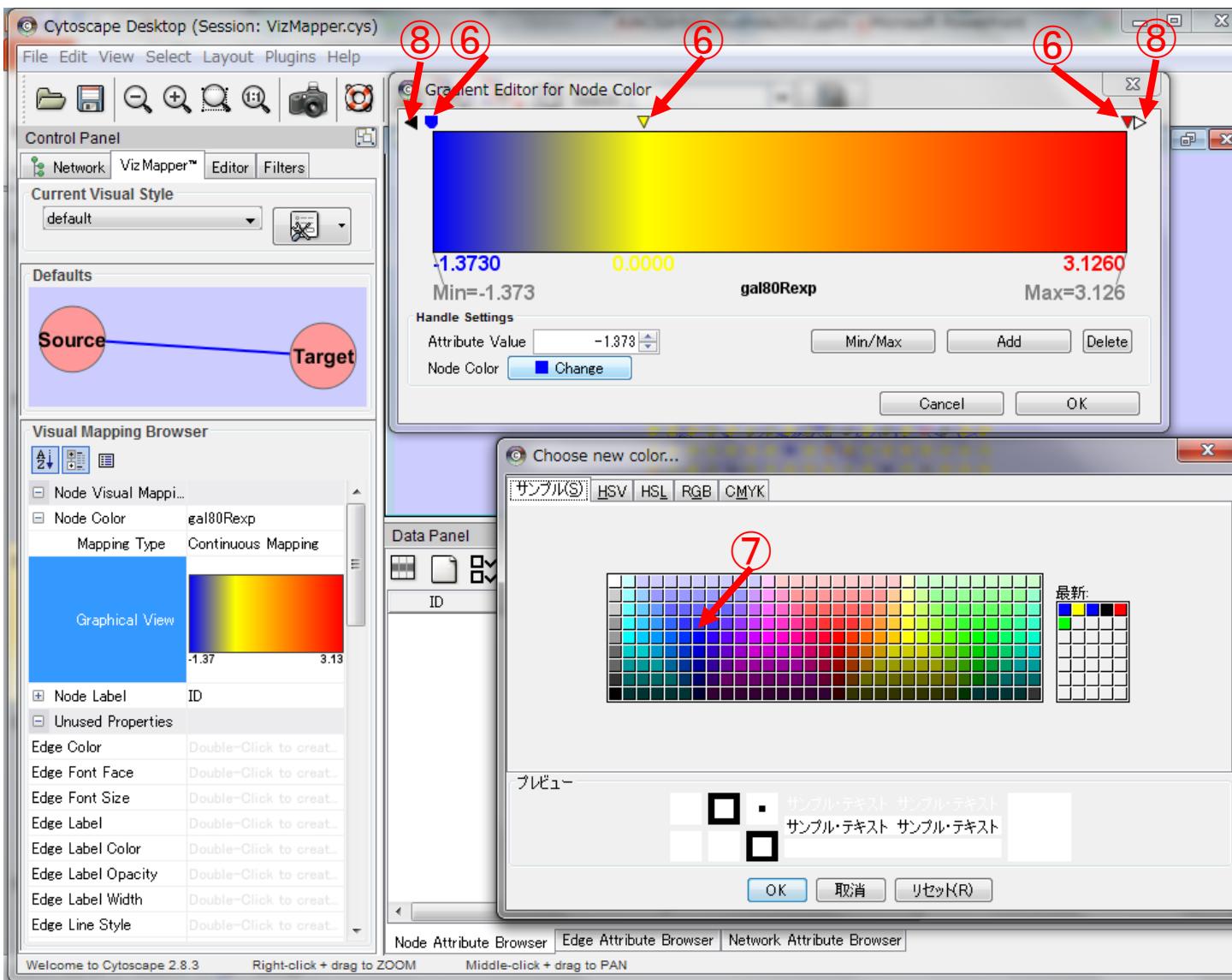
VizMapper™を使ったノード色の編集 1 of 2



VizMapper™では、ノード、エッジの色、形、大きさ、フォント、背景色など多彩に設定が可能

- ① Control Panelで「VizMapper」を選択
- ② Visual Mapping Browserの「Node Color」を選択、ノードの属性が表示されるので、例えば「gal180Rexp」(発現量)を選択。
- ③ 「Mapping Type」から、「Continuous Mapping」を選択「galExpData.csv」を選択
- ④ 「Graphical View」を選択、「Gradient Editor」のウィンドウが表示される

VizMapper™を使ったノード色の編集 2 of 2

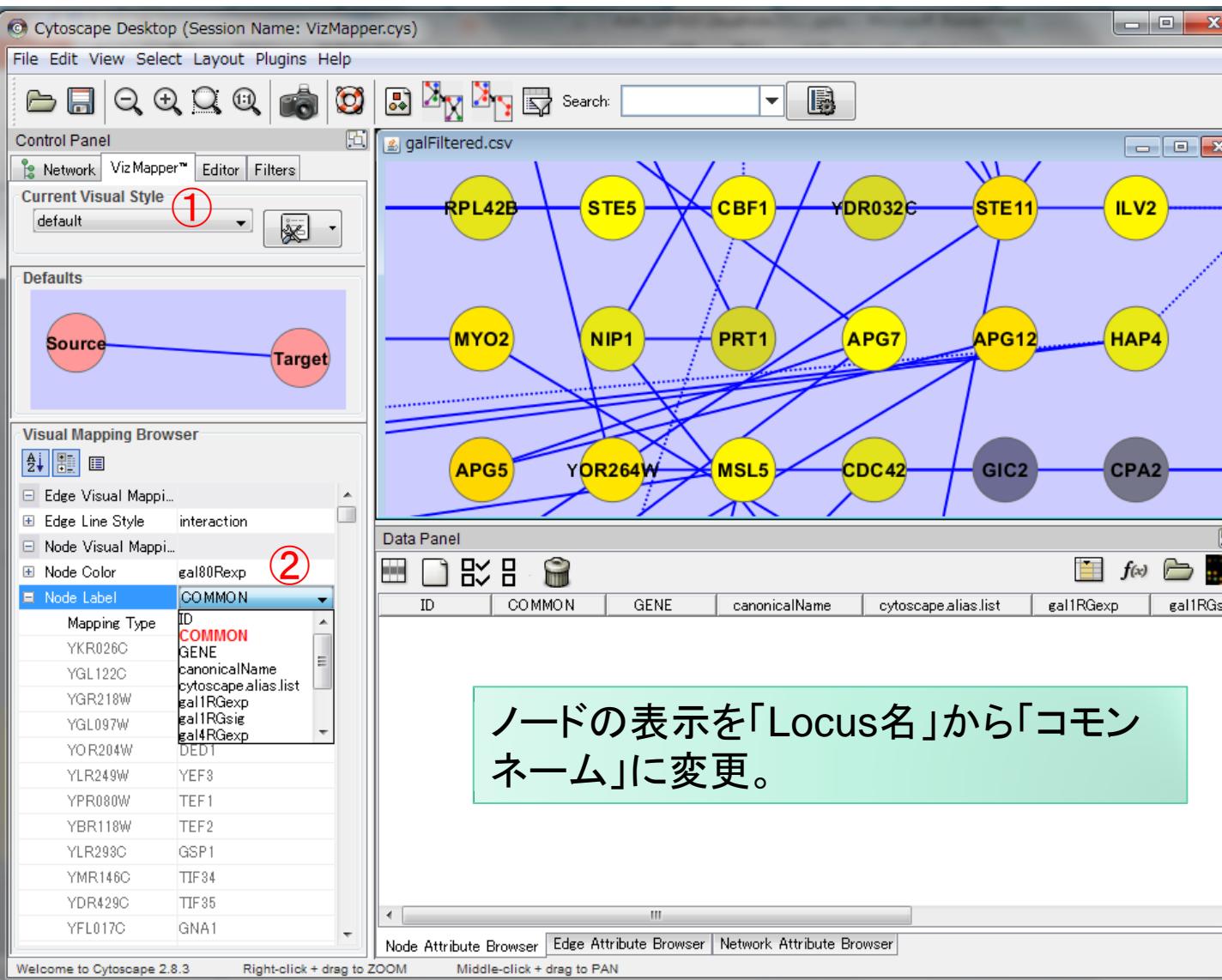


⑥「Gradient Editor」で発現量に応じた色の指定を行う。色帯の上部の三角形を選択し、スライドさせ適当な位置でダブルクリックして、「Choose new color」のウィンドウを表示

⑦色を指定。この例では、発現量の差が最小の場合を青、最大の場合を赤、発現量に差が見られなかった場合(発現量0)を黄色に指定。

⑧最大値以上、最小値以下の色も、赤、青に指定。

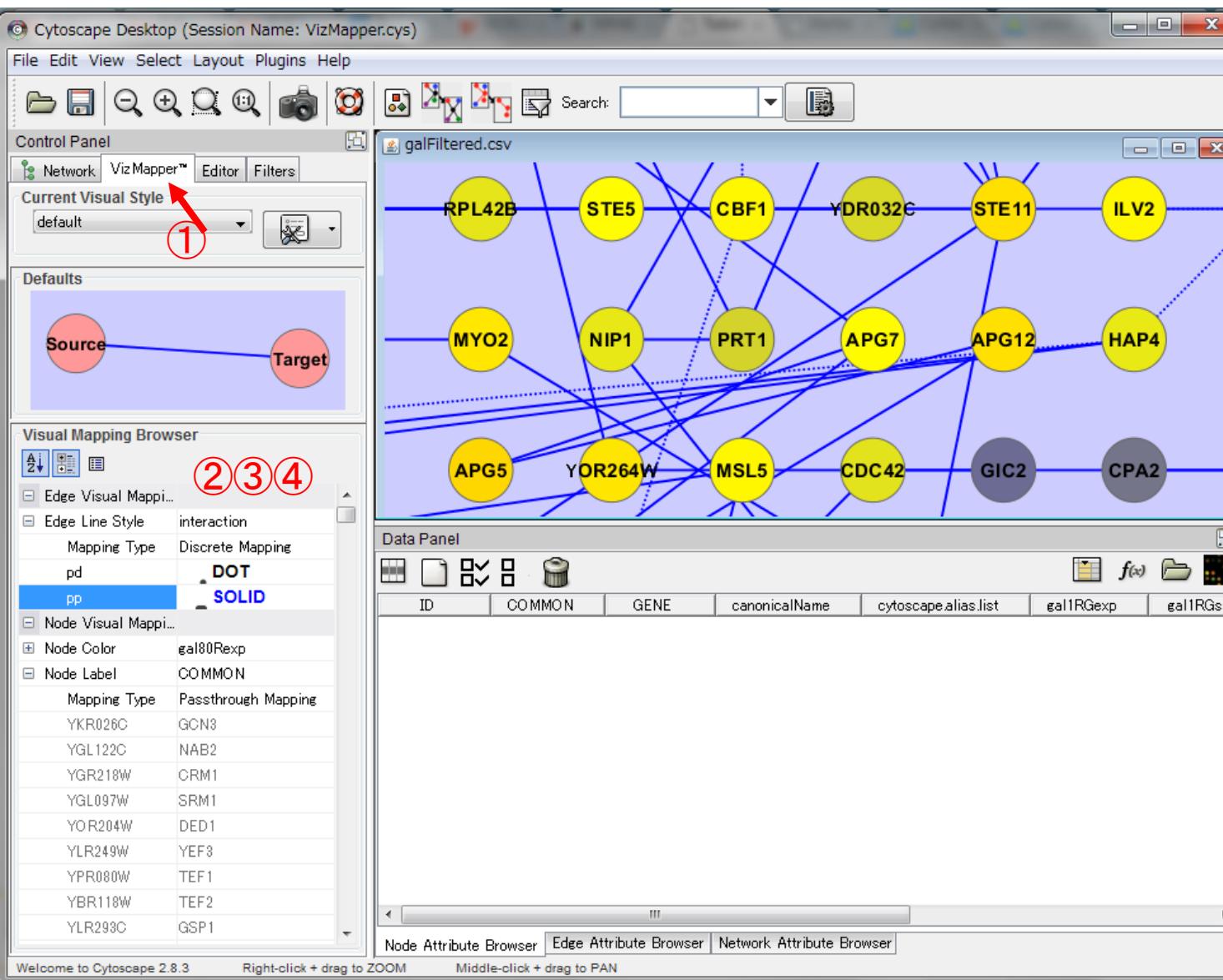
VizMapper™を使ったラベルの編集



- ① Control Panelの「VizMapper」、「Node Label」を選択
- ② 「Mapping Type」で「COMMON」を選択

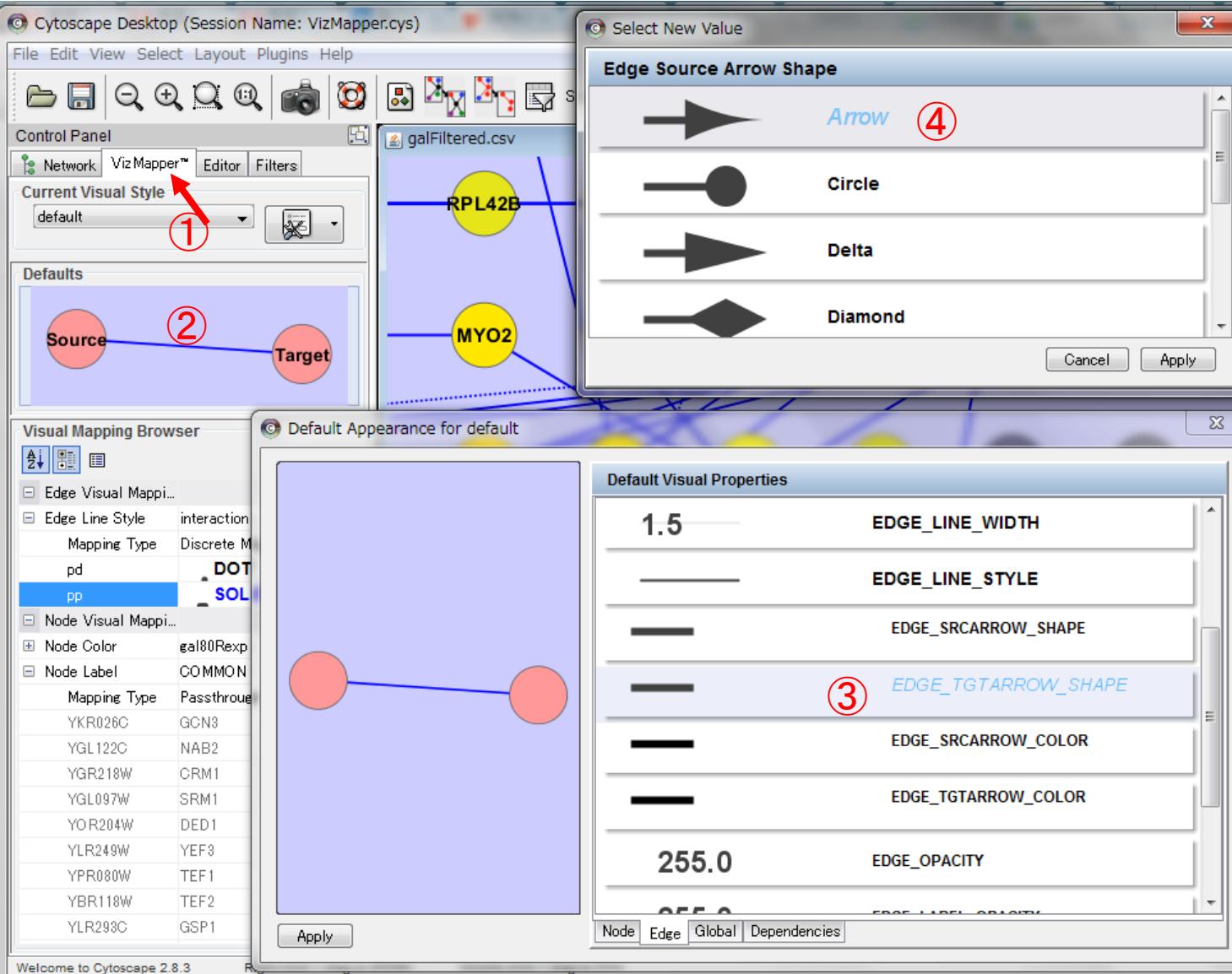
ノードの表示を「Locus名」から「コモンネーム」に変更。

VizMapper™を使ったエッジの編集



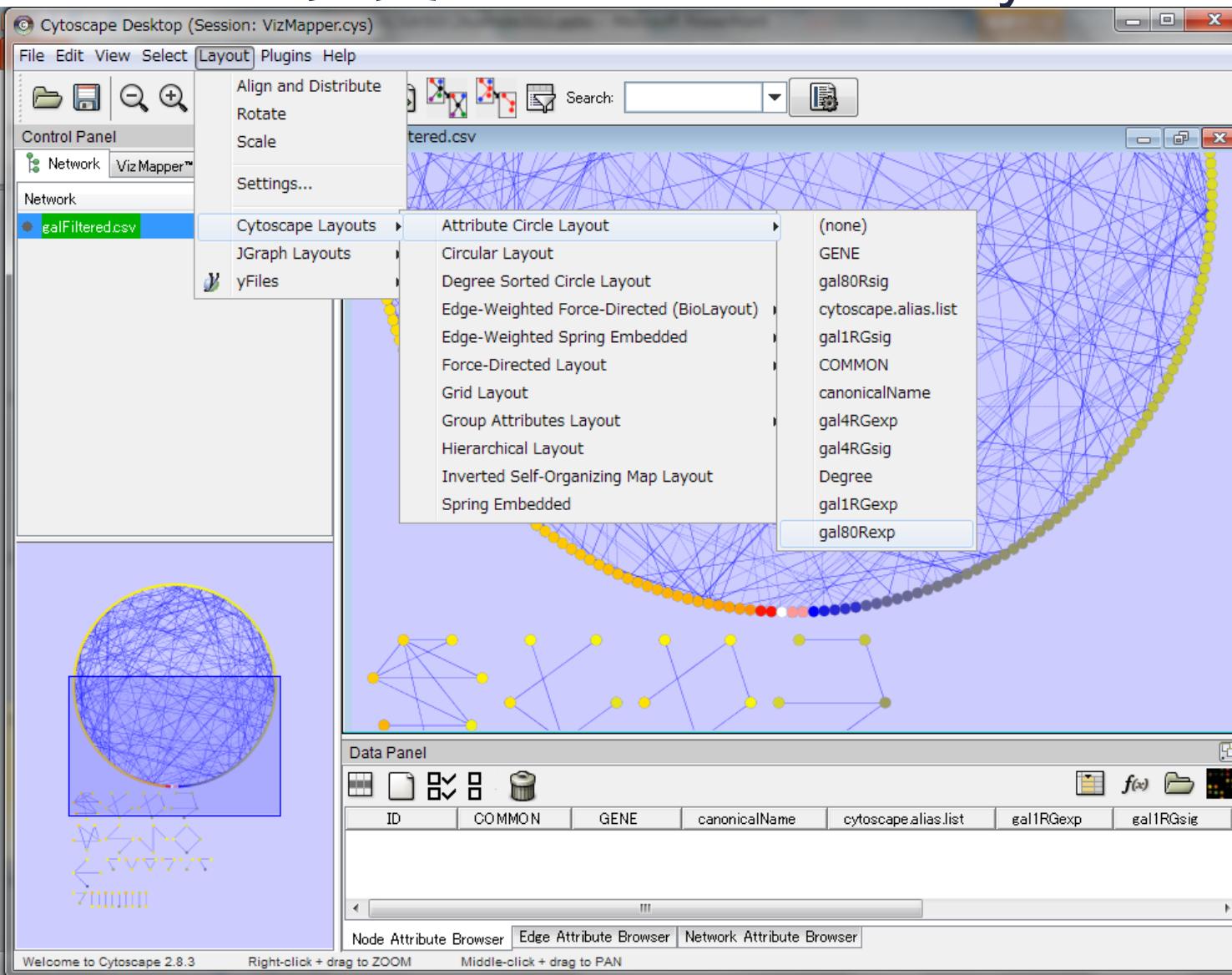
- ① Control Panel で「VizMapper」を選択
- ② Visual Mapping Browser の「Edge Line Style」を選択、エッジの属性が表示されるので、「interaction」を選択。
- ③ 「Mapping Type」から、「Discrete Mapping」を選択
- ④ 「pd」(タンパク質-DNA結合)を「DOT」、「pp」(タンパク質-タンパク質結合)に指定。

VizMapper™を使ったデフォルト値の編集



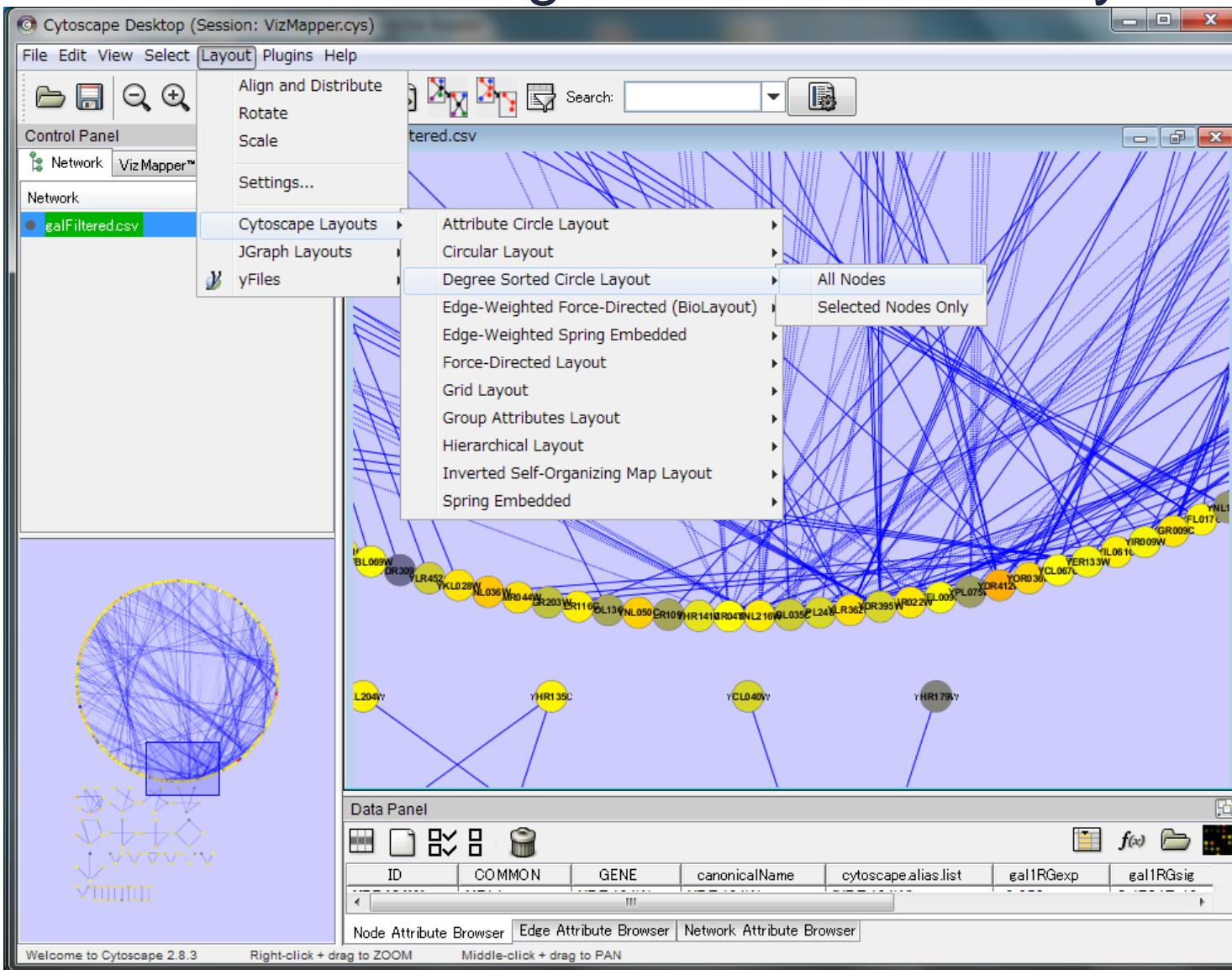
- ① Control Panelで「VizMapper」を選択
- ② Defaultsの図をクリック。
- ③ Default Appearance for defaultの「EDGE_TGTARROW_SHAPE」を選択
- ④ Select New Valueの「Arrow」を選択。

レイアウト変更:Attribute Circle Layout



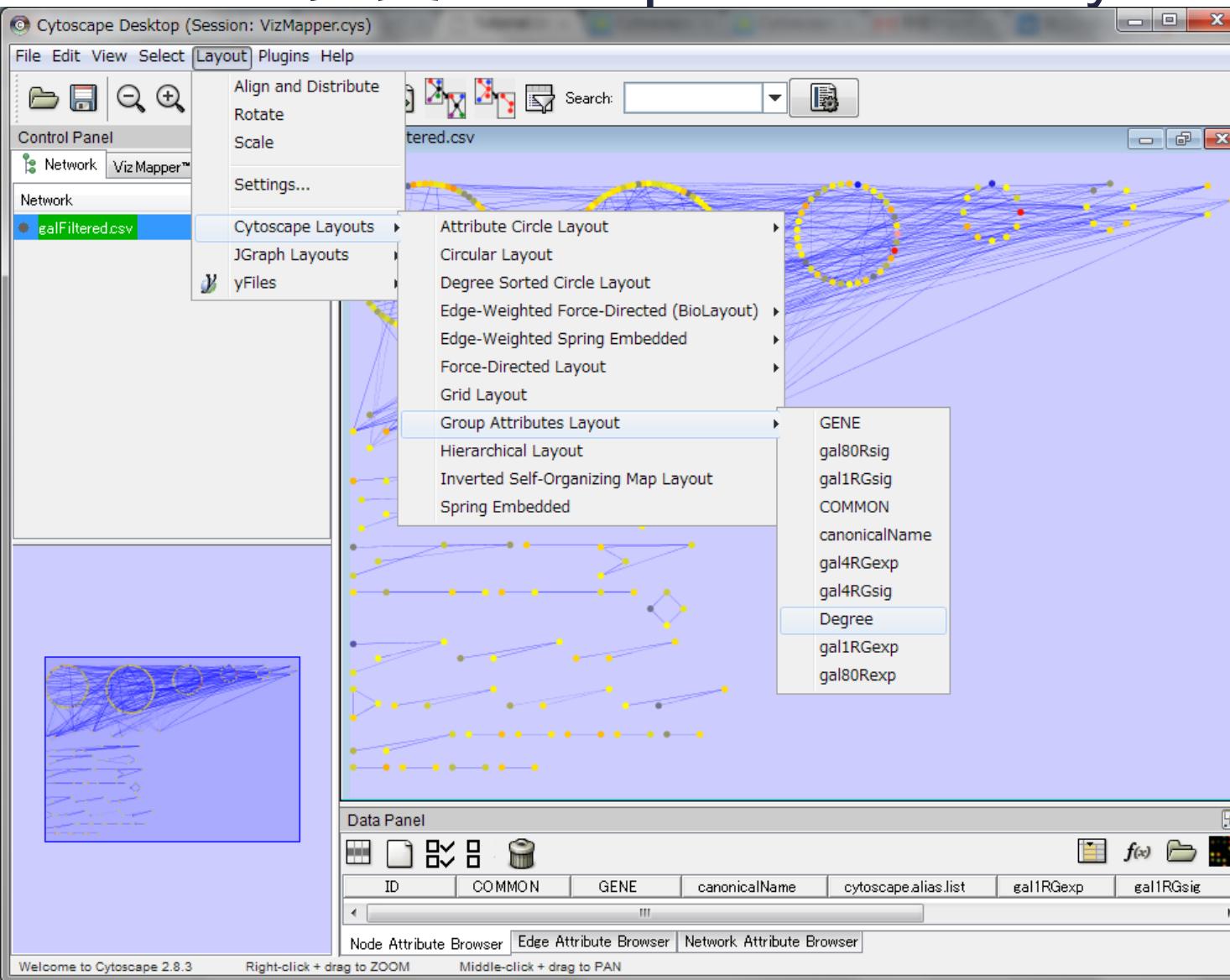
- ①ノードの属性値の順に環状グラフの下部から時計回りに配置するレイアウト
- ②メニュー「Layout」、「Cytoscape Layouts」、「Attribute Circle Layout」、「gal80Rexp」を選択

レイアウト変更: Degree Sorted Circle Layout



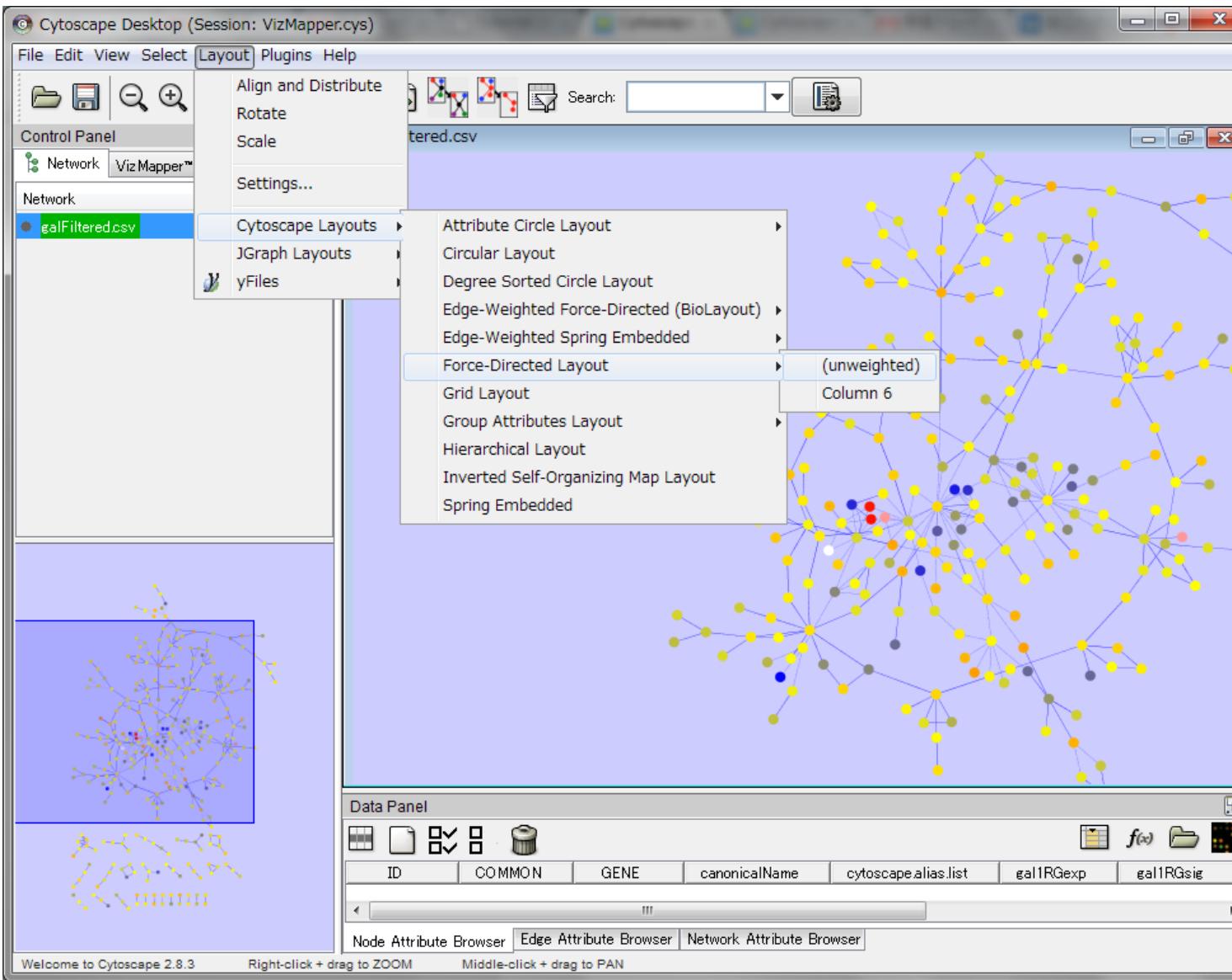
- ①ノードが持つエッジ数の多いものからの環状グラフの下部から反時計回りに配置するレイアウト
- ②メニュー「Layout」、「Cytoscape Layouts」、「Degree Sorted Circle Layout」、「All Nodes」を選択。このとき「Selected Nodes Only」選択した場合、そのノードだけでレイアウト変更する

レイアウト変更: Group Attribution Layout



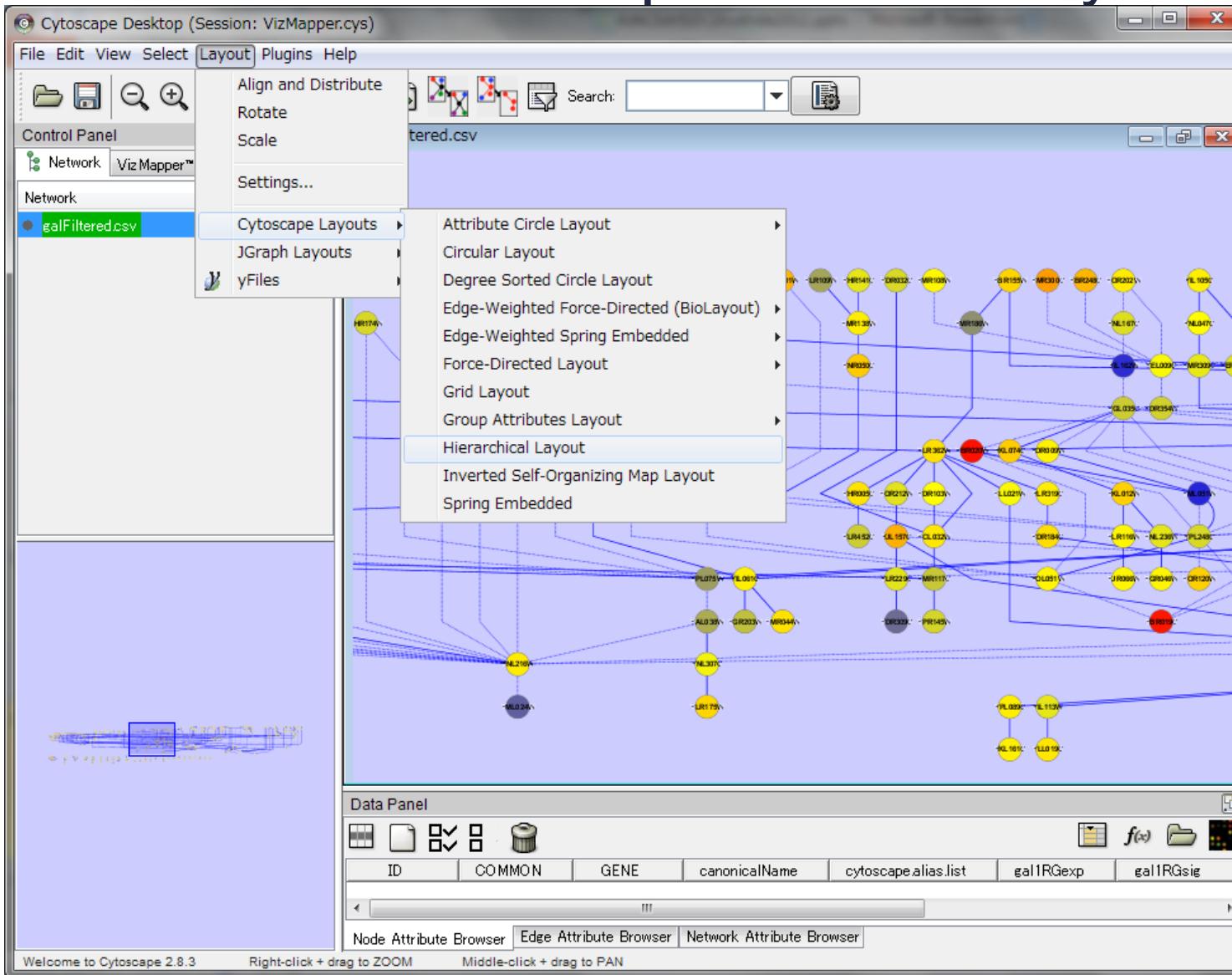
- ① ノードが持つエッジ数の同じものを同じ環状グラフに配置するレイアウト
- ② メニュー「Layout」、「Cytoscape Layouts」、「Group Attribution Layout」、「Degree」を選択。

レイアウト変更: Force-Directed Layout



- ① グラフの詳細な構造を表すのに適したレイアウト
- ② メニュー「Layout」、「Cytoscape Layouts」、「Force-Directed Layout」、「(unweighted)」を選択。
- ③ 「Layout」、「Setting...」でノード間の距離などを変更することが可能。

レイアウト変更: Group Attribution Layout



- ①パスウェイを階層的に表現するレイアウト
- ②メニュー「Layout」、「Cytoscape Layouts」、「Hierarchical Layout」を選択。

データ解析の例

(参照 : [Basic Expression Analysis - Yeast](#))

サンプルデータ概要

- 生物種は出芽酵母
- 転写因子 Gal1, Gal4, Gal80などの遺伝子摂動株を用いたマイクロアレイ遺伝子発現量解析
- タンパク質-DNA相互作用(pd)を用いたネットワーク解析

フィルタ機能を使った絞り込み

Cytoscape Desktop (Session: VizMapper.cys)

File Edit View Select Layout Plugins Help

Control Panel

① Network Viz Mapper™ Editor Filters (1)

Current Filter session: Default

Filter Definition edge.interaction (2)

+ Advanced interaction Not pd (3)

④ Apply Filter Select All Deselect All

Network Nodes Edges

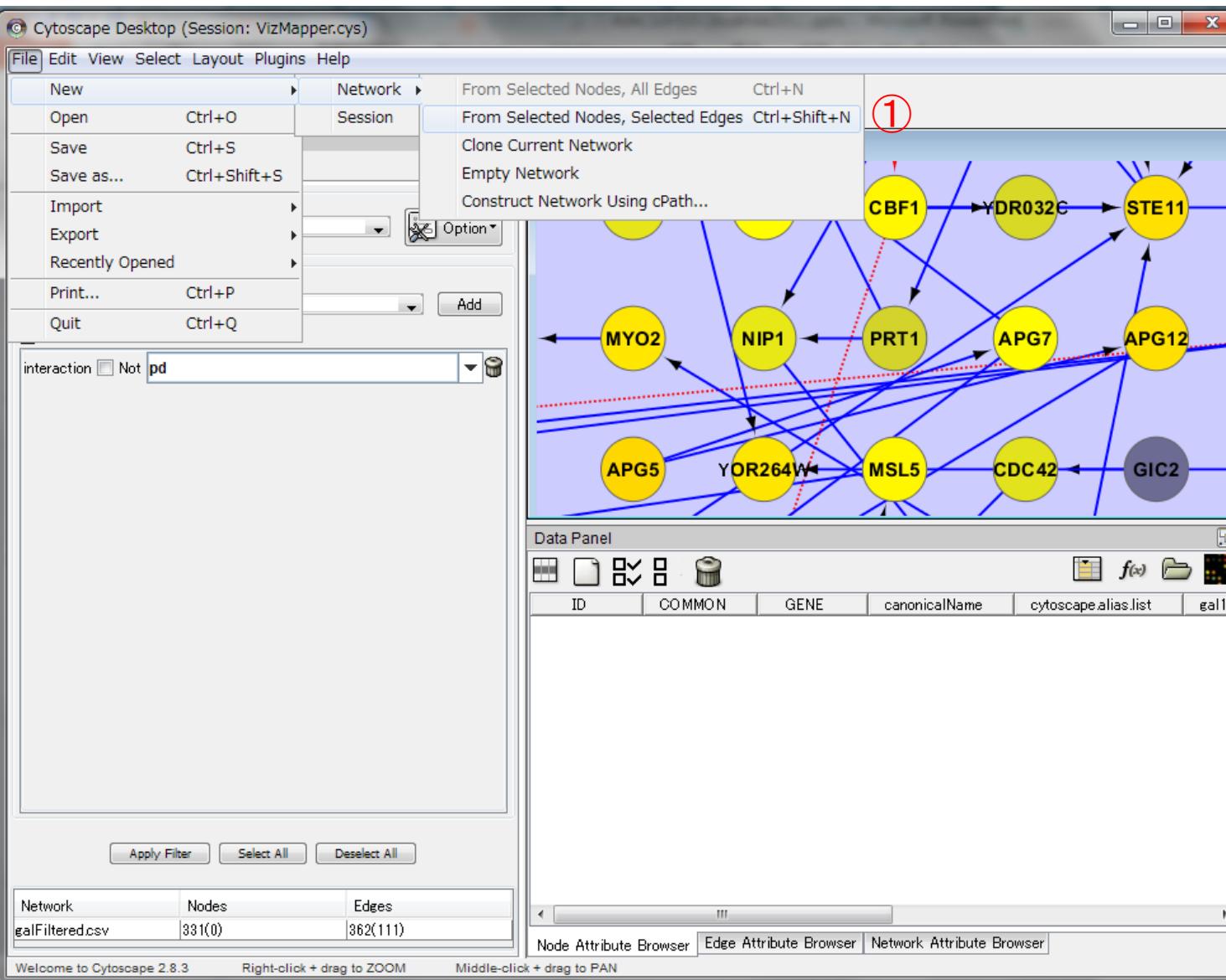
ID	COMMON	GENE	canonicalName	cytoscape.alias.list	gal1R
galFiltered.csv	331(0)	362(111)			

Welcome to Cytoscape 2.8.3 Right-click + drag to ZOOM Middle-click + drag to PAN

Node Attribute Browser Edge Attribute Browser Network Attribute Browser

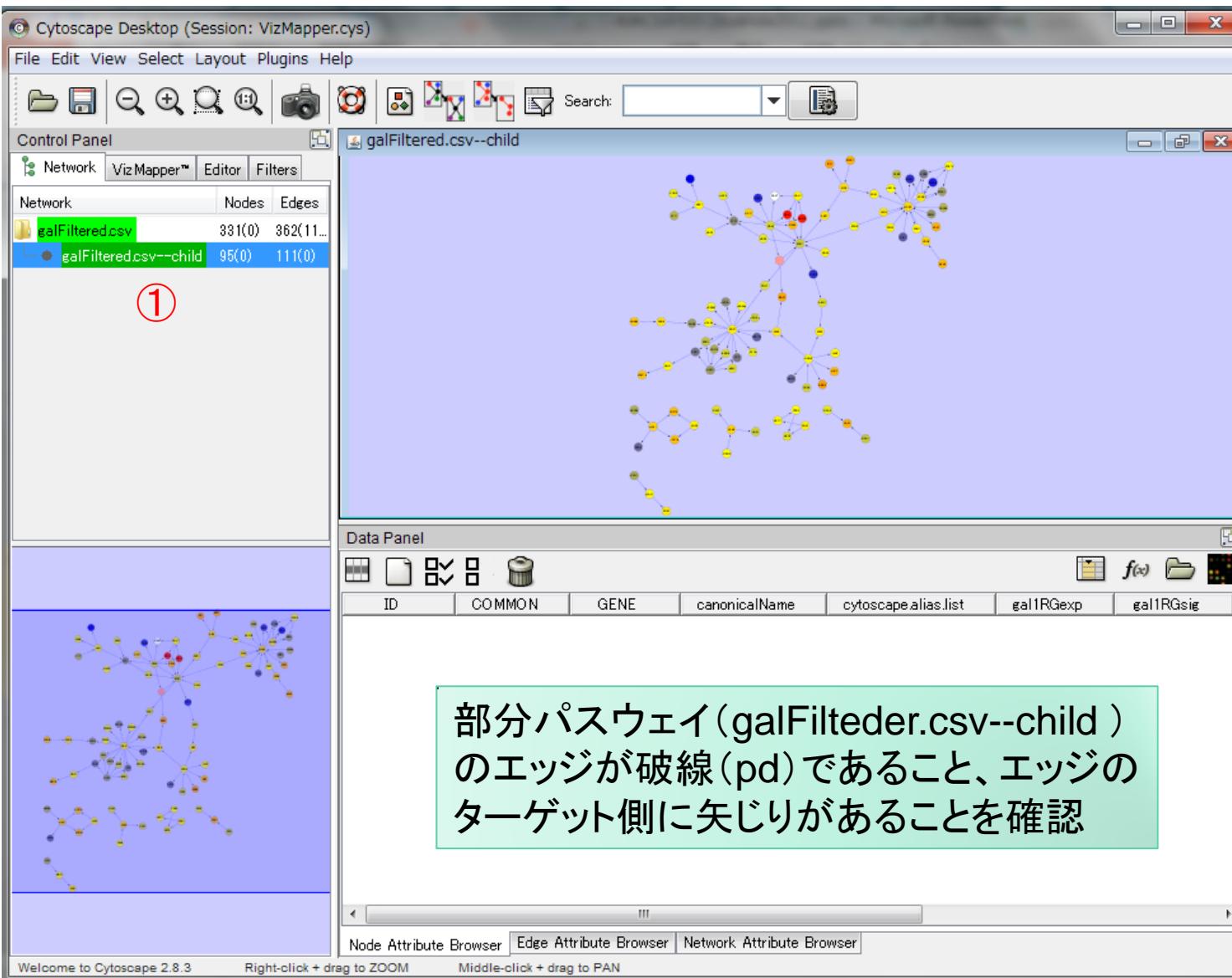
- ① Control Panel の「Filters」を選択
- ② Filter Definitionから「edge:interaction」を選択、次いで「Add」を押す
- ③ Advancedのクリエイター欄に「pd」を入力
- ④ 「Apply Filter」を押す
- ⑤ タンパク質-DNA相互作用(pd)を表す破線が赤く選択されていることを確認

サブパスウェイ(部分パスウェイ)の抽出 1 of 5



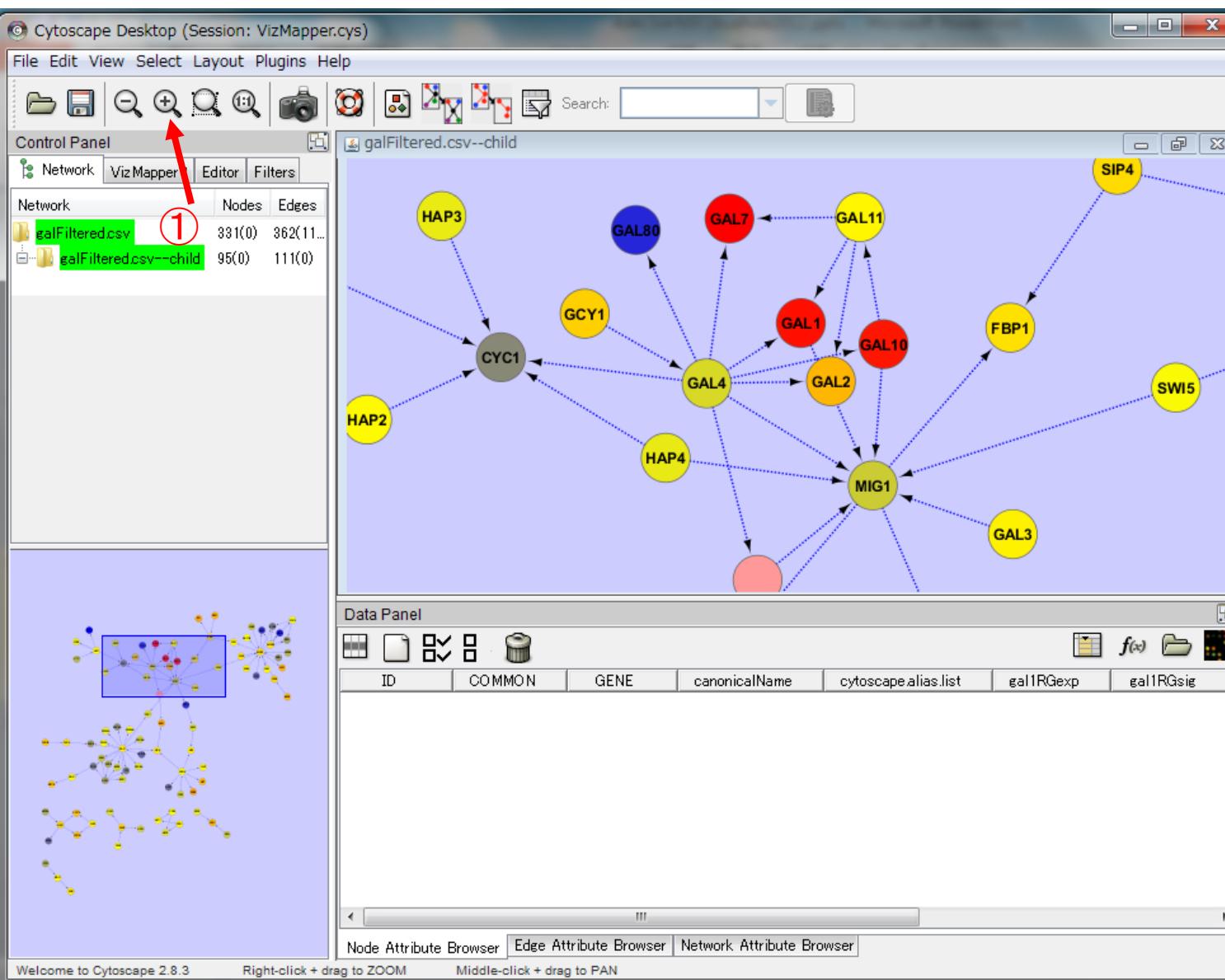
①メニュー
 「File」、
 「New」、
 「Network」、
 「From
 Selected
 Nodes,
 Selected
 Edges」を選
 択

サブパスウェイ(部分パスウェイ)の抽出 2 of 5



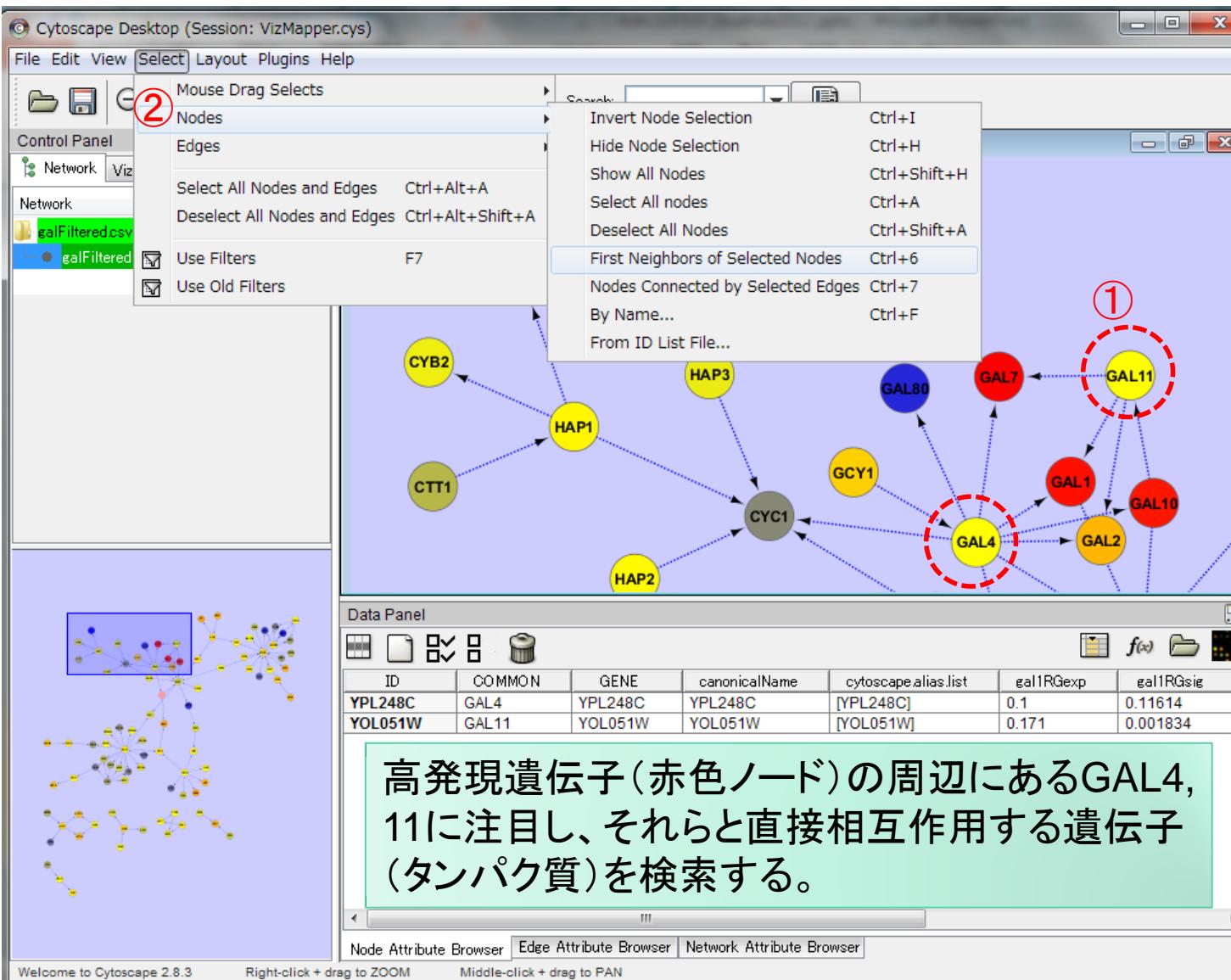
- ①Control Panel の「Network」で、元のパスウェイ (galFiltered.csv) の下位に、部分 パスウェイ (galFiltered.csv-child) が作成されたことを確認
- ②メニュー 「Layout」、「Cytoscape Layouts」、「Force-Directed Layout」、「(unweighted)」を選択してレイアウト変更

サブパスウェイ(部分パスウェイ)の抽出 3 of 5



①メインネットワークビューもしくは、画面左下のネットワーク全体図から、赤色のノード(高発現遺伝子、GAL1, GAL7, GAL10)に注目し、その近辺を拡大

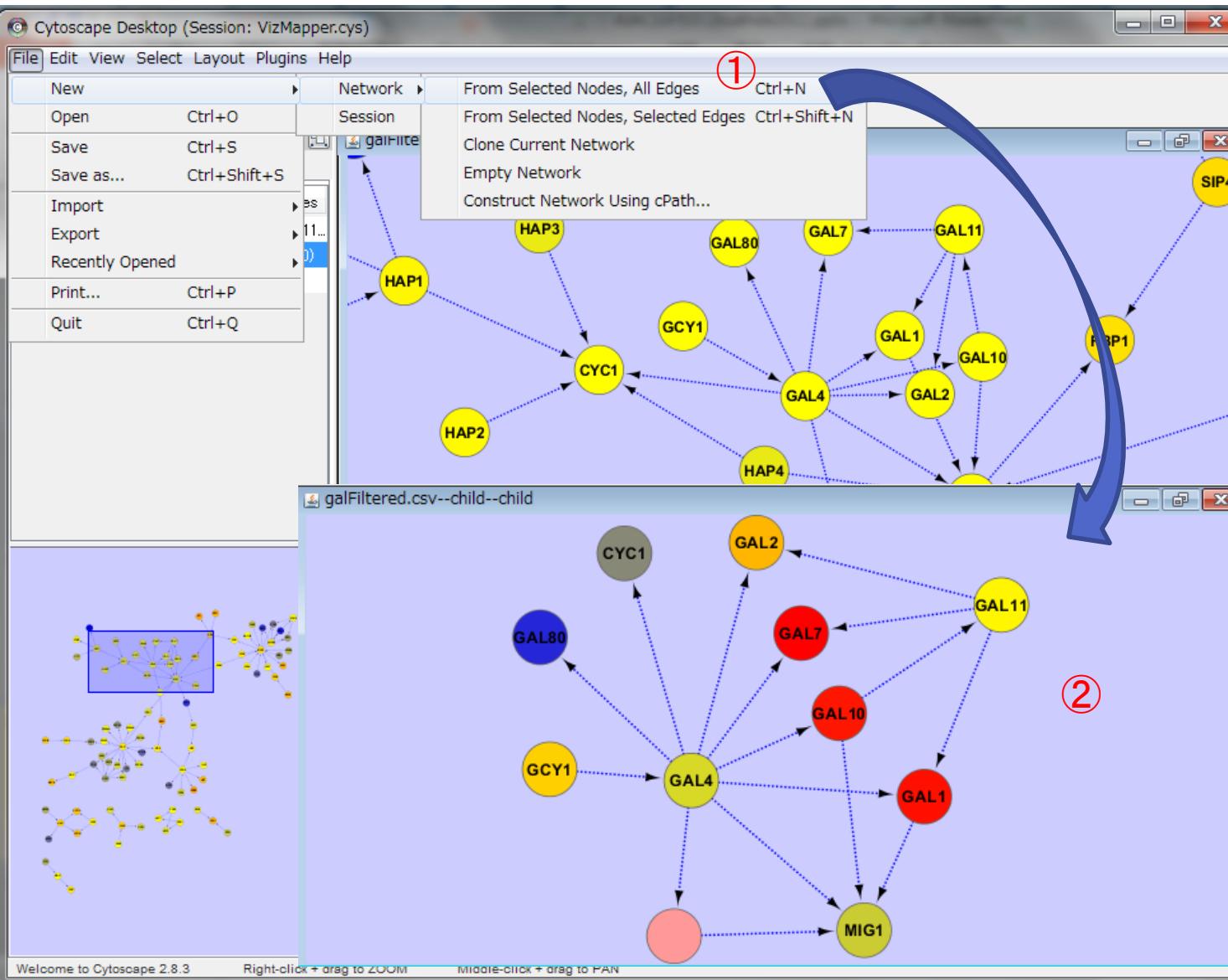
サブパスウェイ(部分パスウェイ)の抽出 4 of 5



①メインネットワークビューで、Shiftキーを押しながらGAL4, 11を複数選択

②メニュー「Select」、「Nodes」、「First Neighbors of Selected Nodes」を選択

サブパスウェイ(部分パスウェイ)の抽出 5 of 5



①メニュー
 「File」、
 「New」、
 「Network」、
 「From
 Selected
 Nodes, All
 Edges」を選
 択

②GAL4, 11と
 直接相互作用
 する遺伝子(タ
 ンパク質)を抽
 出

プラグインの紹介

Manage Plugins

データ解析、ネットワーク解析、データインポート等の拡張機能は「Manage Plugins」で導入、管理される。

①メニュー
「Plugins」、
「Manage
Plugins」を選択

データインポート

ネットワーク解析

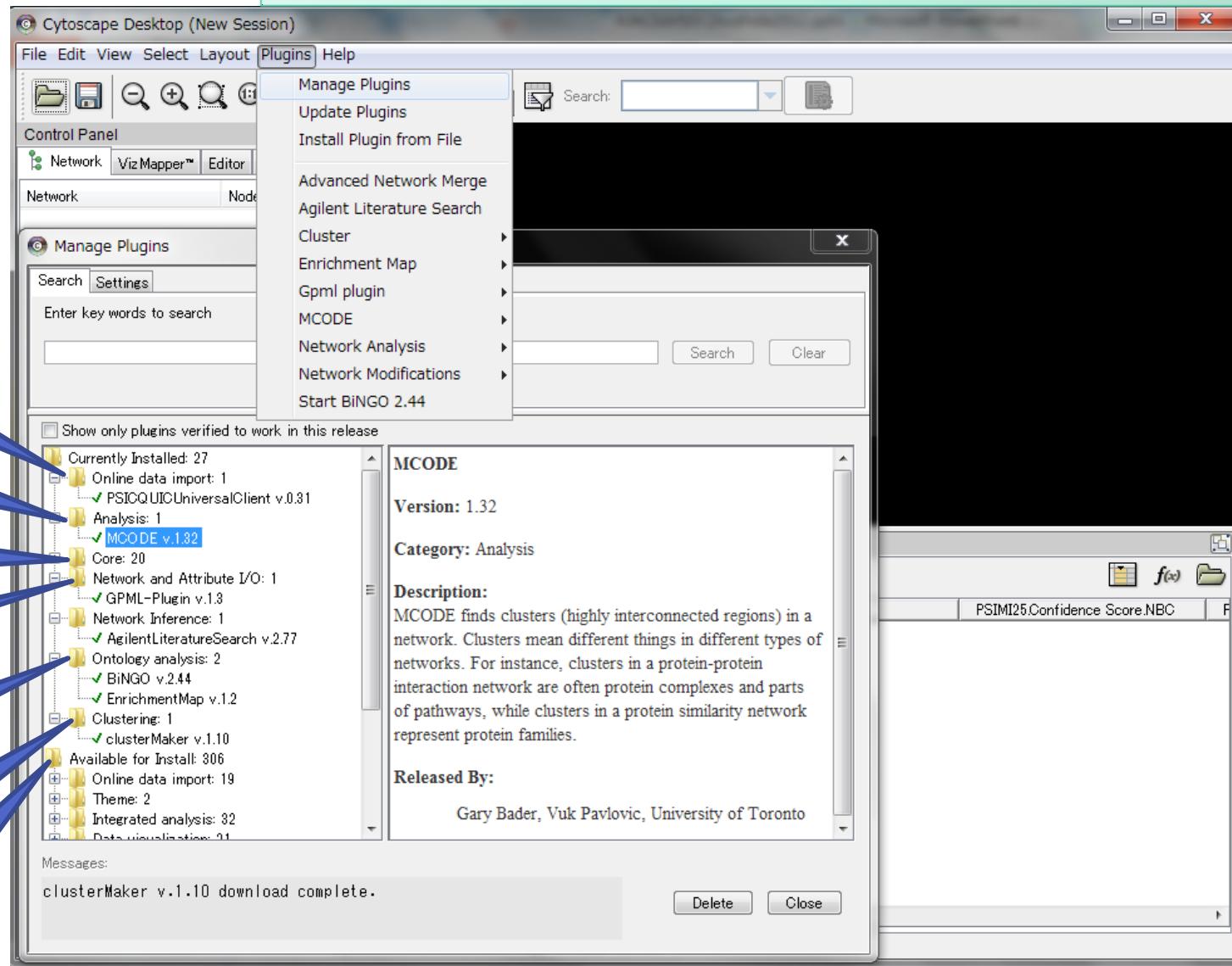
基本プラグイン

パスウェイデータ
取り込み

遺伝子(タンパク
質)機能予測

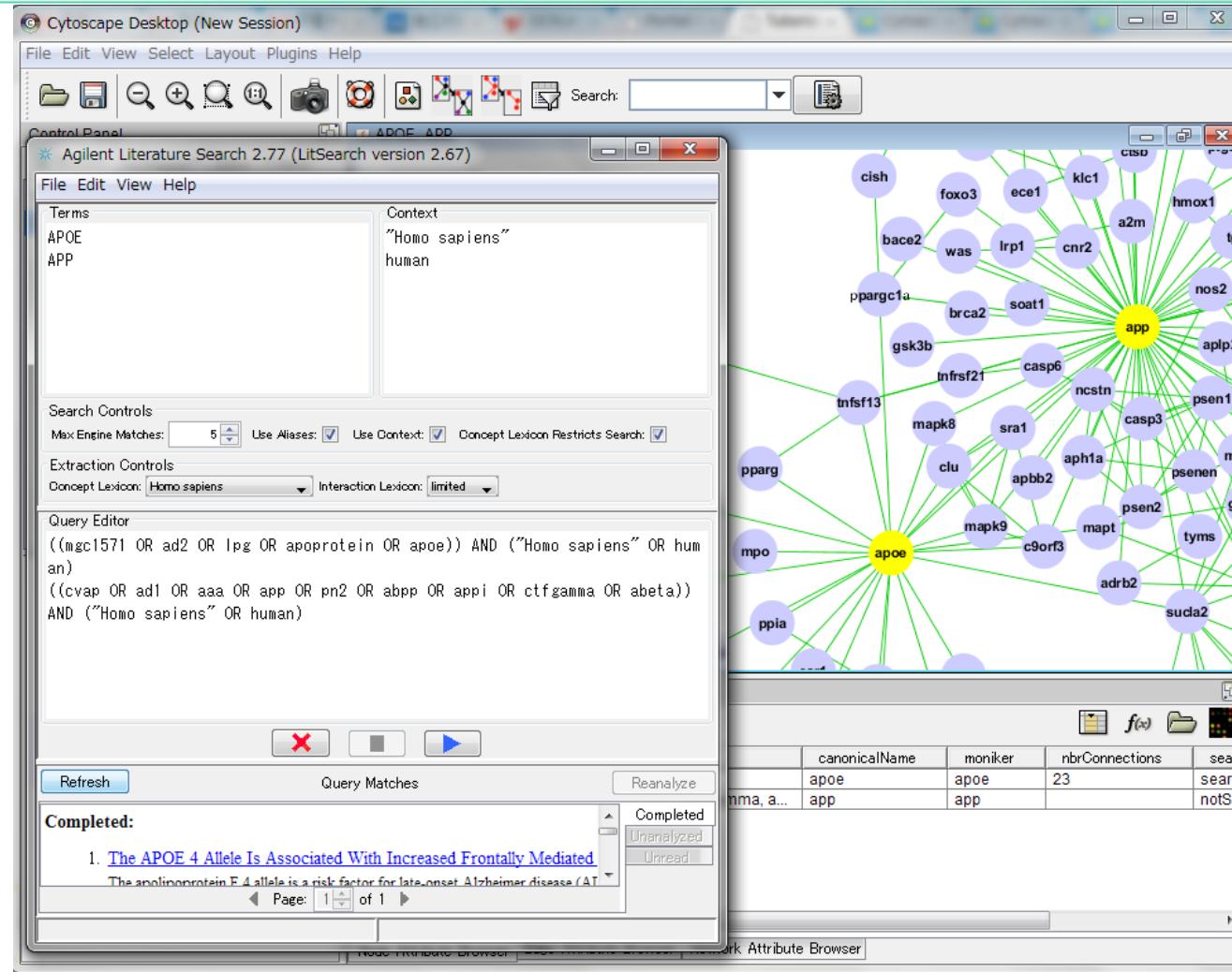
クラスター解析

未導入



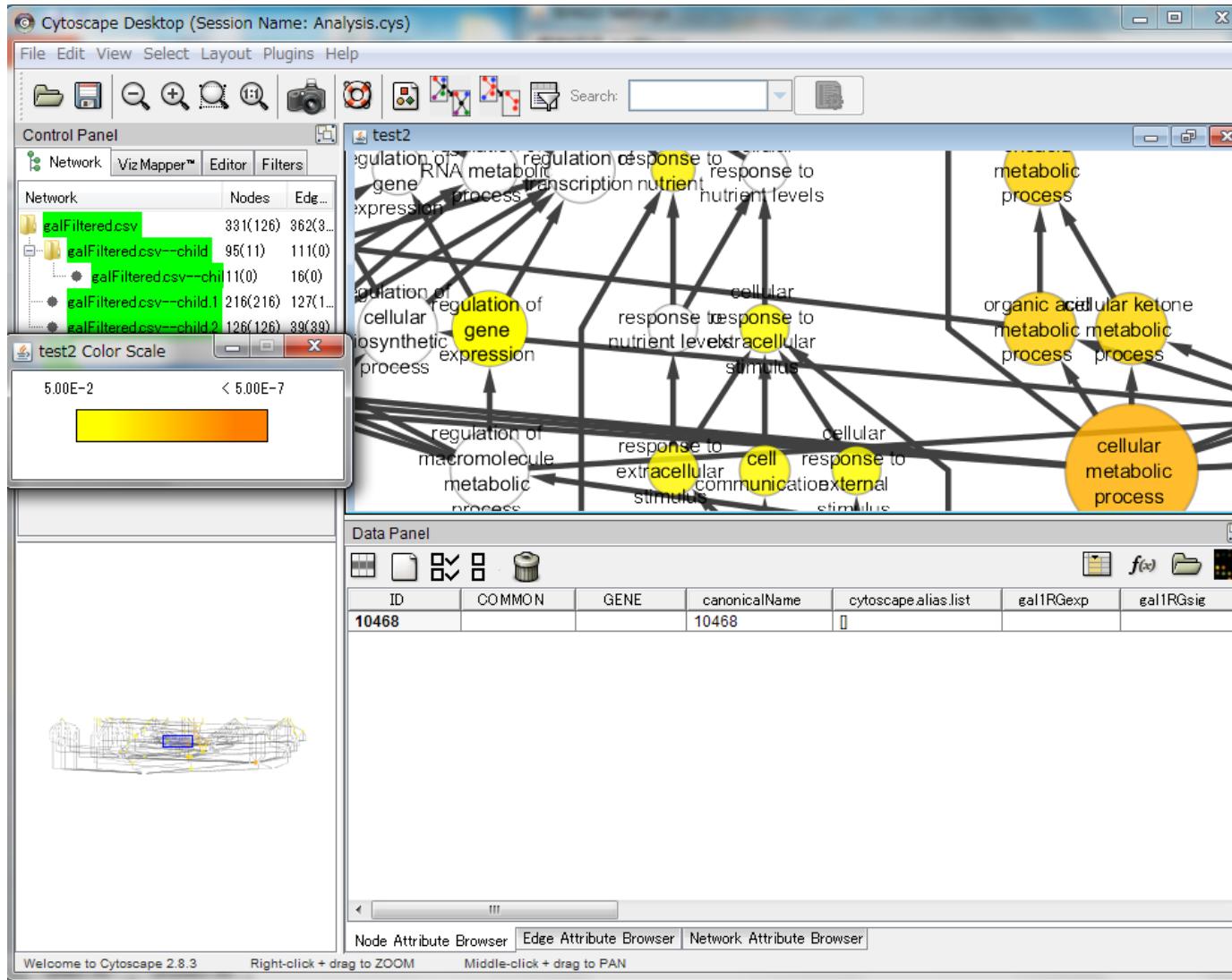
プラグイン例: Agilent Literature Search

Pubmed、OMIM、USPTO(米国特許商標庁)を情報元として、検索キーワードと関係のある相互作用情報をマイニングし、ネットワーク表示するツール



プラグイン例: BiNGO

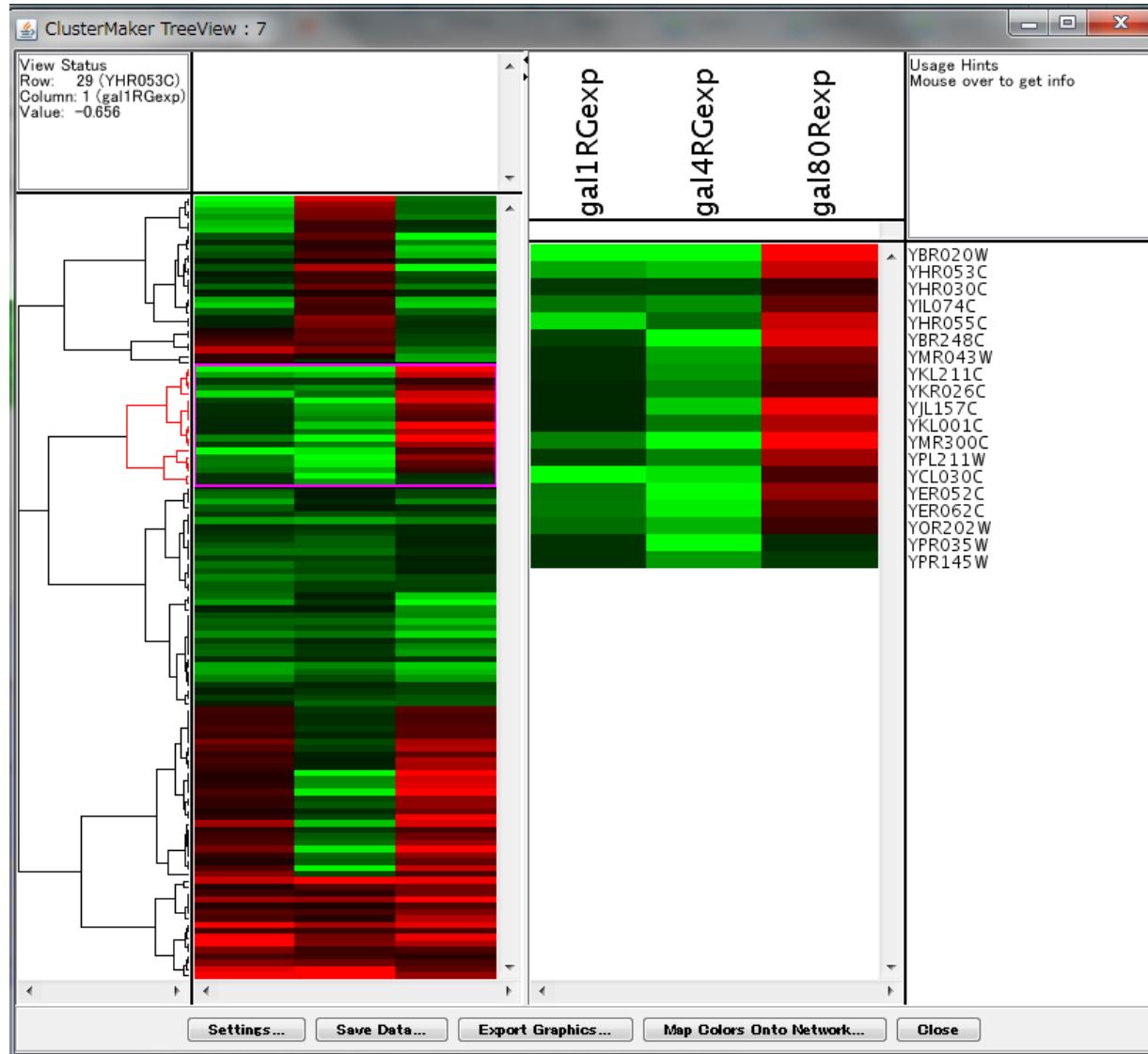
過剰発現遺伝子などを対象に、GeneOntologyを使って機能予測するツール



類似のツール:
Enrichment Map

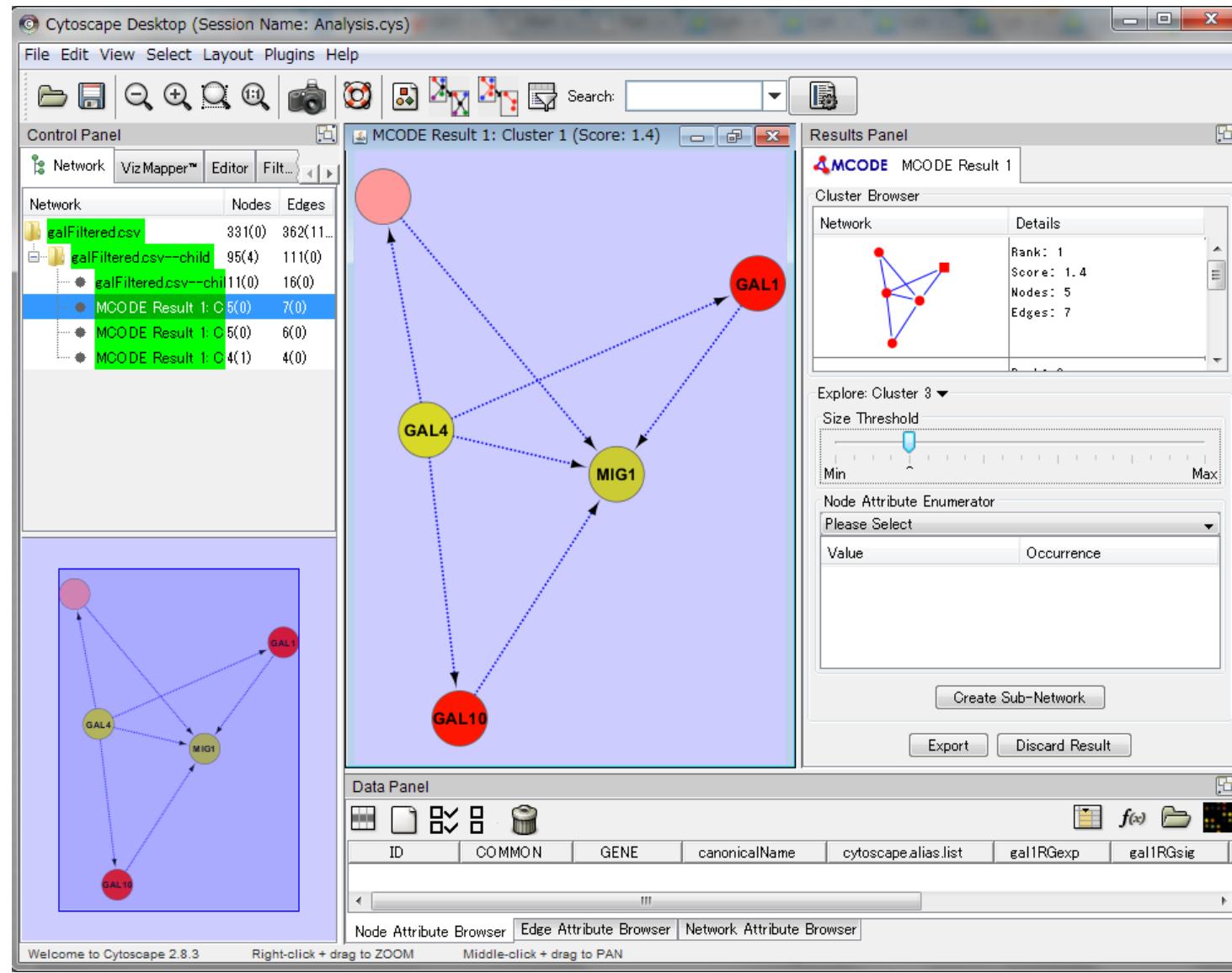
プラグイン例: clusterMarker

階層クラスタリングやk-means法による遺伝子クラスタリングを行うツール



プラグイン例: MCODE

ネットワーク分析により、大規模なネットワークの中からクラスターを発見するツール



知っていると便利なよく使う操作方法

1. あるノードに直接リンクするノードを指定する。
 1. Ctrl + 6
 2. 上記操作を繰り返すことで、数ステップを介してリンクするノードを指定することもできる。
2. 作成したパスウェイを削除する。
 1. Controlパネルの「Network」で、対象のパスウェイを選択。
 2. 右クリックで、「Destroy Network」を選択。
3. 作業(パスウェイの編集、作成)の内容をすべて消去して、最初から作業し直す。
 1. メニュー「File」の「New」、「Session」を選択。
4. ネットワークを結合(マージ)する。
 1. メニュー「Plugins」の「Advanced Network Merge」を選択
 2. Operationで「union」を選択し、マージしたいネットワークを選択。「右向き矢印」を押し、「Merge」ボタンを押す。

参考

情報提供、共有サイト 1 of 2

- Cytoscape 2.x のチュートリアル集
 - <http://opentutorials.cgl.ucsf.edu/index.php/Portal:Cytoscape>
 - チュートリアル1
 - <http://opentutorials.cgl.ucsf.edu/index.php/Portal:Cytoscape>
 - Basic Expression Analysis – Yeast
 - http://opentutorials.cgl.ucsf.edu/index.php/Tutorial:Basic_Expression_Analysis_in_Cytoscape
- Cytoscape Japanese Documentation Project
 - <http://cydoc.sourceforge.jp/cydocwiki/>
- Cytoscapeに関する日本語情報のポータルサイト
 - (新)Cytoscape J
 - <http://cytoscape.wordpress.com/>
 - (旧)Cytoscape Info
 - <http://cytoscape.seesaa.net/>

情報提供、共有サイト 2 of 2

- 統合TV
 - Cytoscape を使い倒す！～基本操作編～
 - <http://togotv.dbcls.jp/20110603.html>
 - Cytoscape を使い倒す！～応用発展編～
 - <http://togotv.dbcls.jp/20110630.html>
- ボランティアによる日本語チュートリアル
 - <http://wiki.livedoor.jp/bioinformatics/d/Cytoscape/Tutorials> (日本語)

謝辞

- ・本資料を作成するに当たり、Cytoscape開発者の大野圭一朗氏(UCSD)からご助言、最新の情報をいただきました。感謝申し上げます。
- ・また、本資料は、National Resource for Network Biology(NRNB) Showcaseの Introduction to Cytoscape (<http://nrnb.org/showcase-intro.html>) および、Basic Expression Analysis in Cytoscape (<http://nrnb.org/showcase-expression.html>) 他を参考に作成しました。