Integration of biological networks and gene expression data using Cytoscape

Melissa S Cline[1,2], Michael Smoot[3], Ethan Cerami[4], Allan Kuchinsky[5], Nerius Landys[3], Chris Workman[6], Rowan Christmas[7], Iliana Avila-Campilo[7,8], Michael Creech[9], Benjamin Gross[4], Kristina Hanspers[10], Ruth Isserlin[11,12], Ryan Kelley[3], Sarah Killcoyne[7], Samad Lotia[3], Steven Maere[13,14], John Morris[15], Keiichiro Ono[3], Vuk Pavlovic[11,12], Alexander R Pico[10], Aditya Vailaya[5,16], Peng-Liang Wang[3], Annette Adler[5], Bruce R Conklin[10], Leroy Hood[7], Martin Kuiper[13,14], Chris Sander[4], Ilya Schmulevich[7], Benno Schwikowski[1], Guy J Warner[17], Trey Ideker[3] & Gary D Bader[11,12]

[1]Institut Pasteur, 25-28 rue du Docteur Roux, 75724 Paris cedex 15, France. [2]Department of Molecular, Cell and Developmental Biology, University of California, Santa Cruz, California 95064, USA. [3]Department of Bioengineering, 9500 Gilman Drive, Mail Code 0412, La Jolla, California 92093-0412, USA. [4]Computational Biology Center, Memorial Sloan-Kettering Cancer Center, 1275 York Avenue, Box 460, New York, New York 10021, USA. [5]Agilent Laboratories, 3500 Deer Creek Road, MS 26U-16, Palo Alto, California 94304, USA. [6]BioCentrum-DTU, Technical University of Denmark, Building 208, DK-2800 Lyngby, Denmark. [7]The Institute for Systems Biology, Seattle, Washington 98103, USA. [8]Rosetta Inpharmatics LLC, Merck & Co. Inc., 401 Terry Ave N, Seattle, Washington 98109, USA. [9]Blue Oak Software, 1734 Austin Ave, Los Altos, California 94024-6103, USA. [10]Gladstone Institute of Cardiovascular Disease, 1650 Owens Street, San Francisco, California 94158, USA. [11]Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, 160 College Street, Toronto, Ontario, Canada M5S 3E1. [12]Banting and Best Department of Medical Research, University of Toronto, 112 College Street, Toronto, Ontario, Canada M5G 1L6. [13]Department of Plant Systems Biology, VIB, Technologiepark 927, 9052 Ghent, Belgium. [14]Department of Molecular Genetics, Ghent University, Technologiepark 927, 9052 Ghent, Belgium. [15]Department of Biopharmaceutical Sciences, University of California, San Francisco, 1700 Fourth Street, San Francisco, California 94143-2250, USA. [16]Retrevo Inc., 440 N Wolfe Road, Sunnyvale, California 94085, USA. [17]Unilever Safety and Environmental Assurance Centre, Colworth Park, Sharnbrook, Bedfordshire MK44 1LQ, UK. Correspondence should be addressed to T.I. (trey@bioeng.ucsd.edu).

Published online 27 September 2007; doi:10.1038/nprot.2007.324

Cytoscape is a free software package for visualizing, modeling and analyzing molecular and genetic interaction networks. This protocol explains how to use Cytoscape to analyze the results of mRNA expression profiling, and other functional genomics and proteomics experiments, in the context of an interaction network obtained for genes of interest. Five major steps are described: (i) obtaining a gene or protein network, (ii) displaying the network using layout algorithms, (iii) integrating with gene expression and other functional attributes, (iv) identifying putative complexes and functional modules and (v) identifying enriched Gene Ontology annotations in the network. These steps provide a broad sample of the types of analyses performed by Cytoscape.

INTRODUCTION

Functional genomic and proteomic techniques enable routine measurement of expression profiles and functional interactions from the cells and tissues of many different organisms[1–4]. These measurements have significant potential to map cellular processes and their dynamics, given the appropriate computer software to filter and interpret the resulting large amount of data. Commonly used expression analysis methods identify active biological processes from expression profiles by finding enriched gene annotation terms in the lists of differentially expressed genes[5–8]. By combining expression profiles with cellular network information, including protein–protein and protein–DNA interactions, we can begin to explain the control mechanisms underlying the observed changes in activity of a biological process. For instance we can identify a transcription factor known to regulate a set of affected genes.

An important benefit of integrating expression and network data is biologically relevant signals supported by both data types are more likely to be correct than those supported from either data source alone. This is important because expression profiles can be noisy and difficult to reproduce when expression levels are low[9], while protein interaction assays are known to contain false positives and negatives. For instance, it is estimated that up to 50% of unfiltered yeast two-hybrid data are spurious[10], although this is improving as experimental protocols and automated reliability measures that combine multiple data sets of a given type evolve[11,12].

Many sources of expression profiles and cellular networks exist. The Gene Expression Omnibus[13] and ArrayExpress[14] are both large public repositories of gene expression profiles. Protein–protein interactions mapped either by focused studies or by high-throughput techniques are increasingly available in public repositories such as IntAct[15], HPRD[16] and MINT[17] (as reviewed in ref. 18). Protein–DNA interactions mapped at the genome scale using ChIP-Chip and ChIPSeq technology[19] provide potential links between transcription factors and their regulated genes. When information is not available in databases but is in the literature, text-mining techniques can extract functional relationships between recognized genes that, while not always accurate, are useful for analysis in aggregate[20]. In these networks, two genes are linked if they are frequently mentioned in the same sentence[21]. This link may indicate a biochemical association, such as catalysis, or a genetic, colocalization or coexpression relationship. Literature association networks are also useful as a general literature search tool, since each link is tied to the supporting publication. These public data repositories are growing rapidly as the underlying measurement technology improves. For example, the HPRD repository more than doubled in size between 2003 and 2005[22].

A number of software tools are available for network visualization and analysis, including Osprey[23], VisANT[24], CellDesigner[25] GenMAPP[26], PIANA[27], ProViz[28] BioLayout[29], PATIKA[30] and Cytoscape[31]. Each tool has a distinct set of features, which are highlighted in **Table 1**. Here, we describe the application of Cytoscape within a workflow for integration of functional genomics data with biological networks.

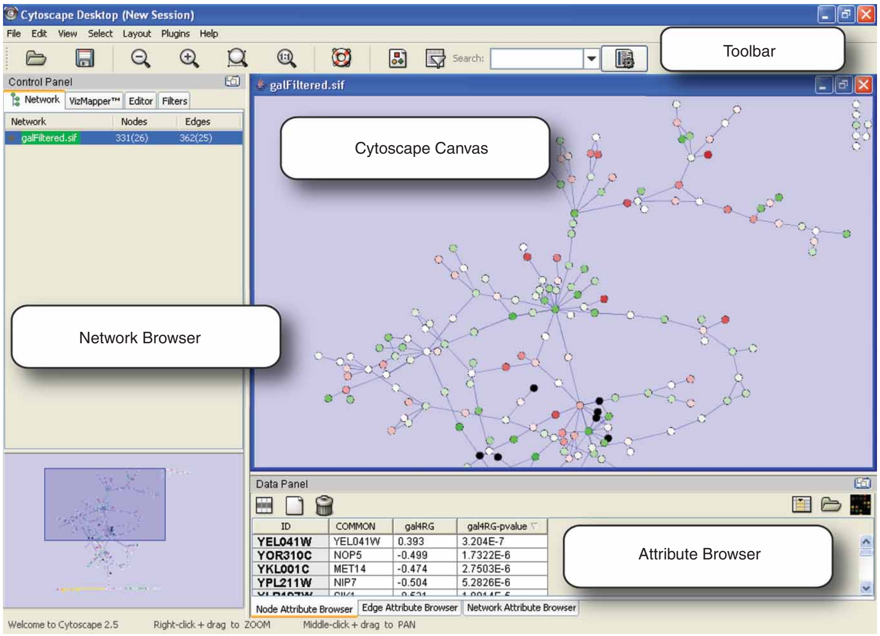
TABLE 1 | Comparison of network analysis platforms.

Feature	CY	GM	VA	OS	CD	AR	IN	GG	PI	PR	BL	PA
Free for academic use	√	√	√	√	√				√	√	√	√
Free for commercial use	√	√	√		√				√	√	√	
Open source	√	√							√	√	√	
Curated pathway/network content		√		√		√	√	√				
Standard file format support	√		√		√				√	√		√
User-defined networks/pathways	√	√	√	√	√	√	√	√	√	√	√	√
Functionality to infer new pathways	√		√			√		√	√			
GO/pathway enrichment analysis	√	√	√				√	√				
Automated graph layout	√		√	√	√	√	√	√		√	√	√
Complex criteria for visual properties		√				√	√	√		√	√	√
Multiple visual styles	√		√	√		√	√			√		
Advanced node selection	√		√	√		√	√	√	√	√	√	√
Customizable gene/protein database		√	√			√		√	√			
Rich graphical annotation		√	√				√	√				√
Statistical network analysis	√		√				√	√	√		√	
Extensible functionality: plugins or API	√		√		√	√	√	√	√			
Quantitative pathway simulation					√	√						

CY, Cytoscape[31]; GM, GenMAPP[26]; VA, VisANT[24]; OS, Osprey[23] (http://biodata.mshri.on.ca/osprey/); CD, CellDesigner[25]; AR, Ariadne Genomics Pathway Studio; IN, Ingenuity Pathways Analysis; GG, GeneGo; PI, PIANA (http://sbi.imim.es/piana/); PR, ProViz (http://cbi.labri.fr/eng/proviz.htm); BL, BioLayout; PA, PATIKA.

Cytoscape is freely distributed under the open-source GNU Lesser General Public License, which allows any use of the software, including feature extension by programming (http://www.gnu.org/licenses/lgpl.html). In Cytoscape nodes representing biological entities, such as proteins or genes, are connected with edges representing pairwise interactions, such as experimentally determined protein–protein interactions (**Fig. 1**). Nodes and edges can have associated data attributes describing properties of the protein or interaction. A key feature of Cytoscape is its ability to set visual aspects of nodes and edges, such as shape, color and size, based on attribute values. This data-to-visual attribute mapping allows biologists to synoptically view multiple types of data in a network context. Additionally, Cytoscape allows users to extend its functionality by creating or downloading additional software modules known as 'plugins'. These plugins provide additional functionality in areas such as network data query and download services[32–35]; network data integration and filtering[12]; attribute-directed network layout[36,37]; Gene Ontology (GO) enrichment analysis[7]; and network motif[38,39], functional module[40–42], protein complex[43] or domain interaction detection[44]. Links to these plugins can be found at http://www.cytoscape.org. Altogether, Cytoscape and its plugins provide a powerful tool kit designed to help researchers answer specific biological questions using large amounts of cellular network and molecular profiling information.

This protocol is modular in its organization, and the five modules can be followed sequentially or as stand-alone protocols (the modular organization is shown schematically in **Fig. 2**). The first module, 'Obtain network data', describes methods to build networks for genes of interest by querying protein interaction databases and text-mining data sources. 'Explore network and generate layout' introduces basic aspects of Cytoscape operation, including network navigation and layout. 'Annotate with attribute and expression data' shows how to link expression profile data to



Network	Nodes	Edges
galFiltered.sif	331(26)	362(25)

ID	COMMON	gal4RG	gal4RG-pvalue
YEL041W	YEL041W	0.393	3.204E-7
YOR310C	NOP5	-0.499	1.7322E-6
YKL001C	MET14	-0.474	2.7503E-6
YPL211W	NIP7	-0.504	5.2826E-6

Welcome to Cytoscape 2.5   Right-click + drag to ZOOM   Middle-click + drag to PAN

Figure 1 | The Cytoscape Desktop. The *Cytoscape canvas* displays network data. The *toolbar* (top) contains the command buttons. The name of each command button is shown when the mouse pointer hovers over it. The *Control Panel* (left) displays the *Network tree viewer,* which lists the available networks by name and size. The *Network Overview Pane* (lower left) shows the current network in white and the displayed portion in blue. Finally, the *Data Panel* (lower right) can be used to display node, edge and network attribute data. The Cytoscape Desktop shows the sample network and expression data, with nodes colored by expression values from lowest (green) to highest (red).

the network for visualization and analysis. This module uses mRNA expression data as an example, but the steps outlined apply to any form of molecular profile such as protein levels. 'Analyze network features' explains how to perform analytical methods that identify putative functional or structural modules within the network that may, for instance, highlight protein complexes active under a profiled experimental condition. Finally, 'Detect enriched gene functions' illustrates methods to identify enriched gene functions, such as those characteristic of biological processes, in previously identified sets of interesting genes or network regions.

These analysis steps have proven useful in multiple studies, such as analyzing networks of genetic interactions[45–48], gene regulatory events[49,50] and protein–protein interactions[51,52], cellular network organization[53,54] and evolution[55] and determining pathways involved in atherosclerosis[56]. A sample protein network and mRNA expression profile resulting from gene knockouts that perturb galactose metabolism in *S. cerevisiae*[57] is provided to illustrate the protocol.

We now turn to each of the five modules in detail, presenting the rationale for each portion of the protocol and indicating viable alternative techniques.

Obtain network data

This section describes three ways to import network data into Cytoscape.

The first method is to query protein interaction databases such as cPath[33] with a list of genes of interest. cPath queries the IntAct[15] and MINT[17] databases. This is an appropriate method for users who are interested in assessing the connections between genes with significant experimental responses, in well-studied organisms such as *S. cerevisiae* or *Homo sapiens*. Cytoscape users can interrogate additional protein interaction databases via MiMi[34] or BioNetBuilder[32], plugins that are similar in application to cPath. In each of these cases, following the steps in this workflow section yields a network that contains known and putative functional associations between the genes of interest.

The second method is to build a text-mining association network using the Agilent Literature Search plugin[21]. This method is most appropriate for those who are working in organisms that are not well represented in protein interaction databases, or want to restrict the network to associations observed only in specific contexts. For Literature Search the user builds a set of queries by entering terms, such as gene names, and contexts, such as an organism or disease name. The query set is submitted to selected search engines, for example PubMed or OMIM. The resulting documents are fetched, parsed into sentences, and analyzed for known interaction terms, like 'binding' or 'activate'. Agilent Literature Search uses a lexicon set for defining gene names (concepts) and aliases, drawn from Entrez Gene, and interaction terms (verbs) of interest. An association is extracted for every sentence containing at least two concepts and one verb. Associations are then converted into interactions with corresponding sentences and source hyperlinks, and added to a Cytoscape network. Interaction network and text-mining association network data are complementary: protein interaction databases contain experimentally determined interactions; whereas text-mining association networks contain more general association types and offer an alternative network source where interaction data are limited.

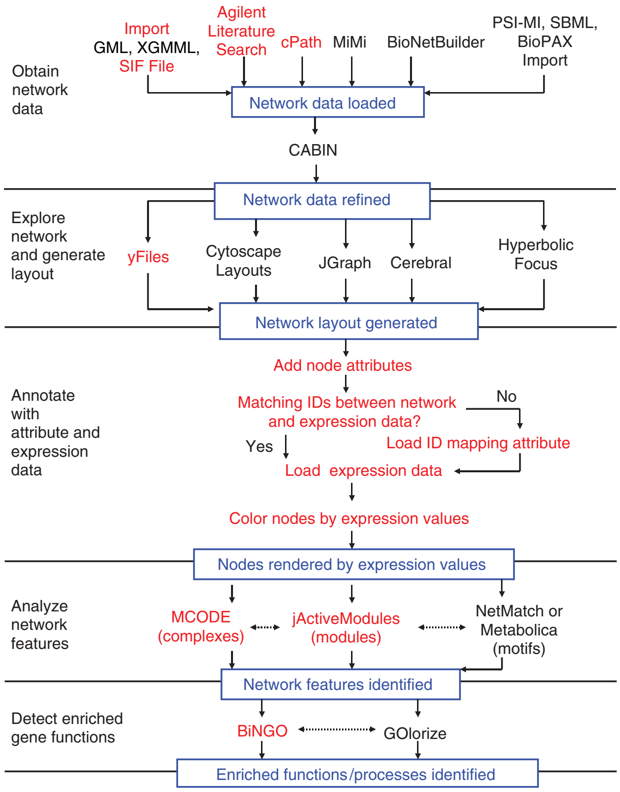

Obtain network data
Import GML, XGMML, SIF File
Agilent Literature Search
cPath
MiMi
BioNetBuilder
PSI-MI, SBML, BioPAX Import
Network data loaded
CABIN
Network data refined
Explore network and generate layout
yFiles
Cytoscape Layouts
JGraph
Cerebral
Hyperbolic Focus
Network layout generated
Annotate with attribute and expression data
Add node attributes
Matching IDs between network and expression data?
No
Yes
Load ID mapping attribute
Load expression data
Color nodes by expression values
Nodes rendered by expression values
Analyze network features
MCODE (complexes)
jActiveModules (modules)
NetMatch or Metabolica (motifs)
Network features identified
Detect enriched gene functions
BiNGO
GOlorize
Enriched functions/processes identified

Figure 2 | Outline of the protocol. The steps in red are included in this protocol, while analyses listed in black represent useful alternatives that achieve related goals. Dotted lines represent steps that can be done in any order, while solid lines represent steps that must be done in sequence. Expression data are not required for finding putative complexes (with MCODE) or for network motifs (with NetMatch or Metabolica), but can aid in interpreting the results by suggesting system dynamics.

The third method described for importing network data in to Cytoscape is to import a network file, such as a SIF (Simple Interaction Format) file. The SIF file format is detailed in **Box 1**. These files are straightforward for a user to create with a standard text editor. SIF file import is the most appropriate method for users who want to focus their analysis on network data identified in advance, such as those who are interested in the impact of the experimental conditions on sets of specific interactions or pathways.

In addition to the import methods described in this protocol, Cytoscape users can also import pathways from repositories such as KEGG[58], Reactome[59] via the PSI-MI, BioPAX, or SBML data exchange formats (as reviewed in ref. 60), although such pathway data contain non-pairwise interactions between molecules. For instance, there is a single interaction between multiple substrates and products in a biochemical reaction, which must be mapped to pairwise interactions in Cytoscape.

After importing interaction data, the user may optionally filter the resulting network to reduce the network size by selecting just the types of network information of interest. For instance, the CABIN plugin[12] enables the user to merge network data from multiple experimental sources and select the interactions

BOX 1 | CYTOSCAPE INPUT AND OUTPUT FILE FORMATS

Cytoscape can import and export data in a variety of formats, from simple delimited text formats to XML and other sophisticated formats for sharing data with other programs. This box provides a brief overview of these formats. For complete information, please refer to the Cytoscape manual at http://www.cytoscape.org/.

Network import and export

The standard file that Cytoscape opens and saves is the Cytoscape Session File (.cys). This file stores all information in your current session including multiple network layouts, attribute values and setting information. Cytoscape can also import network data in the following formats:

- Simple Interaction File (SIF or .sif)
- Cytoscape Node and Edge Attribute File Format (.noa and .eda)
- Graph Markup Language (GML or .gml)
- eXtensible Graph Markup Language (XGMML or .xgmml)
- Systems Biology Markup Language (SBML)
- Biological PAthways eXchange (BioPAX)
- Proteomics Standards Initiative Molecular Interaction (PSI-MI) Level 1 and 2.5
- Delimited Text Table
- Excel Workbook (.xls)

For network export, Cytoscape can output SIF, GML, XGMML and PSI-MI formats. Users may also create an image file of their network data via File → Export → Network view as graphics. This feature includes many standard image formats, including JPEG, PNG and PDF.

SIF

The SIF format is a straightforward format to allow users to define network data with a text editor. Each line of the file contains three or more tokens. The first token is the source node. The second token is the interaction type. This token is an arbitrary text string that describes the interaction between the two nodes. The third token, and all subsequent tokens on the same line, specify the target nodes. A sample file might look like this:

```
nodeA interactionType1 nodeB
nodeA interactionType2 nodeC
NodeF
```

In this network nodeA has an edge to nodeB labeled interactionType1; nodeA has an edge to nodeC labeled interactionType2; nodeF is defined but has no edges. In practice the nodes will be the names of proteins or genes in the network, and the labels given to the interaction type will be some tag that defines that relationship, such as 'protein-protein', 'degrades' or 'phosphorylates'. Because of its basic text format, a SIF file is easily created either manually by a user (e.g., in Excel) or programmatically by a text-processing script. The Supplementary Data contain a sample SIF file, galFiltered.sif (**Supplementary Data 2**).

Other network file formats

The GML format stores information on the network connectivity (like SIF) but also preserves the visual layout and appearance of the network in the Cytoscape view. The XGMML format is the XML extension of the GML format and is generally preferable to GML. These formats are not amenable to human editing. However, once a network is loaded into Cytoscape from a (human-edited) SIF file, the attributes and visual style settings can be modified using the GUI interface. Then the network can be stored along with its visual properties and data attributes using the GML or XGMML format. Cytoscape can also read files in the SBML, BioPAX, and PSI-MI data exchange formats, allowing the use of networks created in other programs.

Attribute file formats

Data attributes on nodes and edges can be imported from delimited text files or Excel spreadsheets via Cytoscape's Table Import functionality. Text files are given the filename suffix .noa for Node Attribute and .eda for Edge Attribute. For a node attribute, a sample .noa file might appear like this:

```
AttributeName
nodeA = value1
nodeB = value2
nodeC = value1
```

The first line is the name of the attribute, for example, 'SubcellularLocation'. Each subsequent line contains the name of the node, an equal to sign and the node's attribute value. Attribute values can be numeric or textual such as '4.12' or 'nucleus'. An edge attribute file (.eda) is formatted similarly, where the edge name is specified by the source node, the interaction value in parentheses and the target node. For example:

```
AttributeName
nodeA (interactionType) nodeB = 0.56
nodeB (interactionType) nodeC = 0.918
nodeB (interactionType) nodeA = 0.3412
```

Each attribute is stored in a separate file.

BOX 1 | CONTINUED

Expression data file format

Cytoscape reads expression data from tab-delimited text files that can be exported from a spreadsheet program or created by the user in a text editor. These files must be renamed to have the extension either '.mrna' or '.pvals' to be recognized by Cytoscape. The choice of extension does not matter between these two. The data are organized as a matrix, with each row representing the expression results for one gene/protein in the network (**Fig. 3**). The first row provides column labels. The first column holds the gene/protein identifier, while the second column contains arbitrary text, such as a descriptive annotation. The subsequent columns contain expression data: one experiment per column, with the experiment names provided in the first row. If the expression data consist of one measure per gene per experiment, then there is one column per experiment. If the data also contain *P*-values or other significance measures, as generated by expression analysis packages such as RMA[73], then each experiment is represented with two columns: the first is assumed to contain the expression measure, the second contains the significance measure, and the two columns must have exactly the same label:

Gene	Label	Experiment1	Experiment2	Experiment1	Experiment2
geneA	labelA	valueA_1	valueA_2	pvalueA_1	pvalueA_2
geneB	labelB	valueB_1	valueB_2	pvalueB_1	pvalueB_2
geneC	labelC	valueC_1	valueC_2	pvalueC_1	pvalueC_2

GeneA identifies a gene, labelA provides a descriptive name, valueA_1 contains its expression level as measured in Experiment1, valueA_2 contains its expression level as measured in Experiment2, and pvalueA_1 and pvalueA_2 contain *P*-values (or other measures of significance of differential expression) for Experiment1 and Experiment2, respectively. The *P*-value columns are optional for most Cytoscape functionality. However, to identify active modules with the jActiveModules plugin, the expression data must contain *P*-values of significance ranging between 0 (most significant) and 1 (least significant). The Supplementary Data contain a sample expression data file, galExpData.pvals (**Supplementary Data 1**). **Figure 3** illustrates the first lines of this file.

observed multiple times, which are likely more reliable than those observed once.

Explore network and generate layout

This section introduces the user to basic aspects of Cytoscape operation, including network navigation and layout. Apart from basic operation, basic filter features are available to identify genes as important by virtue of their high number of connections in the network[61]. Essential genes are also associated with nodes, which occupy central positions in large interaction networks, particularly those frequently found to connect other nodes together[62]. This section uses the yFiles layout algorithm, although many other layout algorithms exist, some of which support more specialized operations.

Annotate with attribute and expression data

Expression profiles can provide powerful insights into cellular state and dynamics when integrated with network data. For example, if two nodes consistently show similar changes in expression levels, or they consistently show changes in opposite directions, then one gene might regulate the production of the other. Also, a group of connected nodes characterized by large fold changes may represent a signal-transduction cascade or protein complex that is repressed or induced under the experimental condition[63]. This section shows how to apply mRNA expression data to a network as an example, but the steps outlined apply to any form of molecular profile data.

Analyze network features

This section explains how to run analysis methods that identify putative functional or structural modules within the network that may highlight protein complexes active under a profiled experimental condition as complex topologies of interaction networks make this difficult to do by eye. The jActiveModules plugin[42] automates this analysis and identifies connected sections of the network in which the nodes have significant *P*-values. This indicates a group of nodes that may be coregulated, suggesting a module whose activity is influenced by the experimental context of the expression data. A *complex* is a module in which the macromolecules involved form a structure to execute some function called a molecular machine. Densely connected regions in protein interaction networks tend to correspond to protein complexes. The MCODE plugin[43] identifies putative complexes by finding regions of significant local density. At this stage, the user could also identify over-represented network motifs with the Metabolica or NetMatch plugins[39], although these alternatives are not covered here.

Detect enriched gene functions

The function of interesting gene sets or network regions can be summarized by finding significantly enriched functional annotation terms. This may be used to support module or complex predictions. This section outlines steps required to identify enriched GO processes with the BiNGO plugin[7]. Users can optionally use the GOlorize plugin[36] to refine the network layout according to selected GO classes.

Cytoscape can be used in many additional visualization and analysis workflows, including published protocols on the reverse engineering of regulatory[64] and metabolic[65] networks, and can be extended using Java programming to implement new features and analysis methods. Over 40 excellent plugins contributed by many software development groups provide examples. While these topics are not covered here, Cytoscape has a community of software developers and users who can answer questions about potential new Cytoscape uses, as detailed at http://www.cytoscape.org.

MATERIALS

EQUIPMENT

- Internet access

EQUIPMENT SETUP

Hardware requirements Cytoscape hardware requirements depend on the size of the networks to be imported and analyzed. For networks up to 5,000 edges, we recommend a 1 GHz CPU or higher, a high-end graphics card, 60 MB of available hard disk space, at least 512 MB of free physical RAM and a minimum screen resolution of 1,024 × 768. At least 1 GB of RAM is required for larger networks.

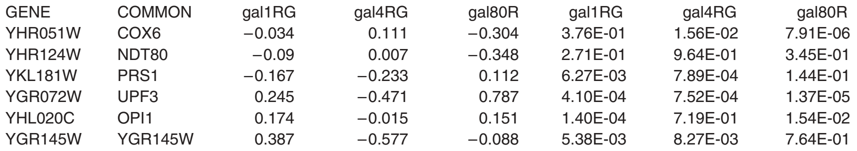
Java 2 platform Standard Edition, version 5.0 or higher (http://java.sun.com/javase/downloads/index.jsp), which includes Java Web Start (http://java.sun.com/products/javawebstart/).

A three-button mouse This is recommended (but not required) for Mac users as an aid in network navigation.

Data files This protocol begins with an expression data set to be analyzed. While Cytoscape has direct links to several previously published expression data sets via the Internet, loading new data into Cytoscape requires assembling a tab-delimited file as detailed in **Box 1**. A sample expression file is made available in **Supplementary Data 1** (galExpData.pvals; see **Fig. 3**). In addition to expression data, several additional data files are available for readers wishing to follow this protocol as a tutorial:

- **galFiltered.sif** contains a protein–protein and protein–DNA interaction network pertaining to the galactose-utilization pathway in yeast, as previously published[57]. This file illustrates the SIF, which offers a straightforward means to import networks into Cytoscape as text. Because this is a yeast data set, genes/proteins are identified by their systematic Open Reading Frame locus tags.
- **galGeneNames.csv** provides common gene names corresponding to the Open Reading Frame gene names in galFiltered.sif. This file also provides an example of how Cytoscape can import attribute data from delimited text files.
- **SampleData.cys** is a Cytoscape session file containing the network data from galFiltered.sif along with a second network constructed for the same genes using the Agilent Literature Search plug-in to query literature associations, as described below.

Cytoscape For simplest operation of this protocol, Cytoscape can be executed over the web via Java Web Start by navigating to http://www.cytoscape.org/nature.protocols/ and clicking on the web start link. See http://java.sun.com/products/javawebstart/ for more information on running Java Web Start applications. Alternatively, Cytoscape can be installed on a local computer by following the steps in **Box 2**.



GENE	COMMON	gal1RG	gal4RG	gal80R	gal1RG	gal4RG	gal80R
YHR051W	COX6	−0.034	0.111	−0.304	3.76E-01	1.56E-02	7.91E-06
YHR124W	NDT80	−0.09	0.007	−0.348	2.71E-01	9.64E-01	3.45E-01
YKL181W	PRS1	−0.167	−0.233	0.112	6.27E-03	7.89E-04	1.44E-01
YGR072W	UPF3	0.245	−0.471	0.787	4.10E-04	7.52E-04	1.37E-05
YHL020C	OPI1	0.174	−0.015	0.151	1.40E-04	7.19E-01	1.54E-02
YGR145W	YGR145W	0.387	−0.577	−0.088	5.38E-03	8.27E-03	7.64E-01

Figure 3 | Sample expression input file, showing genes and associated columns of expression data for multiple experiments.

PROCEDURE

Obtain network data

1| Start Cytoscape. For webstart users, go to http://www.cytoscape.org/nature.protocols/ and click on the indicated link to launch Cytoscape. This will automatically download Cytoscape to the local computer and start it. For users with Cytoscape already installed locally, navigate to the Cytoscape home directory and execute cytoscape.bat (Windows users) or cytoscape.sh (Linux and Mac OS X users).

2| After Cytoscape has started, you should see a Cytoscape Desktop window, as shown in **Figure 1**. (a) The main *Cytoscape canvas* displays network data and is initially blank. (b) The toolbar, at the top of the Cytoscape desktop, contains the command buttons. A description of each command button is shown when the mouse pointer hovers over it. (c) The *Control Panel*, at the left, displays the *Network tree viewer*. This lists the available networks by name and number of nodes and edges. (d) The *Network Overview Pane*, at the lower left, shows the current network in white with the displayed portion overlaid in blue. (e) The *Data Panel,* at the lower right, can be used to display node, edge and network attribute data.

3| Identify a list of genes of interest. Typically, these will be genes with a pronounced response to your experimental conditions of interest. This protocol describes three options for importing network data related to these genes into Cytoscape: querying interaction databases (option A), building an association network through text mining (option B) and loading your own network data from a text file (option C). Continue with one or more of these steps based on which is most appropriate for your use case. Users wishing to follow this protocol as a tutorial should download the **Supplementary Data 2** file (galFiltered.sif) and continue with Step 6. The steps in this protocol are most effective with a network of at least

BOX 2 | CYTOSCAPE INSTALLATION

1. Go to http://www.cytoscape.org/, click on the Download Cytoscape link (fill out the form) and accept the terms of the LGPL license.
2. Download the platform-specific installation bundle for your operating system. After downloading, execute the installation bundle to install Cytoscape on the local computer. As you do so, make careful note of the directory in which Cytoscape is installed: the *Cytoscape home directory*.
3. Launch Cytoscape by navigating to the Cytoscape home directory and executing cytoscape.bat (under Microsoft Windows) or Cytoscape.sh (under Max OS X and Linux).

Install the plugins required for this protocol

4. Go to the Plugins menu and then Manage Plugins to activate the Manage Plugins window. Click on Available for Install.
5. Expand Network Inference Plugins, click on Agilent Literature Search 2.52 (or later) and click the Install button. This will bring up the Plugin License Agreement window. Click Accept and then Finish.
6. Repeat with BiNGO 2.0 (or later) under Functional Enrichment, and jActiveModules 2.2 (or later) and MCODE 1.2 (or later) under Analysis. Installing BiNGO may take several minutes; this behavior is normal owing to the large GO database, which is included in the package.
7. Close the Manage Plugins window. Return to the Plugins menu to verify that the four plugins are now listed. After installing the plugins once, they will remain installed for subsequent executions of Cytoscape.

For more information, refer to the Cytoscape manual by navigating to http://www.cytoscape.org and following the link to the manual.

250 interactions. To obtain such a network, start with a set of at least 25 genes and add more genes or import more interactions as needed to obtain a network of the desired size.

(A) Obtain protein interaction data with cPath

(i) cPath[33] enables users to retrieve protein interactions from multiple repositories, including IntAct[15] and MINT[17]. Invoke cPath by selecting the File menu, then New → Network → Construct network using cPath...

(ii) Enter the names of one or more genes of interest in the input box at the upper left of the cPath Plugin dialog box (e.g., **Supplementary Data 3**).

▲ CRITICAL STEP New users who are becoming familiar with this functionality may wish to enter just one gene name at a time, whereas experienced users may wish to enter a list of genes, such as the contents of the **Supplementary Data 3** file (sample.input.genes.txt).

(iii) Select the desired species in the species pull-down menu, which is set to All Organisms by default. For the gene list in sample.input.genes.txt, select *Saccharomyces cerevisiae*.

(iv) The maximum number of records is set to Limit to 10 by default. Change this to Limit to 500. While the default setting is useful for exploratory queries with a single gene of interest, one must typically retrieve a larger number of records to achieve connectivity between a set of genes of interest. Many database records contain more than one interaction of a protein, so the number of interactions retrieved may be greater than the limit. To obtain all interactions for this gene set, select No Limit, remembering that a higher limit will take a longer time to download.

(v) Click the Search button. Shortly, the Cytoscape canvas will show a protein interaction network with proteins (nodes) arranged in a grid, connected by retrieved interactions (edges). By default, Cytoscape displays networks with 10,000 or fewer nodes; for larger networks, the user may request a view by performing a Right-Click on the network label in the *Network tree viewer* and selecting Create View from the pop-up menu.

(vi) Notice the entry for your network in the Network tree viewer, and notice the number of nodes and edges.

(B) Generate an association network

(i) Under the Plugins menu, select Agilent Literature Search.

▲ CRITICAL STEP The first time this plugin is executed, the user will be asked to accept a license agreement. Next, the Agilent Literature Search window should appear.

(ii) In the Terms panel at the upper left of this window, enter the names of up to 100 genes or proteins from your list of genes of interest.

▲ CRITICAL STEP For best results, use standard HUGO gene symbols[66] for human searches or other appropriate official gene symbols. The search will take at least 3 s per gene, to comply with PubMed limits on usage, so we recommend starting with a shorter list (ten or fewer genes) for exploratory work. Notice that the Query Editor panel echoes the gene list. Each line shown in this panel represents one query that will be sent to PubMed.

(iii) Under Extraction Controls, go to the Concept Lexicon pull-down menu and select your species. For the sample data, select *Saccharomyces cerevisiae*. Leave the Interaction Lexicon set at limited; this specifies that the sentences chosen for text extraction are more likely to be related to interactions, and this is the recommended setting except for cases where the literature is sparse.

(iv) Increase the number of publications retrieved per query by adjusting the value in the Max Engine Matches: field. Up to 1,000 queries may be issued in total, to comply with PubMed limits on usage. For the sample data, Max Engine Matches can be increased to 50. In general, requesting more publications per search term is more effective than including more search terms.

(v) Click on the Use Aliases button. Notice that the queries in the Query Editor panel change to search for common aliases of specified gene or protein names, as well as searching for the specified name.

(vi) Refine your search by going to the Context panel and entering additional descriptors, such as tissue type or disease, and clicking the Use Context button. For the sample data, enter the descriptor 'galactose'. If a descriptor is more than one word, enclose it in quotes (e.g., 'transcriptional regulation'). Notice that the Query Editor shows each of your search terms connected with your context term with an AND (e.g., '((ylr044c OR pdc1)) AND galactose'). This will limit the PubMed search to publications relating to both the search terms and the context term.

(vii) Click on the forward arrow to begin the search. When the search has completed, the Query Matches panel will show a list of articles containing the search terms in their abstracts. The Cytoscape desktop will display a new network in which each node represents a gene or protein and each edge represents an association found in the text of the selected articles. These associations are extracted through sentences characterized by two or more gene names and an interaction-related verb such as 'degrade', 'inhibit' or 'methylate'.

(viii) Examine the list of articles shown in the Query Matches panel. Remove articles as desired by right-clicking on the link and selecting Delete Match. This will delete the article from the list, along with the corresponding nodes and edges from the network, if they are not supported by any other article.

(ix) The Show Sentences feature enables the user to view literature references. To do this (i) select one or more node or edge in the network by clicking on them with the mouse and (ii) on the Select menu, select the Evidence from Literature → Show Sentences from the Literature.

This brings up a window with the list of sentences that were extracted from the literature, as pertaining to the selected nodes or edges, showing the key search terms in bold. If an alias matched, then the formal name is shown in square brackets. If the user has selected multiple nodes or edges, then this list will contain the sentences for all the nodes or edges selected. Mouse-clicking on any of these sentences will bring up its abstract in the user's default web browser. Users can delete sentences from the list by right-clicking on the sentence and then clicking on the Delete Sentence button. If all sentences that support a particular edge are deleted, then that edge is removed from the network.

(x) Users can extend the network around any node by right-clicking on the node and then clicking on Select → Evidence from Literature → Extend Network from the Literature. This will issue a new search request for the specified node. If this new search yields any new nodes, edges or sentences, they will be added to the existing network. This is most effective for nodes that were not included in the original set of search terms.

(C) Import network data from a text file

(i) Assemble your data into a SIF file, as described in **Box 1**.

(ii) Under the File menu, select Import → Network (multiple file types). Specify the name of the SIF file and click OK.

Explore the network and generate a layout

4| Reorganize the network by going to the Layout menu and selecting yFiles → Organic. This algorithm implements a variant of the force-directed paradigm[67] in which nodes are modeled as objects with mutually repulsive forces, edges induce attractive forces between the nodes they connect and nodes are placed such that the sum of these forces is minimized. This layout has the effect of exposing structure inherent in the network. In particular, it facilitates identification of clusters of tightly connected nodes, which suggest functional modules, and 'hub nodes', which are involved in many interactions and often represent functionally important proteins. Note that Cytoscape provides a wide variety of alternatives for layout, including hierarchical layouts, circular layouts and attribute-based layouts. Experiment with several of these.

5| To generate more space for the network canvas, 'float' the Control Panel and Data Panel by clicking the *Float Window control*, the icon at the top right corner of each panel in the title bar. This detaches the panels as separate windows. Resize the network canvas by dragging its lower right corner.

6| Select any node by clicking it with the left mouse button. Notice that the selected node turns yellow. Select several nodes by holding down and dragging the left mouse button to define a rectangular selection region.

7| Notice that the Data Panel lists the identifiers of all nodes selected. This browser can also display node attributes. Nodes obtained from cPath have attributes including 'FULL_NAME', a descriptive protein name. Nodes from the Agilent Literature Search plugin have attributes including 'NbrConnections', the degree of the node.

8| Select one of these attributes for display by clicking on the Select Attributes button, the leftmost button of the Data Panel toolbar. This will display the list of available attributes, with selected attributes highlighted. Select or deselect an attribute by clicking on its name. Exit by clicking the right mouse button.

9| Select one or more edges by dragging a rectangular region across them. Edges obtained from cPath have attributes including 'EXPERIMENTAL_SYSTEM_NAME', the type of experimental evidence supporting the interaction. Edges obtained from the Agilent Literature Search plugin have attributes including NbrSentences, the number of distinct sentences that support the association.

10| View these attributes by clicking on the Edge Attribute Browser tab at the bottom of the Data Panel and selecting attributes for display as described in Step 8.

11| Zoom into the network using the Zoom In button on the toolbar, depicted as a magnifying glass with a plus sign ('+'). Notice that in the Control Panel, the blue box in the Network Overview window shrinks to depict the portion of the network displayed. Move this box to navigate the network. Other controls include the Zoom Out button (to the left of Zoom In) and the Zoom Selected Region button (to the right of Zoom In). Alternatively, zooming can be accomplished with a right-click drag in the main Cytoscape canvas.

12| Identify the ID name of any node in the network, such as 'YPL248C' for the sample network in galFiltered.sif. Search for that node ID by going to the Search: field at the right of the Cytoscape toolbar and typing the name. After each character is typed, a drop-down menu displays the list of node IDs matching the portion of the name entered. This field can be configured to search any attribute (not just ID) by clicking the 'Configure search options' button to the right.

13| Cytoscape stores user data as a *session file*, containing all network, layout and attribute data in the program's memory. Save the current session by clicking on the Save Current Session As… button on the toolbar (appearing as a floppy disk) or by going to Save or Save As… under the File menu. Once a session is saved, it can be reloaded with the Open option under the File menu. A sample session file is provided under **Supplementary Data 4**, SampleData.cys, containing a SIF file imported in Step 3C and an association network for the same genes from Step 3B.

Annotate with attribute and expression data

14| Optional step: Additional node or edge attribute data can be integrated with your network by following the steps in **Box 3** (and see **Supplementary Data 5**).

15| The expression data should be formatted as described in **Box 1**. For users wishing to follow this protocol as a tutorial, the Supplementary Data section includes a sample expression data file (**Supplementary Data 1**, galExpData.pvals); this file can be viewed with a standard text editor to reveal the file format. If the identifiers in the first column of your expression file match the node names in the Cytoscape network *exactly* in a case-sensitive match, then skip directly to Step 16 (this applies to users of the sample data). Otherwise, follow the procedure described in **Box 3** to load an attribute file that describes the correspondence between the gene/protein ID names in the expression data file and the ID names of nodes in your network. **Box 4** provides further details of one method to build such an attribute file.

BOX 3 | ADDING NODE OR EDGE ATTRIBUTE DATA

Cytoscape provides many mechanisms for integrating heterogeneous forms of data and for assessing the influence of each data type within a network context. Additional node or edge attribute data can be integrated with your network by following the steps below. Users who are following this protocol as a tutorial should download the **Supplementary Data 5** file (galGeneNames.csv) and follow the steps below to import common gene names (synonyms) for nodes in the network galFiltered.sif.

1. Organize your attribute data into a delimited text file or Excel spreadsheet, in which one of the columns matches the node or edge IDs in your network *exactly* in a case-sensitive match.
2. Go to the File menu, then Import and then Attribute from Table (Text/MS Excel). This will bring up the Import Annotation File window.
3. Click the Select File button and specify your attribute file.
4. If importing an Excel spreadsheet, continue to Step 5. Otherwise, (a) select Show Text File Import Options and (b) select the appropriate delimiter for your file. If the first line of your file contains column headings, select Transfer first line as attribute names.
5. Click on the Import button.
6. Verify that the new attribute has been loaded by returning to the Node Attribute Browser and clicking the Select Attribute button.

Load expression data

16| Load the expression data file by going to the File menu, then Import and then Attribute/Expression Matrix. Select your expression data file in the dialog box. If an ID-mapping attribute was loaded in Step 15, then specify this attribute for expression data import by using the control labeled Assign values to nodes using…. For the sample data, specify the Supplementary Data file galExpData.pvals (**Supplementary Data 1**) and assign values to nodes using the attribute ID. Click on Import.

17| Check the list of node attributes in the Node Attribute Browser. There should be two new sets of attributes per experiment, named 'Xexp' and 'Xsig' where 'X' is the name of your experiment. The expression values will be contained in 'Xexp'. If your expression data set contains *P*-values (see **Fig. 3** and **Box 1**), they will be contained in 'Xsig'; otherwise, 'Xsig' will be blank. Users of the sample data should see attributes for experiments 'gal1RG', 'gal4RG' and 'gal80R': expression fold changes in 'gal1RGexp', 'gal4RGexp' and 'gal80Rexp'; and *P*-values in 'gal1RGsig', 'gal4RGsig' and 'gal80Rsig'. Verify the expression data import by selecting attributes from one or more experiments for display and selecting one or more nodes on the Cytoscape canvas.

Setting visual node properties

18| The Cytoscape *VizMapper* controls the mapping of node and edge attributes to visual properties. Open the VizMapper by navigating to View in the Cytoscape menu and then Open VizMapper. This will bring up the VizMapper in the Control Panel, as shown in **Figure 4a**. Note that if you have not done so already, you can expand the width of the Control Panel by 'floating' it into a separate window, as described in Step 5, and dragging on the resize control at the lower right corner of the window. The VizMapper contains a two-column list, showing the available visual properties at the left and their current settings at the right. Expanding the window may give a better view of the settings.

Color nodes with red-to-green gradient

19| The first step to color your nodes according to expression values is to locate the Node Color property, which is grayed-out initially. Double click on it.

20| Next to Node Color, you will see a message reading 'Please Select an Attribute!'. Click on this message. This will bring up a list of node attributes. Scroll down to your expression data value. For sample data users, this will be 'gal4Rexp'.

21| Underneath Node Color, locate the entry for Mapping Type and click in the right column, directly underneath your expression data attribute. A pull-down menu will appear. Select Continuous Mapping.

22| Underneath Mapping Type, you will see the label Graphical View, located next to an empty rectangle, as shown in **Figure 4a**. Click on the rectangle. This will bring up the window Gradient Editor for Node Color, which shows a bar corresponding to the range of your expression data attribute, with the Min and Max values displayed at the two ends.

23| In the Gradient Editor for Node Color window, click the Add button *twice*. You will see the range bar for your expression data attribute divided in half, with the lower half colored black and the upper half colored white. See **Figure 4b**.

24| Double-click the white downward-facing triangle in the center of the range to bring up a color selection window and select Green. You should now see your range bar divided in half, with the left half colored black and the right half showing a color gradient from green to white. See **Figure 4c**.

25| Slide your green triangle toward the left end of the range bar, leaving a short black portion at the left end of the range bar. See **Figure 4d**. On the Cytoscape canvas, you should see some nodes colored black (those with the lowest expression values), most nodes colored in various shades of darker and lighter green and some nodes colored white (those with the highest expression values).

26| Click the Add button one more time. Locate one of the downward-facing triangles at the right end of the bar and slide it to the center of the range. See **Figure 4e**. There should now be triangles at left, center and right.

27| Double-click the downward-facing triangle at the right end of the bar to bring up a color selection window. Select Red. This will create a gradient of white to red at the right side of the range. See **Figure 4f**. On the Cytoscape canvas, you should see nodes with higher expression values colored pink or red, with darker colors indicating higher expression values.

28| Slide the red triangle a short distance to the left. Double-click on the white right-facing triangle at the right end of the bar and select Blue. See **Figure 4g**.

29| On your Cytoscape canvas (**Fig. 4h**), you should now see nodes colored according to a red–green continuum: (i) Nodes with expression values at the lower end of the range are colored more green, those with expression values at the higher end of the range are colored more red and those in the middle of the range are colored more white. (ii) Lighter colors indicate expression data values near the middle of the range, whereas darker colors indicate values near the ends of the range. (iii) Notice that as you slide the triangles in the range bar, the colors of nodes on the Cytoscape canvas change accordingly. (iv) Close the Gradient Editor window.

30| By default, nodes are colored pink. Consequently, a pink node color could imply an expression data value somewhat higher than the middle of the range, or it could imply that there were no expression data for the node. To remove this ambiguity, change the default node color to gray, as follows: (i) Under the VizMapper, locate the Defaults panel, showing the default node

BOX 4 | ASSEMBLING IDENTIFIER (ID) MAPPING ATTRIBUTES

To import attribute files or expression data into Cytoscape, the gene or protein identifier in the file must exactly match the corresponding Cytoscape node ID (or other Cytoscape attribute that has been previously loaded). If no matching identifiers are present, the situation can be corrected by loading an additional identifier into Cytoscape as a new node attribute. This section describes one method for constructing and loading an identifier-mapping attribute, using an external ID mapping service called Synergizer, which is provided by the Roth laboratory at Harvard University. Additional ID mapping services are described at http://baderlab.org/IdentifierMapping/.

1. In the Cytoscape canvas, select all the nodes in your network by typing CTRL-A.
2. In the Node Attribute Browser, use your mouse to select the contents of the ID column for all nodes. This operation can be streamlined by selecting the first node ID and pressing CTRL-SHIFT-END, which will select all IDs in the column.
3. Go to the Synergizer identifier translation service at http://llama.med.harvard.edu/cgi/synergizer/translate.
4. Go to the box labeled IDs to translate: and paste in the list of identifiers.
5. Select your organism in the Select organism: field (*Saccharomyces cerevisiae* for our sample network). Set Select FROM namespace: to __ANY__. In the field labeled Select TO namespace:, select the identifier used in your expression data file. For the sample data, select systematic. Click the box labeled Output as Spreadsheet: and click Submit.
6. The identifier mapping will be downloaded to your computer in an Excel file named 'translate'. Open the file with a spreadsheet viewer such as Excel. The first column contains the original identifier list. The second column contains the identifiers that these map to in the chosen namespace. Note the label given to this column in the first line.
7. Save this file in the comma-separated (CSV) format.
▲ CRITICAL STEP There are system-dependent issues in producing and reading EXCEL files. This step guards against such issues.
8. Import these data into Cytoscape.
   a. Go to the File menu, then Import and then Attribute from Table (Text/MS Excel). This will bring up the Import Annotation File window.
   b. Click the Select File button and specify your 'translated' CSV file.
   c. Select Show Text File Import Options and select Transfer first line as attribute names.
   d. Click on the Import button.
   e. Verify that the new attribute is loaded by returning to the Node Attribute Browser and click on the Select Attribute button.

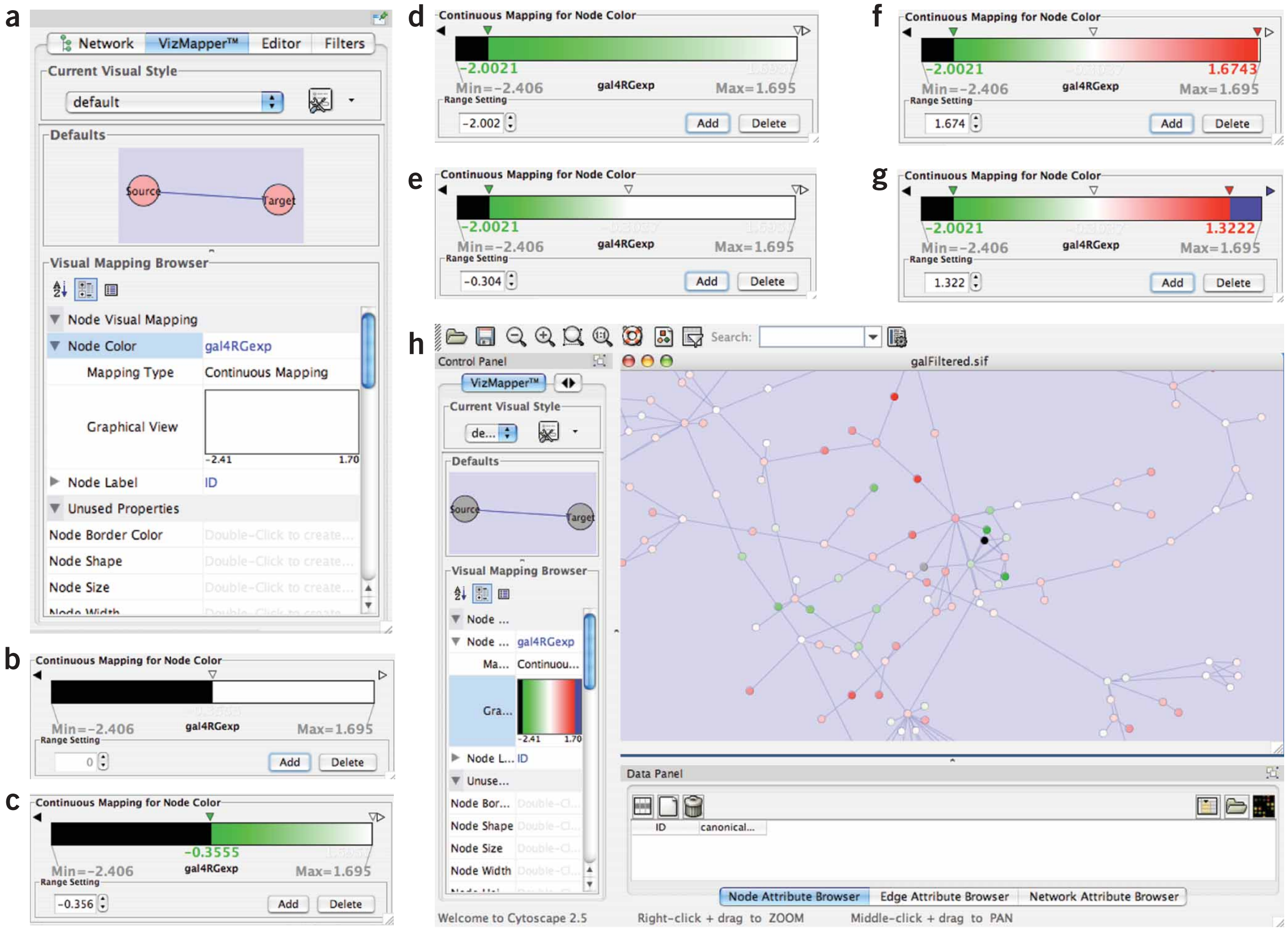


Figure 4 | The steps in creating a green-to-red node color gradient. (**a**) Specify a new attribute-based gradient; (**b**) add two range selector end points; (**c**) change the color at the lower end point to green; (**d**) move this end point to the bottom of the range; (**e**) add another new range selector end point; (**f**) change the color at the top of the range to red; (**g**) change the color at the top of the range to blue; and (**h**) a network with the nodes rendered with a red–green color map.

and edge rendering. (ii) Click on either node in this panel. This will bring up a window showing the default node attribute settings. (iii) In this window, click on the button for Node Fill Color and select Gray in the color selection window. Click Apply.

31| The nodes on the canvas should now be colored according to expression levels, which can suggest patterns of coregulation and modular structure. For example, examine the network for connected regions in which the nodes are predominantly colored red or green (see **Fig. 1**). Such regions are suggestive of protein complexes or signaling pathways whose activities are being modulated in the expression experiment. Several utilities are available to locate such regions automatically within large networks, including the jActiveModules plugin (Step 33) and Cytoscape Filters (not described here—see Cytoscape manual at http://www.cytoscape.org).

32| Supplementary data users: Change the node label from the default ID to the common gene name attribute as follows. (i) Locate the Node Label entry in the VizMapper. (ii) Click on the field opposite Node Label, which is set to 'ID' by default. Select 'GeneName' from the menu. (iii) Zoom in to the network. You should now see nodes labeled by their common gene names.

Analyze network features

33| Functional modules can be identified as highly connected network regions with similar responses across multiple experimental conditions. These patterns can be detected visually, but in practice the high level of connectivity in interaction networks makes visual detection difficult. Putative modules can be identified with the jActiveModules plugin as follows. Further details are available in the original publication[42]. Under the Plugins menu, select jActiveModules. This will bring up a jActiveModules section on the Control Panel, listing the available expression experiments.

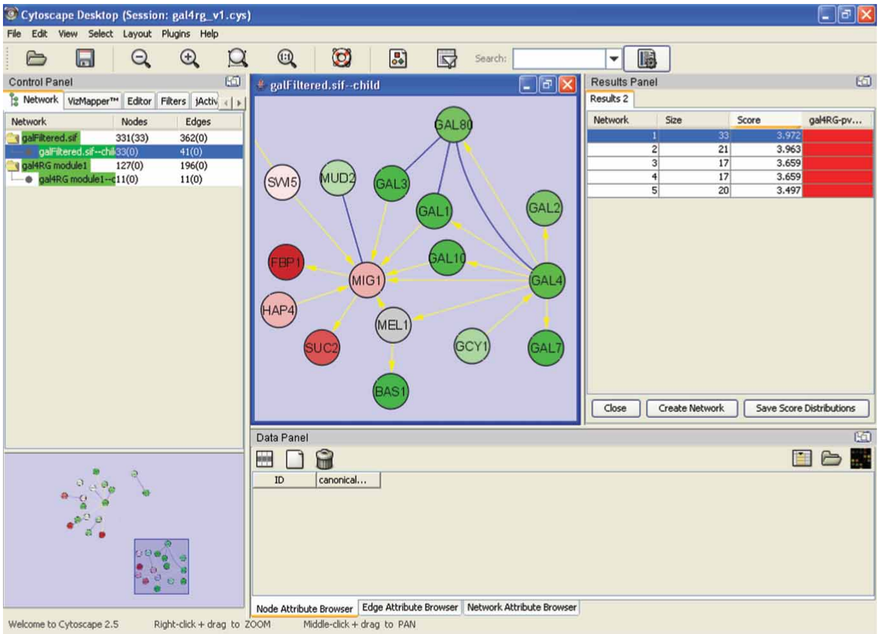
34| Select all of your expression experiments by clicking on them with the left mouse button (one can run jActiveModules on only one expression experiment or on a subset of the experiments as well).

35| In the General Parameters panel, keep the default values, as these are effective for most initial analyses. (a) ‘Number of modules’ indicates the number of putative modules that will be reported. (b) ‘Adjust score for size’ corrects for the fact that a larger putative module is more likely to contain nodes with significant *P*-values by random chance. (c) Regional scoring affects how the score of a given module is calculated. Instead of scoring only those nodes within the module, the neighboring nodes of the module are also included. This aids in the identification of active modules in networks that contain nodes with many neighbors (i.e., hubs). Consider a transcription factor with many targets: even if the transcription factor is not active, it is likely that some of its targets will be expressed simply due to random chance. Without regional scoring, this subset of targets can be selected as an active module, even though this arrangement is not unexpected. Regional scoring prevents this problem by forcing all targets to be scored simultaneously.

36| In the Strategy panel, the user can select the search strategy used to identify high-scoring modules. (a) Search: By this strategy, local (greedy) searches are initiated from single nodes in the network. (b) Search Depth: At each step in the greedy search, this parameter determines how close (as determined by the shortest path) a node must be to the current active module to be considered for inclusion. (c) Max Depth: This parameter determines how close a node must be to the initial seed node to be considered for inclusion. (d) Search from Selected Nodes: By default, a separate search is initiated for each node in the network. Using this option, searches are initiated only from those nodes selected by the user. (e) Anneal: In this strategy, all active modules in the network are discovered simultaneously using the method of simulated annealing as previously described[42]. The Annealing Parameters define the simulated annealing schedule. Annealing Extensions enable various modifications to the simulated annealing search procedure. For a detailed description of all parameters, click the Help button in the jActiveModules main panel.

37| Click the Find Modules button.

38| Shortly, the Results panel will appear, as illustrated in **Figure 5**. This window contains a table in which each row represents one putative module, listed according to the number of nodes, and an associated *Z*-score. *Z*-scores greater than 3.0 are generally considered significant. The table contains one column per expression experiment; for these columns, any row in which the cell is filled indicates an active module that showed a significant response under the conditions of the experiment. Clicking on any row causes the corresponding nodes to be selected in the Cytoscape canvas. To view these nodes as a separate network, click on the Create Network button. Lay out this new network using Layout → yFiles → Organic from within the new view.



Network	Nodes	Edges
galFiltered.sif	331(33)	362(0)
galFiltered.sif--chi	33(0)	41(0)
gal4RG module1	127(0)	196(0)
gal4RG module1--c	11(0)	11(0)

Network	Size	Score	gal4RG-pv...
1	33	3.972	🟥
2	21	3.963	🟥
3	17	3.659	🟥
4	17	3.659	🟥
5	20	3.497	🟥

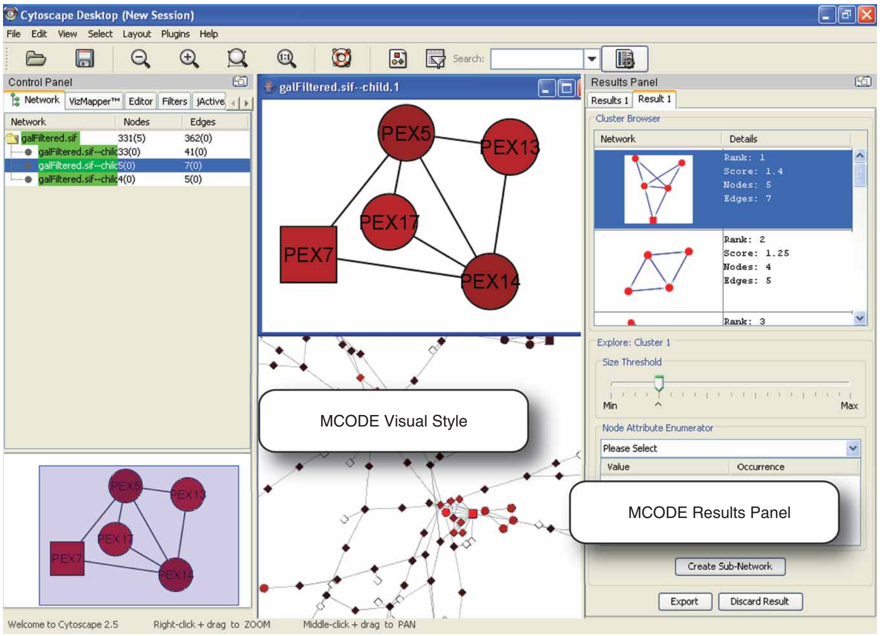
ID	canonical...

Figure 5 | JActiveModules output with the top-scoring module selected and displayed as a separate child network. Here, jActiveModules has analyzed the gal4RG expression condition to identify a module of predominantly downregulated genes (green) involved in galactose metabolism. Blue links: protein–protein interactions; yellow links: protein–DNA interactions.

39| Repeat the Find Modules step a few times. Depending on the parameter settings, jActiveModules relies on random sampling and is not guaranteed to return the same result with each execution. Consequently, it is worthwhile to run jActiveModules repeatedly to ensure that the approach is converging, that is, the identified modules are reproducible across runs. One method for validating module predictions is to determine if the nodes in the putative module are enriched for any GO biological processes, since it is expected that all genes in a module are involved in the same biological process. See Step 43.

40| Highly connected network regions can indicate protein complexes. As with the differentially expressed modules in the previous step, complexes can in principle be detected visually, but in practice this is difficult due to complex topologies of interaction networks. Putative complexes can be automatically identified using the MCODE plugin, as follows. Further details are available in the original MCODE publication[43]. Note that MCODE analysis does not require an expression data set, and thus can alternatively be performed immediately after Step 13. At this point, the Cytoscape session might contain several networks. Complexes are easier to detect in larger networks. Move to your largest interaction network by selecting the Network control in the Control Panel and then mouse-click on this network.

41| Run MCODE by navigating to the Plugins menu, then MCODE → Start MCODE and then clicking the Analyze button in the Control Panel. After a few seconds, the results will appear in the Results Panel, as shown in **Figure 6**.



Network	Nodes	Edges
galFiltered.sif	331(5)	362(0)
galFiltered.sif--chilc	33(0)	41(0)
galFiltered.sif--chilc	5(0)	7(0)
galFiltered.sif--chilc	4(0)	5(0)

Network	Details
	Rank: 1 Score: 1.4 Nodes: 5 Edges: 7
	Rank: 2 Score: 1.25 Nodes: 4 Edges: 5
	Rank: 3

Figure 6 | MCODE output, showing a densely connected network region that is a putative complex. This complex contains five PEX proteins involved in the function of the yeast peroxisome.

42| The results present a diagram of each putative complex and provide details including a score, the number of nodes and edges, and a diagram of the complex. Significant predictions are typically those with scores of greater than 2.0, and at least four nodes. Clicking on any row will highlight the corresponding nodes in the network canvas. The Cytoscape canvas shows the network according to the *MCODE visual style*, in which red nodes belong to high-scoring predicted complexes and black nodes do not. Squares and circles signify nodes in predicted complexes: squares indicate the 'seed' of the prediction, which are usually the nodes in the complex with the highest connectivity. Diamonds indicate nodes that are not part of a predicted complex. Note that the visual style can be changed back to its original settings by going to the VizMapper in the Control Panel and selecting a visual style using the Current Visual Style control. One method to validate a prediction of a complex is to determine if the nodes of the putative complex are enriched for any GO biological processes. The next section outlines the steps required.

Detect enriched gene functions

43| At this point, you should have identified one or more sets of nodes that may be functionally related, such as nodes in an expression module (jActiveModules) or a putative complex (MCODE). We will use the BiNGO plugin to assess if the nodes in a set are enriched for any biological processes recorded in the Gene Ontology database[68]. Further details are available in the original BINGO publication[7]. Select a group of nodes with a putative functional association. For instance, select a putative module from the jActiveModules results (Results Panel).

44| Under the Plugins menu, select BiNGO 2.0. This will bring up the BiNGO Settings window.

45| In the Cluster name: field, choose a name (arbitrary) of your selected node cluster. In the field labeled Select organism/annotation: select the organism (*Saccharomyces cerevisiae* for the sample data). The default values shown in the other menu options are appropriate for most cases. For more information on any parameter, click the Help button and follow the links to the BiNGO User Guide.

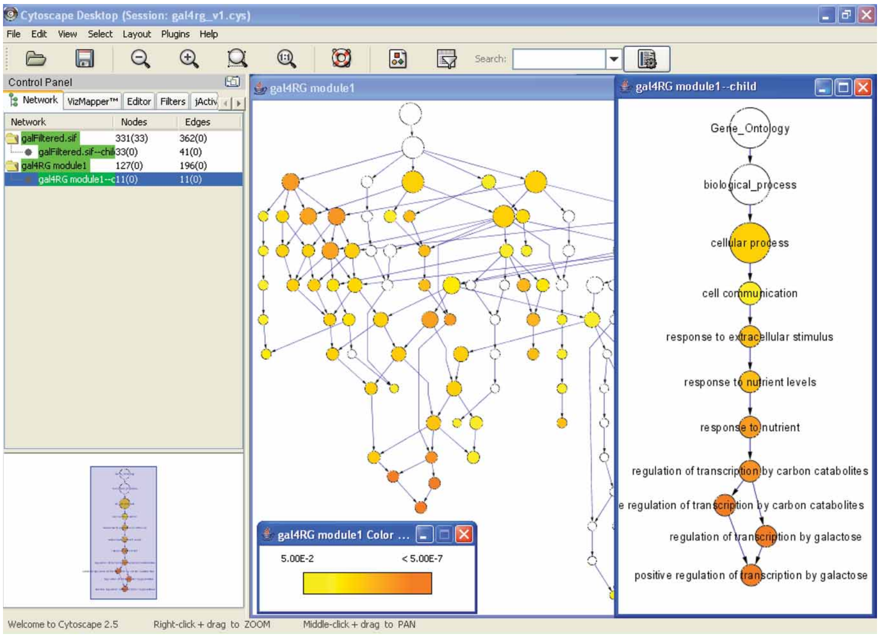
46| Click on the Start BiNGO button.

47| After a few minutes, the BiNGO Output window should appear listing the *P*-value (corrected for multiple testing) for each enriched GO class and the corresponding nodes from the network. On the Cytoscape desktop, a new network should appear as shown in **Figure 7**. In this network, each node represents a class in GO, and an edge between two nodes indicates that one class is the parent of the other. The rendering of each node indicates any significant enrichment for the corresponding GO class: (a) The size of each node is proportional to the number of proteins from the input set that belong to that class. (b) Nodes are color-coded according to *P*-values reporting the significance of enrichment of that term in the input set. Darker colors indicate a more significant degree of enrichment.

? TROUBLESHOOTING

● TIMING

The time required to execute this protocol is most strongly related to the time required to download the sample data and software over the Internet. With a 1.5 Mbs ADSL connection under favorable operating conditions, it takes approximately 9 min to download Cytoscape under Java Web Start, 2.5 min to download a network of ~1,000 nodes and ~2,000 interactions from cPath and 8 min to download and process 1,000 articles with the Agilent Literature Search plugin. On a Macintosh Powerbook G4 with 768 MB of memory, a network of ~3,000 nodes and ~7,000 edges requires approximately 1.50 min for loading, 2.25 min for analysis with jActiveModules using one expression experiment and 0.75 min for analysis with MCODE. An experienced user can execute the full protocol described within 30 min.


Cytoscape Desktop (Session: gal4rg_v1.cys)
File
Edit
View
Select
Layout
Plugins
Help
Search:
Control Panel
Network
VizMapper™
Editor
Filters
jActiv
gal4RG module1
gal4RG module1--child
Gene_Ontology
biological_process
cellular process
cell communication
response to extracellular stimulus
response to nutrient levels
response to nutrient
regulation of transcription by carbon catabolites
e regulation of transcription by carbon catabolites
regulation of transcription by galactose
positive regulation of transcription by galactose
gal4RG module1 Color ...
5.00E-2
< 5.00E-7
Welcome to Cytoscape 2.5
Right-click + drag to ZOOM
Middle-click + drag to PAN

Network	Nodes	Edges
galFiltered.sif	331(33)	362(0)
galFiltered.sif--chil	33(0)	41(0)
gal4RG module1	127(0)	196(0)
gal4RG module1--c	11(0)	11(0)

Figure 7 | BiNGO output, analyzing the active module shown in **Figure 5**. This analysis shows significant enrichment for processes relating to metabolism and galactose.

? TROUBLESHOOTING

Limited network data

The effectiveness of network analysis depends strongly on the amount of network data available. Species such as yeast, fly, *Escherichia coli* and human have tens of thousands of measured interactions available. For other species, the methods outlined here will likely become more effective over time as the amount of interaction data increases. Even in these cases, a gene association network can be constructed from published abstracts, the usefulness of which depends on the number of available publications, not the number of protein interaction assays. Another option when protein interaction measurements are lacking is to perform network analysis on a closely related organism for which more interactions have been measured (or for which a greater volume of publications are available). Protein identifiers can be mapped via orthology relationships, that is, as calculated by services such as PhyloFacts[69] or Roundup[70].

Mismatched gene identifiers

The most common problems with Cytoscape arise from mismatched gene identifiers across different input data files. This problem occurs most often when integrating attribute or expression data with an existing network. In Cytoscape, the identifiers in attribute or expression data files should match exactly to node names in the network in a case-sensitive match. If there is a case-sensitive match but data are still not loaded correctly, the following steps can help isolate the source of the problem:
(1) Under the Node or Edge Attribute Browser, select the new attribute from the list of available attributes by clicking on the leftmost button in the attribute browser toolbar (Select Attributes; square button with horizontal gray central stripe). If the name of the attribute does not appear, then the input file was not correctly loaded. Verify the format of your input file against the sample files provided in Supplementary Data or in the sampleData subdirectory within the Cytoscape home directory.
(2) If the name of the attribute appears, then select several nodes (or edges) in the Cytoscape canvas and check their attribute values in the Node/Edge Attribute Browser against the expected values. If any node or edge has a blank attribute value, then the identifier of the node or edge could not be matched against the input file. Ensure that the input attribute file contains exact matches to the nodes in the network.
(3) If attempting to load expression data using an ID-mapping identifier, first verify that the probeset attribute file was loaded correctly using the steps above. If the nodes have the probeset identifiers assigned correctly but have blank values for expression results, there is probably a mismatch between the probeset identifier in the probeset attribute file and the expression data file.

Multiple values for a node attribute

Sometimes a node attribute is a list of multiple values rather than just a single value (e.g., multiple gene function terms per node). However, in most other cases, Cytoscape assumes that a node or edge will be assigned only one value per attribute. In these cases, if a node or edge attribute file contains two values for the same node or edge, such as two probesets for one gene, the latter will be retained and the former overwritten. Consequently, if you discover that the expression data loaded for some node(s) are not as expected, verify that multiple attribute entries have not been specified.

Out-of-memory errors

As is true with many programs, Cytoscape can fail when the computer runs out of memory. This problem is manifest in a myriad of ways. Typically in these cases, Cytoscape will display an error message such as a Java Null Pointer Exception. In other cases, no error will be reported, but a pop-up window may fail to appear on the screen or a network may fail to appear in the Cytoscape canvas. If Cytoscape exhibits such behavior when analyzing a very large network, or running alongside many other applications on the same computer, insufficient memory may be the cause. The problem may be addressed by freeing memory: deleting extraneous networks, closing other applications that are running on the same computer or saving the session and rebooting the computer. If sufficient memory is available, Cytoscape may need to be set to use it. Further instructions on this are available at http://www.cytoscape.org/cgi-bin/moin.cgi/How_to_increase_memory_for_Cytoscape.

Failure to display very large networks

If a loaded network fails to display in the Cytoscape canvas, it may be that the loaded network is very large. By default, Cytoscape displays (or 'creates views') for networks of up to 10,000 nodes. If the user attempts to load a network of over 10,000 nodes, the network will be loaded into memory but by default no view will be created. In such cases, the network is listed in the control panel in red. At this point, the user can create a view by right-clicking on the network name in the Control Panel and selecting Create View. Before doing so, the user should bear in mind that visual analysis of very large networks is not usually productive, and that it may be more expedient to issue a more modest query or subdivide the input file. Also, the user can filter to select nodes and edges of interest in a loaded network even if it does not have a view and these can be extracted to create a smaller, more manageable network that can be reasonably viewed.

ANTICIPATED RESULTS

Network collection and layout

We have obtained the following average numbers of interactions for networks constructed based on the following methods and representative species. Numbers represent (mean ± s.d.) interactions per gene sampled from ten randomly chosen genes in July 2007. cPATH: yeast 90.0 ± 194.1, human 63.1 ± 81.6, mouse 3.3 ± 5.5, rat 0.5 ± 1.1. Agilent Literature Search: yeast 18.9 ± 22.7, human 178.8 ± 134.0, mouse 152.5 ± 143.1, rat 60.3 ± 92.1. If the network sizes obtained are much smaller than expected, see TROUBLESHOOTING. Successful import of the sample SIF file and application of the yFiles Organic layout algorithm produces a Cytoscape view similar to that shown in **Figure 1**.

In the near future, Cytoscape and its plugins will expand to make network loading easier. For instance, the cPath[33] plugin currently serves protein interaction data from recent releases of the IntAct[15] and MINT[17] public data repositories. The cPath site will grow to include a larger number of databases that share data in the BioPAX or PSI-MI[71] standard formats, providing a convenient point of access for public biological interaction and pathway information.

Analysis of network features

Execution of jActiveModules on the sample network with the expression condition gal4RG produces five modules (as specified by the default parameters), all having significant *Z*-scores >3.0. The top module (shown in **Fig. 5**) contains the genes encoding the known galactose enzymes (*GAL1*, *7*, *10*), the galactose transporter (*GAL2*) and the galactose pathway transcriptional regulators (*GAL4*, *80*, *3*). Gal4, the central transcriptional activator of the pathway, is connected to other *GAL* genes through protein–DNA interactions (yellow), which are strongly differentially expressed since the *GAL4* gene is deleted in this experiment. In this way, expression values superimposed on a protein network in the Cytoscape canvas can identify expression-activated pathways or protein complexes.

Execution of MCODE on the sample network produces 46 protein interaction clusters with scores ranging from 1 to 5.5, with the five top scores >2.5. **Figure 6** displays one such cluster containing five PEX genes that encode components of a protein complex associated with the peroxisomal membrane. This set of proteins has been detected as a putative complex by MCODE based on its high level of enrichment for protein–protein interactions (seven observed interactions out of ten possible among five proteins).

The key regulatory "control points" of a biological system may not always be genes with the most pronounced expression ratios; more often, the most differentially expressed genes are in fact regulated by the controller, which lies upstream. An effective approach to finding upstream agents is to identify groups of genes that exhibit large measured changes under the experimental conditions and then to examine the genes adjacent to those in the network. When doing so, bear in mind that existing network data may be incomplete, and so one should not expect all connections between the causative agent(s) and the affected genes to be present in the network.

Evidence suggests that the activity of a complex may be controlled by restricting the production of one component, even if most are produced ubiquitously[72]. Thus, if one gene in the putative complex shows decreased expression levels, the complex may be less active under the given experimental conditions (although it is known that mRNA–protein levels do not always correlate). Similarly, signal-transduction pathways are often branched with multiple alternatives, yet all that is necessary for signaling is that the proteins in some branch are available and activated. The expression levels on each node suggest which alternatives may be produced under the given experimental conditions. In cases where the signaling activity can be measured, gene knockdown methods such as RNA interference can be used to verify these hypotheses experimentally.

Note: Supplementary information is available via the HTML version of this article.

ACKNOWLEDGMENTS Many research groups have developed plugins to Cytoscape and provided them for download free of charge from http://www.cytoscape.org/. These plugins represent key contributions to the overall utility of Cytoscape, and we gratefully thank the authors for their contributions. Cytoscape is developed through an ongoing collaboration between the University of California at San Diego, the University of Toronto, the Institute for Systems Biology, Memorial Sloan-Kettering Cancer Center, Institut Pasteur, Agilent Technologies and the University of California at San Francisco. Many developers have contributed to Cytoscape, and we gratefully acknowledge the contributions of former developers including Nada Amin, Mark Anderson, Richard Bonneau, Larissa Kamenkovich, Andrew Markiel, Owen Ozier, Paul Shannon, Robert Sheridan and Jonathan Wang. We thank Tero Aittikalio and Cricket Sloan for assistance with the manuscript. Funding for Cytoscape is provided by the US National Institute of General Medical Sciences of the National Institutes of Health under award number GM070743-01. Corporate funding is provided through a contract from Unilever PLC. Cytoscape contributions by G.D.B. were funded in part by Genome Canada. The BiNGO plugin was developed at the Department of Plant Systems Biology at the University of Ghent, with partial funding through the Research Foundation Flanders in Belgium.

AUTHOR CONTRIBUTIONS Cytoscape development is a large community effort, with principal efforts under the direction of A.A. at Agilent Technologies; I.S. and L.H. at the Institute for Systems Biology; C.S. at Memorial Sloan Kettering Cancer Center; B.S. at Institut Pasteur; G.J.W. at Unilever; T.I. at University of California, San Diego; B.R.C. at University of California, San Francisco; and G.D.B. at University of Toronto. Every author has made significant contributions to the software, without which this protocol would not be possible. Specific contributions are as follows. E.C., B.G., G.D.B. and C.S. developed the cPath plugin. A.K., A.V. and M.C. developed the Agilent Literature Search plugin. S.M., R.I. and M.K. developed the BiNGO plugin. M.S., T.I. and R.K. developed the jActiveModules plugin. V.P., G.D.B. and C.S. developed the MCODE plugin. N.L., R.C., I.A.C., S.K., S.L., M.S., K.O. and P.-L.W. developed much of the improvements to the Cytoscape v2.5 core, including automatic loading of plugins and the VizMapper. K.H., J.M. and A.R.P. were instrumental in the usability analysis, specification and prototyping of many of these features. This protocol was written by M.S.C., M.S., A.K., S.M., C.W., G.D.B. and T.I.

Published online at http://www.natureprotocols.com
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Kislinger, T. *et al.* Global survey of organ and organelle protein expression in mouse: combined proteomic and transcriptomic profiling. *Cell* **125**, 173–186 (2006).
2. Galbraith, D.W. & Birnbaum, K. Global studies of cell type-specific gene expression in plants. *Annu. Rev. Plant Biol.* **57**, 451–475 (2006).
3. Butcher, E.C., Berg, E.L. & Kunkel, E.J. Systems biology in drug discovery. *Nat. Biotechnol.* **22**, 1253–1259 (2004).
4. Bader, G.D. *et al.* Functional genomics and proteomics: charting a multidimensional map of the yeast cell. *Trends Cell Biol.* **13**, 344–356 (2003).
5. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550 (2005).
6. Doniger, S.W. *et al.* MAPPFinder: using gene ontology and GenMAPP to create a global gene-expression profile from microarray data. *Genome Biol.* **4**, R7 (2003).
7. Maere, S., Heymans, K. & Kuiper, M. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* **21**, 3448–3449 (2005).
8. Zeeberg, B.R. *et al.* GoMiner: a resource for biological interpretation of genomic and proteomic data. *Genome Biol.* **4**, R28 (2003).
9. Quackenbush, J. Weighing our measures of gene expression. *Mol. Syst. Biol.* **2**, 63 (2006).
10. Sprinzak, E., Sattath, S. & Margalit, H. How reliable are experimental protein–protein interaction data? *J. Mol. Biol.* **327**, 919–923 (2003).
11. D'Haeseleer, P. & Church, G.M. Estimating and improving protein interaction error rates. *Proc. IEEE Comput. Syst. Bioinform. Conf.* 216–223 (2004).
12. Singhal, M. & Domico, K. CABIN: collective analysis of biological interaction networks. *Comput. Biol. Chem.* **31**, 222–225 (2007).
13. Barrett, T. & Edgar, R. Gene expression omnibus: microarray data storage, submission, retrieval, and analysis. *Methods Enzymol.* **411**, 352–369 (2006).
14. Parkinson, H. *et al.* ArrayExpress—a public database of microarray experiments and gene expression profiles. *Nucleic Acids Res.* **35**, D747–D750 (2007).
15. Hermjakob, H. *et al.* IntAct: an open source molecular interaction database. *Nucleic Acids Res.* **32**, D452–D455 (2004).
16. Peri, S. *et al.* Human protein reference database as a discovery resource for proteomics. *Nucleic Acids Res.* **32**, D497–D501 (2004).
17. Zanzoni, A. *et al.* MINT: a Molecular INTeraction database. *FEBS Lett.* **513**, 135–140 (2002).
18. Bader, G.D., Cary, M.P. & Sander, C. Pathguide: a pathway resource list. *Nucleic Acids Res.* **34**, D504–D506 (2006).
19. Johnson, D.S., Mortazavi, A., Myers, R.M. & Wold, B. Genome-wide mapping of *in vivo* protein–DNA interactions. *Science* **316**, 1497–1502 (2007).
20. Krallinger, M. & Valencia, A. Text-mining and information-retrieval services for molecular biology. *Genome Biol.* **6**, 224 (2005).
21. Vailaya, A. *et al.* An architecture for biological information extraction and representation. *Bioinformatics* **21**, 430–438 (2005).
22. Mishra, G.R. *et al.* Human protein reference database—2006 update. *Nucleic Acids Res.* **34**, D411–D414 (2006).
23. Breitkreutz, B.J., Stark, C. & Tyers, M. Osprey: a network visualization system. *Genome Biol.* **4**, R22 (2003).
24. Hu, Z. *et al.* VisANT: data-integrating visual framework for biological networks and modules. *Nucleic Acids Res.* **33**, W352–W357 (2005).
25. Funahashi, A., Morohashi, M., Kitano, H. & Tanimura, N. CellDesigner: a process diagram editor for gene-regulatory and biochemical networks. *Biosilico* **1**, 159–162 (2004).
26. Dahlquist, K.D., Salomonis, N., Vranizan, K., Lawlor, S.C. & Conklin, B.R. GenMAPP, a new tool for viewing and analyzing microarray data on biological pathways. *Nat. Genet.* **31**, 19–20 (2002).
27. Aragues, R., Jaeggi, D. & Oliva, B. PIANA: protein interactions and network analysis. *Bioinformatics* **22**, 1015–1017 (2006).
28. Iragne, F., Nikolski, M., Mathieu, B., Auber, D. & Sherman, D. ProViz: protein interaction visualization and exploration. *Bioinformatics* **21**, 272–274 (2005).
29. Goldovsky, L., Cases, I., Enright, A.J. & Ouzounis, C.A. BioLayout(Java): versatile network visualisation of structural and functional relationships. *Appl. Bioinformatics* **4**, 71–74 (2005).
30. Demir, E. *et al.* PATIKA: an integrated visual environment for collaborative construction and analysis of cellular pathways. *Bioinformatics* **18**, 996–1003 (2002).
31. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
32. Avila-Campillo, I., Drew, K., Lin, J., Reiss, D.J. & Bonneau, R. BioNetBuilder: automatic integration of biological networks. *Bioinformatics* **23**, 392–393 (2007).
33. Cerami, E.G., Bader, G.D., Gross, B.E. & Sander, C. cPath: open source software for collecting, storing, and querying biological pathways. *BMC Bioinformatics* **7**, 497 (2006).
34. Jayapandian, M. *et al.* Michigan Molecular Interactions (MiMI): putting the jigsaw puzzle together. *Nucleic Acids Res.* **35**, D566–D571 (2007).
35. Salwinski, L. & Eisenberg, D. The MiSink Plugin: Cytoscape as a graphical interface to the database of interacting proteins. (2007).
36. Garcia, O. *et al.* GOlorize: a Cytoscape plug-in for network visualization with gene ontology-based layout and coloring. *Bioinformatics* **23**, 394–396 (2007).

37. Barsky, A., Gardy, J.L., Hancock, R.E. & Munzner, T. Cerebral: a Cytoscape plugin for layout of and interaction with biological networks using subcellular localization annotation **23**, 1040–1042 (2007).
38. Yip, K.Y., Yu, H., Kim, P.M., Schultz, M. & Gerstein, M. The tYNA platform for comparative interactomics: a web tool for managing, comparing and mining multiple networks. *Bioinformatics* **22**, 2968–2970 (2006).
39. Ferro, A. *et al.* NetMatch: a Cytoscape plugin for searching biological networks. *Bioinformatics* **23**, 910–912 (2007).
40. Vlasblom, J. *et al.* GenePro: a Cytoscape plug-in for advanced visualization and analysis of interaction networks. *Bioinformatics* **22**, 2178–2179 (2006).
41. Luo, F. *et al.* Modular organization of protein interaction networks. *Bioinformatics* **23**, 207–214 (2007).
42. Ideker, T., Ozier, O., Schwikowski, B. & Siegel, A.F. Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics* **18** (Suppl 1): S233–S240 (2002).
43. Bader, G.D. & Hogue, C.W. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* **4**, 2 (2003).
44. Albrecht, M., Huthmacher, C., Tosatto, S.C. & Lengauer, T. Decomposing protein networks into domain–domain interactions. *Bioinformatics* **21** (Suppl 2): ii220–ii221 (2005).
45. Tong, A.H. *et al.* Global mapping of the yeast genetic interaction network. *Science,* **303**, 808–813 (2004).
46. Pan, X. *et al.* A DNA integrity network in the yeast *Saccharomyces cerevisiae*. *Cell* **124**, 1069–1081 (2006).
47. Kelley, R. & Ideker, T. Systematic interpretation of genetic interactions using protein networks. *Nat. Biotechnol.* **23**, 561–566 (2005).
48. Drees, B.L. *et al.* Derivation of genetic interaction networks from quantitative phenotype data. *Genome Biol.* **6**, R38 (2005).
49. Gilchrist, M. *et al.* Systems biology approaches identify ATF3 as a negative regulator of Toll-like receptor 4. *Nature* **441**, 173–178 (2006).
50. Yeang, C.H. *et al.* Validation and refinement of gene-regulatory pathways on a network of physical interactions. *Genome Biol.* **6**, R62 (2005).
51. Krogan, N.J. *et al.* Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* **440**, 637–643 (2006).
52. Rhodes, D.R. *et al.* Probabilistic model of the human protein-protein interaction network. *Nat. Biotechnol.* **23**, 951–959 (2005).
53. Gutierrez, R.A. *et al.* Qualitative network models and genome-wide expression data define carbon/nitrogen-responsive molecular machines in *Arabidopsis*. *Genome Biol.* **8**, R7 (2007).
54. Gunsalus, K.C. *et al.* Predictive models of molecular machines involved in *Caenorhabditis elegans* early embryogenesis. *Nature* **436**, 861–865 (2005).
55. Sharan, R. *et al.* Conserved patterns of protein interaction in multiple species. *Proc. Natl. Acad. Sci. USA* **102**, 1974–1979 (2005).
56. King, J.Y. *et al.* Pathway analysis of coronary atherosclerosis. *Physiol. Genomics* **23**, 103–118 (2005).
57. Ideker, T. *et al.* Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* **292**, 929–934 (2001).
58. Wixon, J. & Kell, D. The Kyoto encyclopedia of genes and genomes—KEGG. *Yeast* **17**, 48–55 (2000).
59. Joshi-Tope, G. *et al.* Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res.* **33**, D428–D432 (2005).
60. Stromback, L., Jakoniene, V., Tan, H. & Lambrix, P. Representing, storing and accessing molecular interaction data: a review of models and tools. *Brief Bioinform.* **7**, 331–338 (2006).
61. Wuchty, S., Barabasi, A.L. & Ferdig, M.T. Stable evolutionary signal in a yeast protein interaction network. *BMC Evol. Biol.* **6**, 8 (2006).
62. Yu, H., Kim, P.M., Sprecher, E., Trifonov, V. & Gerstein, M. The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. *PLoS Comput. Biol.* **3**, e59 (2007).
63. Ghaemmaghami, S. *et al.* Global analysis of protein expression in yeast. *Nature* **425**, 737–741 (2003).
64. Margolin, A.A. *et al.* Reverse engineering cellular networks. *Nat. Protoc.* **1**, 662–671 (2006).
65. Fu, J., Swertz, M.A., Keurentjes, J.J. & Jansen, R.C. MetaNetwork: a computational protocol for the genetic study of metabolic networks. *Nat. Protoc.* **2**, 685–694 (2007).
66. Eyre, T. *et al.* The HUGO gene nomenclature database, 2006 updates. *Nucleic Acids Res.* **1**, D319–D321 (2006).
67. Di Battista, G., Eades, P., Tamassia, R. & Tollis, I. *Graph Drawing: Algorithms for the Visualization of Graphs* (Prentice-Hall, Upper Saddle River, NJ, USA, 1999).
68. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The gene ontology consortium. *Nat. Genet.* **25**, 25–29 (2000).
69. Krishnamurthy, N., Brown, D.P., Kirshner, D. & Sjolander, K. PhyloFacts: an online structural phylogenomic encyclopedia for protein functional and structural classification. *Genome Biol.* **7**, R83 (2006).
70. Deluca, T.F. *et al.* Roundup: a multi-genome repository of orthologs and evolutionary distances. *Bioinformatics* **22**, 2044–2046 (2006).
71. Hermjakob, H. *et al.* The HUPO PSI's molecular interaction format—a community standard for the representation of protein interaction data. *Nat. Biotechnol.* **22**, 177–183 (2004).
72. de Lichtenberg, U., Jensen, L.J., Brunak, S. & Bork, P. Dynamic complex formation during the yeast cell cycle. *Science* **307**, 724–727 (2005).
73. Irizarry, R.A. *et al.* Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.* **31**, e15 (2003).