

Assembling global maps of cellular function through integrative analysis of physical and genetic networks

Rohith Srivas^{1,3}, Gregory Hannum^{1,3}, Johannes Ruscheinski^{1,2}, Keiichoro Ono^{1,2}, Peng-Liang Wang^{1,2}, Michael Smoot^{1,2} & Trey Ideker^{1,2}

¹Department of Bioengineering, University of California, San Diego, La Jolla, California, USA. ²Department of Medicine, University of California, San Diego, La Jolla, California, USA. ³These authors contributed equally to this work. Correspondence should be addressed to T.I. (tideker@ucsd.edu).

Published online 11 August 2011; doi:10.1038/nprot.2011.368

To take full advantage of high-throughput genetic and physical interaction mapping projects, the raw interactions must first be assembled into models of cell structure and function. PanGIA (for physical and genetic interaction alignment) is a plug-in for the bioinformatics platform Cytoscape, designed to integrate physical and genetic interactions into hierarchical module maps. PanGIA identifies ‘modules’ as sets of proteins whose physical and genetic interaction data matches that of known protein complexes. Higher-order functional cooperativity and redundancy is identified by enrichment for genetic interactions across modules. This protocol begins with importing interaction networks into Cytoscape, followed by filtering and basic network visualization. Next, PanGIA is used to infer a set of modules and their functional inter-relationships. This module map is visualized in a number of intuitive ways, and modules are tested for functional enrichment and overlap with known complexes. The full protocol can be completed between 10 and 30 min, depending on the size of the data set being analyzed.

INTRODUCTION

Genetic interactions are defined as functional relationships between genes that result when the phenotypic effect of one gene is altered by one or several other genes^{1,2}. Such interactions have been used to uncover pathway architecture in model organisms^{3–6}. In humans, genetic interactions are thought to influence numerous phenotypes of interest, from expression⁷ to complex diseases⁸ to drug resistance⁹. Recently, a number of technologies such as synthetic genetic arrays^{6,10–12} and heterozygote diploid-based synthetic lethality analysis with microarray¹³ have facilitated the rapid screening of genetic interactions in model organisms. In human cell lines, combinatorial RNA interference screening technologies have begun to show promise in uncovering genetic interactions^{14,15}. As a result of these high-throughput technologies, the amount of genetic interaction data available in the public domain has increased rapidly. As of December 2010, the BioGRID interaction database housed nearly 175,000 genetic interactions spanning 11 different species¹⁶.

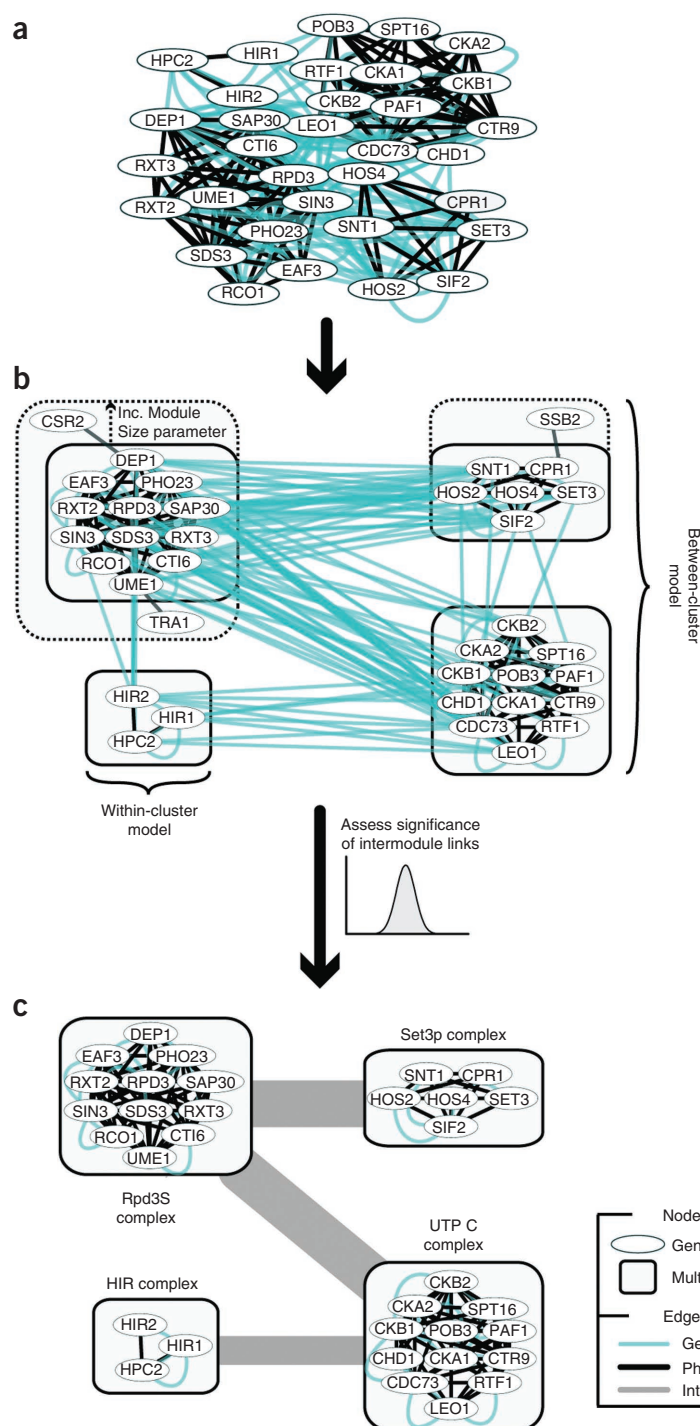
Interpreting the functional significance of each genetic interaction remains a daunting task. One promising solution has been to interpret genetic interactions in the context of their relationships to physical protein-protein interactions (Fig. 1a)^{17–20}. At least two distinct models have been put forth to reconcile genetic and physical interactions. The ‘within-cluster’ model seeks to identify clusters of proteins that are enriched for both physical and genetic interactions (Fig. 1b). We refer to such clusters of proteins and the interactions occurring among them as a module. Modules are often interpreted as functional protein complexes^{6,17–19} or signaling pathways⁹. In contrast, the ‘between-cluster’ model seeks genetic interactions that are enriched across two clusters of interacting proteins (Fig. 1b). Such intermodule links have been shown to identify synergistic or compensatory relationships between protein complexes or signaling pathways^{3,18,20}. Figure 1c shows an example module map consisting of four modules connected by

three intermodule links. The genes in each of these four modules are associated with a strong within-cluster signal, and, furthermore, they coincide with known *Saccharomyces cerevisiae* physical complexes (Fig. 1c). Set3p and Rpd3s are both histone deacetylase complexes involved in transcriptional regulation. The Hir complex functions in replication-independent nucleosome assembly, whereas the UTP-C complex is a component of the 90S preribosome. The intermodule link between Set3p and Rpd3s suggests a functional synergy between the two complexes. Consistent with this hypothesis, several studies have illustrated that the two are jointly responsible for the activation of DNA damage response genes via the recruitment of RNA Polymerase II (ref. 21).

Several methods have been previously published^{17–19,22} for analyzing interactions to identify both within-cluster and between-cluster functional organization. However, these methods have not yet been made available through a publicly accessible software package. Here we introduce a novel software tool, PanGIA, along with a general bioinformatics protocol for integrative analysis of genetic interactions. PanGIA implements a previously published framework²⁰ as a plug-in for the open-source network analysis platform, Cytoscape^{23,24}, and allows the user to easily generate maps of modules and module inter-relationships from genetic and physical interaction data (see Fig. 1 for an overview). A number of options are available to the user for constructing and visualizing the resulting module map. PanGIA is built on the new Cytoscape 2.8 architecture²⁵, which features the ability to view and manipulate nested networks, thereby enabling the user to explore both the global map as well as individual modules in an intuitive manner. Finally, individual modules can be interrogated using a number of functional enrichment options.

The computational workflow presented here has been used in the analysis of genetic networks centered on genes involved in chromosomal biology^{3,20}, RNA processing²⁶, secretory pathways⁵

Figure 1 | Overview of PanGIA's method for identifying a module map of cellular function from physical and genetic networks. **(a)** PanGIA takes as input a physical and genetic network. Black edges refer to physical interactions, whereas turquoise edges refer to genetic interactions. **(b)** Both within-cluster and between-cluster models are identified using the physical and genetic network. A within-cluster model or module consists of a set of genes connected by a large number of physical and genetic interactions. In this example four within-cluster models are identified. A between-cluster model or intermodule link consists of two within-cluster models spanned by a bundle of genetic interactions. Here, five putative between-cluster models have been identified. The size of within-cluster models can be controlled via the Module Size parameter. Higher values of the Module Size parameter lead to larger complexes (denoted by the dashed line). **(c)** If quantitative interaction data have been made available, the significance of each between-cluster model can be assessed. Only significant intermodule links are displayed in the final module map (three of the five putative intermodule links are significant in this example). The thickness of the line reflects the score of the intermodule link, which is based on the number of physical and genetic edges spanning the two modules. If a biological annotation set is provided, PanGIA will check the overlap between the set of genes comprising the annotation and the set of genes comprising each module. If the overlap exceeds a user-specified threshold, the module will be labeled with the name of the annotations. Here, all four modules overlap with known complexes and are labeled accordingly.



and DNA damage response⁹. This analysis has also been used in comparing genetic networks across two different species²⁷. In each case, the module maps generated have helped to identify novel pathways as well as new components and functions for existing complexes^{9,18–20,27}. While this workflow has proven useful in the analysis of numerous genetic interaction data sets, the module search process works best when there is a high density of protein and genetic interactions among the set of genes being studied. For species in which there is a scarcity of either genetic interaction or physical interaction data, this protocol may not identify a significant number of modules or intermodule relationships. This limitation will become less relevant as large-scale interaction screens continue to populate the scientific databases.

This protocol is divided into five basic sections (Fig. 2). The first section, 'Importing physical and genetic networks into Cytoscape', describes the available sources of interaction data and means of acquiring these data within Cytoscape. Second, 'Generating a module map using the PanGIA plug-in' covers the use of the PanGIA plug-in and is further divided into four

subsections covering the various aspects of its use ('Selecting a physical and genetic network', 'Setting the module size and edge reporting parameters', 'Training PanGIA' and, finally, 'Labeling modules'). The third section, 'Visualization of the module map using nested networks', introduces ways in which the user can navigate and visualize the resulting module map. Fourth, 'Functional enrichment of the modules' illustrates methods to identify enriched biological functions and pathways among the identified modules. Finally, 'Exporting the results' covers the various ways in which the module map can be exported from Cytoscape for further analysis or for inclusion as figures in a publication.

Figure 2 | Outline of the protocol. Analyses listed in black indicate required steps in the protocol. Analyses listed in orange represent optional steps, which may be performed if quantitative interaction data are present; those listed in light blue are optional steps, which may be performed if a biological annotation data set is present. The yellow boxes indicate the desired outcome at the end of each major section in the protocol.

Importing physical and genetic networks into Cytoscape

This section of the PROCEDURE (Steps 1–18) describes the various ways in which a physical or genetic network can be imported for analysis into Cytoscape. A previous protocol has outlined the various file formats Cytoscape can recognize as well as provided detailed instructions on how each file type can be imported²⁴. The present protocol will instead focus on importing networks in a tab-delimited format (**Box 1**). **Table 1** provides examples of several different databases from which interaction data (both genetic and physical) can be downloaded in a tab-delimited format for over 50 organisms.

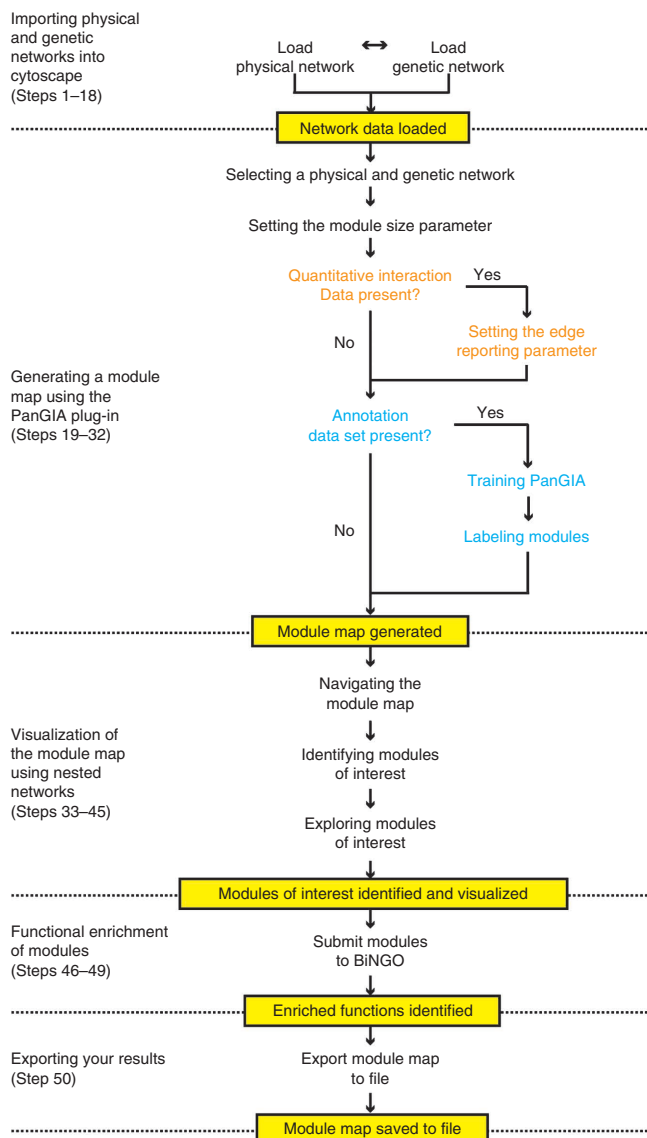
Generating a module map using the PanGIA plug-in

Selecting a physical and genetic network. This section of the PROCEDURE (Steps 19–23) describes the steps necessary to select which physical and genetic networks are to be analyzed. At this point, PanGIA is fully configured and the module search process can be initiated. However, PanGIA is designed with four optional features designed to fine-tune and enhance the search process. We describe these optional features in the subsequent sections.

Setting the module size and edge reporting parameters (Steps 24–26). The first optional feature is the ‘module size’ parameter. This parameter helps to control both the size and number of modules by rewarding the formation of larger modules. Thus, higher values of this parameter results in the formation of larger, but fewer modules. Lower values produce the opposite effect (**Fig. 1b**). It is recommended that the module size parameter initially be left at the default value. If the resulting module map contains very large modules, the module size parameter can be suitably altered and the module search process re-run to produce smaller and more biologically meaningful modules.

The second optional feature is dependent on the presence of quantitative genetic interaction data. Many of the recent experimental technologies for measuring genetic interactions go beyond reporting interactions in a simple binary format (interacting or noninteracting) and provide some measure of confidence in a given interaction. For example, in the synthetic genetic array technology¹² and a recent variant called epistatic mini-array profiles^{3,11}, each double mutant is assigned a quantitative signed score, where positive scores indicate that the double mutant grew better than expected (e.g., suppression) and negative scores indicate pairs for which the double mutant grew worse than expected (e.g., synthetic sick or synthetic lethal)^{11,12}. **Table 1** outlines numerous databases that contain quantitative interaction data.

If quantitative genetic interaction data are provided, each inter-module link can be assessed for significance. A *P* value is assigned by comparing the sum of the interaction confidence values for all genetic interactions spanning two modules (i.e., intermodule link) to a distribution of the sums of confidence values of an equal number of genetic interactions drawn at random²⁰ (**Fig. 1c**). The edge reporting parameter serves as a threshold; only those interactions with a *P* value less than this threshold are displayed in the final module map. By default, this parameter is set to 0.1, thus displaying only those intermodule links with *P* < 0.1.



Training PANGIA (Steps 27–29). The next optional feature relies on the presence of a biological annotation set. Examples of an annotation set that can be used include physical complexes, signaling pathways, metabolic pathways or even broad biological processes. **Table 2** provides a list of databases where an annotation set can be downloaded for a range of different organisms.

The optional training procedure built into PanGIA is designed to help identify modules that are more likely to be biologically relevant, i.e., modules that contain genes that operate in the same complex or biological process. By default, the module search process is designed to identify sets of genes that are densely connected by physical and genetic interactions. However, some interactions can be given more or less influence based on their quantitative score. PanGIA can determine how likely a certain interaction (either physical or genetic) is to connect two genes within a known complex or biological process using an existing annotation set. Examples of such a set include physical complexes (e.g., INO80 complex), signaling pathways (e.g., the mitogen-activated protein kinase (MAPK) pathway), metabolic pathways (e.g., glycolysis) or biological processes (e.g., DNA damage response genes). Using

BOX 1 | THE TAB-DELIMITED NETWORK FORMAT

In the tab-delimited network format each line in the file represents a single interaction and consists of two or three parts separated by a tab-character. The first part is the source node. The second part is the target node. The third part, if present, represents the quantitative value (i.e., confidence) attached to this particular interaction. A sample file might look like this:

```
nodeA nodeB 32.14
```

Or, if no quantitative interaction data are present:

```
nodeB nodeC
```

In the first example network, nodeA interacts with nodeB. The strength or confidence of this interaction is 32.14. In the second example, nodeB is interacting with nodeC. No quantitative value has been attached to this interaction. The Supplementary Data contains two tab-delimited network files: *Collins_physical_network_example.txt* (**Supplementary Data 1**) and *Collins_genetic_network_example.txt* (**Supplementary Data 2**).

this annotation set, PanGIA assigns each interaction a weight based on the unsigned logistic regression of all interaction confidence scores of a given type (physical, genetic) against its proteins' co-membership in an annotation. If no quantitative scores are available, PanGIA uses logistic regression to assign a constant confidence score for all interactions of a given type. For specific details regarding the regression procedure, please see Bandyopadhyay *et al.*²⁰. The module search process will now seek to identify sets of genes that are connected by highly weighted physical and genetic interactions. As the weight of an interaction corresponds to how likely it is to connect two genes belonging to the same physical complex or pathways, the modules identified will contain genes that are functionally similar.

Labeling modules. The genes composing a module may function in the same biological process or encode members of the same protein complex. If a biological annotation set is provided, PanGIA will check to see if the module gene set overlaps with the annotation gene set. Here overlap is defined using the Jaccard similarity coefficient (intersection/union), which ranges from 0 (no overlap) to 1 (perfect overlap). If the Jaccard coefficient exceeds a user-specified threshold, then the module will be labeled with the name of the annotation in the final module map (**Fig. 1c**). This PROCEDURE subsection (Steps 30–32) covers how this labeling feature can be enabled and provides instructions on how to set the overlap threshold.

Visualization of the module map using nested networks

PanGIA is built on the new Cytoscape 2.8 architecture, which features the ability to view nested networks (i.e., each node in a network can represent an entire subnetwork). Instructions are provided for laying out the network of modules and intermodule links and for probing individual modules. This PROCEDURE section is divided into three subsections, 'Navigating the module map' (Steps 33–35), 'Finding modules of interest' (Step 36) and 'Exploring modules of interest' (Steps 37–45), which cover the various ways in which both the module map and individual modules can be interrogated.

Functional enrichment of the modules

Modules will often contain genes of unknown function. One way to dissect the function of modules uncovered in this workflow is to examine if they are substantially enriched for any functional annotations. This can be used to identify new components of existing complexes or to identify entirely new physical complexes or pathways^{3,18,20}. This PROCEDURE section (Steps 46–49) outlines the steps for checking for enriched Gene Ontology (GO) functional terms²⁸ using the BiNGO plug-in²⁹.

Exporting your results

This PROCEDURE section (Step 50) covers the various options for exporting the resulting module map.

TABLE 1 | List of databases of physical and genetic interaction data.

Database name	URL	No. of organisms covered	Physical interaction data available?	Genetic interaction data available?	Quantitative interaction data available?
STRING	http://string-db.org/	630	Yes	No	Yes
DIP	http://dip.doe-mbi.ucla.edu/dip/Main.cgi	372	Yes	No	Yes
IntAct	http://www.ebi.ac.uk/intact/main.xhtml	305	Yes	No	Yes
ConsensusPathDB	http://cpdb.molgen.mpg.de/	3	Yes	No	No
BioGRID	http://thebiogrid.org/	18	Yes	Yes	No
MINT	http://mint.bio.uniroma2.it/mint/Welcome.do	30	Yes	No	Yes
DroID	http://www.droidb.org	1	Yes	Yes	No
DRYGIN	http://drygin.ccb.utoronto.ca	1	No	Yes	Yes

TABLE 2 | Examples of databases from which to obtain annotation data.

Database name	URL	No. of organisms covered	Annotation type
Gene Ontology (GO)	http://www.geneontology.org/GO.downloads.annotations.shtml	48	Physical complexes, biological processes, signaling pathways, metabolic pathways
MIPS CORUM	http://mips.helmholtz-muenchen.de/genre/proj/corum	3	Physical complexes
KEGG	http://www.genome.jp/kegg/pathway.html	833	Metabolic pathways, signaling pathways
CYC2008	http://wodaklab.org/cyc2008/	1 (<i>S. cerevisiae</i>)	Physical complexes
SGD Pathways	http://pathway.yeastgenome.org	1 (<i>S. cerevisiae</i>)	Metabolic pathways
MetaCyc	http://metacyc.org/	2,000	Metabolic pathways
Reactome	http://www.reactome.org/	20	Metabolic pathways

MATERIALS

EQUIPMENT

• Personal computer with Internet access and an Internet browser.

EQUIPMENT SETUP

Hardware requirements PanGIA hardware requirements depend on the size of the physical and genetic networks to be imported and analyzed. For networks containing up to 200,000 edges, we recommend a 2.0-GHz CPU or higher, a medium-end graphics card, 150 MB of available hard disk space and at least 2 GB of free physical RAM. If you are analyzing very large networks (>500,000 interactions), at least 8 GB of free physical RAM is recommended. To view the modular map produced by PanGIA, we recommend a monitor with a minimum screen resolution of 1024 × 768.

Operating system PanGIA and Cytoscape are supported on Windows (XP, Vista and Windows 7), Mac OS X (version 10.6 (Snow Leopard) or higher) and Linux. **Java standard edition** Version 1.6 or higher is required (can be downloaded from <http://www.java.com/>).

A three-button mouse This is recommended (but not required) as an aid in navigating the module map.

Cytoscape v2.8.0 PanGIA requires Cytoscape version 2.8.0 or higher. The steps for downloading and installing the latest version of Cytoscape can be found in a previously published protocol²⁴ or online at http://www.cytoscape.org/documentation_users.html.

Plug-ins The analysis capabilities of Cytoscape are expandable and extensible through add-on software packages called plug-ins. This protocol requires the installation of four plug-ins: PanGIA, BiNGO²⁹, Enhanced Search³⁰ and CyThesaurus³¹. Instructions for installing these plug-ins are outlined in PROCEDURE Steps 2–4.

MeV version 4.6 or higher MeV or MultiExperiment Viewer³² is an integrated toolkit for clustering and visualizing large-scale genomic data. This protocol uses MeV to view modules as a hierarchically clustered heat map. Instructions for downloading and installing MeV can be found at <http://www.tm4.org/mev/>.

Data files PanGIA requires both a physical and genetic network in a tab-delimited format (Box 1). Sample protein and genetic interaction networks are provided as examples to illustrate the protocol. The physical interaction network (Supplementary Data 1) was taken from a recent integration of two high-throughput protein interaction screens³³. Each physical interaction was assigned a Purification Enrichment score, with larger values representing greater confidence in the physical interaction. The genetic interaction network (Supplementary Data 2) was obtained from a large epistatic mini-array profile screen, which measured all possible genetic interactions among 743 genes involved in yeast chromosomal biology³. Each genetic interaction was assigned an S-score representing both the magnitude and confidence in

BOX 2 | THE NODE ATTRIBUTE FILE FORMAT

Annotation and physical complex/pathway data are read into Cytoscape using the node attribute file. This file maps individual genes to a given annotation or physical complex. The first line in this file represents the name of the annotation set being imported. Each subsequent line represents a mapping between a gene and an annotation and consists of two parts. The first part represents the gene or protein name. The second part represents the annotation to which the gene belongs to. If a gene maps to multiple annotations, annotations should be separated two colons (i.e., '::'). The two parts (gene and annotation name) are separated by an equal sign (i.e., '='). The annotation name is surrounded by parentheses. A sample file might look like this:

```
CYC2008
YDR473C = (U4/U6 × U5 tri-snRNP complex::TRAPP complex)
YOR373W = (SPB components)
```

The first line specifies the name of the annotation set being imported into Cytoscape. In this case, the annotation set is called 'CYC2008'. The gene YDR473C has been mapped to two annotations, 'U4/U6 × U5 tri-snRNP complex' and 'TRAPP complex'. The gene YOR373W has been mapped to a single annotation, 'SPB components'. The Supplementary Data contains a node attribute file named CYC2008_yeast_complexes.txt (Supplementary Data 3).

the interaction. Additional supplementary information can also be accessed at <http://prosecco.ucsd.edu/PanGIA/>. **Table 1** lists several public databases where protein and genetic interaction data can be downloaded for many different species.

Additional data files The file *CYC2008_yeast_complexes.txt* (**Supplementary Data 3**) contains a list of 408 protein complexes in the yeast *S. cerevisiae* hosted by the CYC2008 database^{34,35}. This file illustrates an example of a Cytoscape node attribute file, which allows nodes in a network to be mapped to a particular attribute (**Box 2**). In this case, yeast genes are mapped to the

various physical complexes in which they participate. This file is used to demonstrate how a set of known biological modules can be used to train PanGIA to identify more biologically meaningful modules and intermodule relationships (covered in the 'Training PanGIA' PROCEDURE subsection, Steps 27–29). Additionally, this file is used during the 'Module labeling' section of this protocol (Steps 30–32) to check if the identified modules correspond to known protein complexes. **Table 2** outlines several different public databases from which an annotation set can be downloaded for a variety of species.

PROCEDURE

Importing physical and genetic networks into Cytoscape

1 | Start Cytoscape. If Cytoscape is not yet installed on your computer, instructions for downloading and installing the latest version can be found at http://www.cytoscape.org/documentation_users.html. Cytoscape can be started by navigating to the directory in which it was installed and executing the file *cytoscape.bat* (Windows users) or *cytoscape.sh* (Linux and Mac OS X users).

▲ CRITICAL STEP PanGIA requires Cytoscape version 2.8.0 or higher. If your current installation of Cytoscape does not meet this requirement, download and install the latest version from <http://www.cytoscape.org/>.

? TROUBLESHOOTING

2 | Next, install the required plug-ins by navigating to the Plug-ins menu and clicking on Manage Plug-ins.

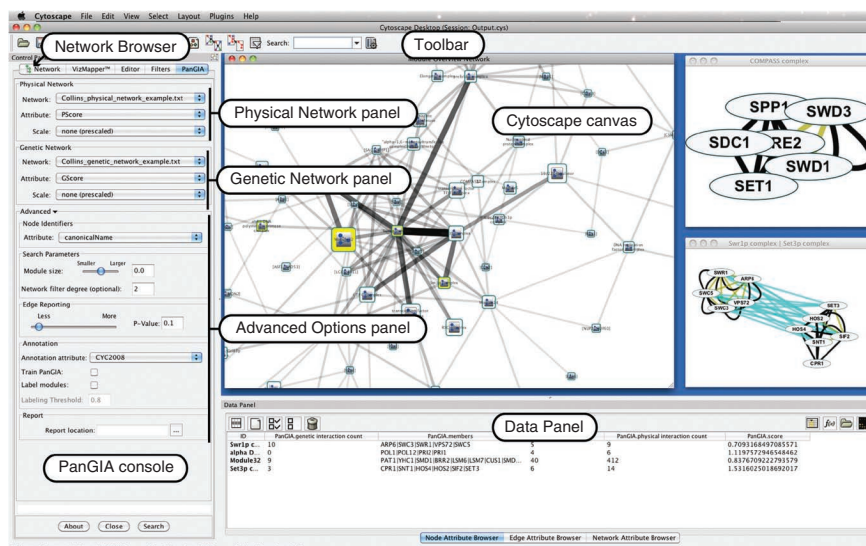
3 | Double-click on the Analysis folder located under the Available for Install folder and select the plug-in for PanGIA version 1.1 or later. Click Install. Accept the plug-in license agreement and then click Finish.

4 | Repeat the above step with BiNGO²⁹ version 2.42 or later (located in the Functional Enrichment folder), Enhanced-Search³⁰ version 1.2 or later (located in the Analysis folder) and CyThesaurus version 1.2 or later (located in the Network and Attribute I/O Folder).

5 | After installing the required plug-ins, start the PanGIA plug-in by navigating to the Plug-ins menu and selecting Module Finders → PanGIA.

6 | After PanGIA has started, the PanGIA console will appear (**Fig. 3**). The console is divided into three main panels: the Physical Network panel, where details regarding the physical network will be entered; the Genetic Network panel, where details regarding the genetic network will be entered; and the Advanced Options panel, which can be expanded by clicking on the triangle located next to the word 'Advanced'. This panel contains multiple advanced options for tuning the module-finding process. Four additional areas of interest are the Cytoscape canvas, which displays network visualizations and may be initially blank; the Data Panel, which is used to display node, edge and network attribute data; the Toolbar, which contains numerous command buttons; and the Network Browser, which can be accessed by clicking on the tab titled

Figure 3 | The PanGIA console. The Cytoscape canvas displayed the network data and may initially be blank. The Data Panel (bottom) is used to display node, edge and network attribute data. The Toolbar (top) contains numerous command buttons used for navigating the network. The PanGIA console (left) is divided into three main panels, including the Physical Network panel, the Genetic Network panel and the Advanced Options panel. The Network Browser may be accessed by clicking on the Network tab located to the left of the PanGIA console tab.



'Network' (**Fig. 3**). The Network Browser provides a list of networks currently available along with the number of nodes and edges in each network.

7| Next, we import both a physical and a genetic network to be used in the analysis. Assemble the data in a tab-delimited format as described in **Box 1**. Users wishing to follow this protocol as a tutorial should download the **Supplementary Data 1** (Collins_physical_network_example.txt) and **Supplementary Data 2** (Collins_genetic_network_example.txt) and continue with Step 8.

▲ **CRITICAL STEP** PanGIA is designed to work with both quantitative and nonquantitative interaction data. However, any single network (either physical or genetic) must consist of a single type of interactions (i.e., either all quantitative interactions or all non-quantitative interactions).

8| Click on the File menu, then select Import → Network from Table (Text/MS Excel). The Import Network and Edge Attributes from Table window will appear.

9| Click on the button titled 'Select File(s)' and specify the file containing the physical interaction network. A preview of the file should appear in the Preview panel located at the bottom. Select the column number representing the gene, which is the source node in the selection box titled 'Source Interaction'. Select the column number representing the target node in the Target Interaction selection box. If the example files (**Supplementary Data 1** and **Supplementary Data 2**) are being used, the source and target nodes are, respectively, columns 1 and 2.

10| Specify an interaction type that will enable Cytoscape to differentiate between protein and genetic interactions. Check the box titled 'Show Text File Import Options' and, under Network Import Options, enter a meaningful string character in the Default Interaction box (e.g., 'pi' or 'gi', depending on whether physical interactions or genetic interactions are being imported).

11| *Optional step:* Use this step if quantitative interaction strengths are attached to the network. In the Preview panel launched in Step 9, left-click the column, which represents the quantitative attribute under the Preview panel, to enable the import of this attribute into Cytoscape. Right-click the same column, and, when prompted, type in an appropriate Attribute name (e.g., PScore or GScore, depending on whether the physical or genetic network is being imported); click OK. Make sure to note the name used. You will need it later when selecting the attribute to be used in the training process. If the sample data are being used, the quantitative attribute for each interaction will be present in the third column.

▲ **CRITICAL STEP** The quantitative attribute provided should be either an integer (e.g., numbers such as 1, -2 or 514) or a floating point (e.g., numbers such as 2.343, -45.7687 or 74.3).

12| Click the Import button located in the lower right-hand corner. The physical network should now appear in the Cytoscape canvas area. The title of the network should be the name of the file provided.

13| Repeat Steps 8–12 to import the genetic network.

14| *Optional step:* Steps 14–18 should be used if the physical and genetic networks use different gene identifier systems (e.g., UniProt ID versus Ensembl ID). PanGIA requires that the two networks use the same gene identifier system. To convert between two gene identifier systems, assemble an ID translation file into a tab-delimited format as described in **Box 3**. This file should contain a map between the gene identifier system currently being used and the target gene identifier system. Users following this protocol as a tutorial using the sample data provided should skip to Step 19.

15| *Optional step:* Start the CyThesaurus plug-in by clicking on the Plug-ins menu and then selecting CyThesaurus. A window titled 'CyThesaurus plug-in' should appear.

16| *Optional step:* Configure the CyThesaurus plug-in to use the ID mapping file generated in Step 14 by clicking on ID Mapping Resources Configuration. A new window titled 'ID Mapping Source Configuration' will open up. In the left panel of this window, click on the folder titled 'Local Remote Files', which will bring up another window titled 'File-based ID Mapping Resources Configuration'. Under the panel named 'Data source', click Select file to specify the location of the ID mapping file. Click on Open, then OK, and finally Close.

17| *Optional step:* Select both the physical and genetic networks by clicking on them in the Available Networks panel, then click the right arrow button. The two networks will appear in the Selected Networks panel.

BOX 3 | THE TAB-DELIMITED ID MAPPING FILE

PanGIA requires that the gene identifiers used in the physical and genetic network are of the same type. If the gene identifier used in these two networks differ from one another (e.g., the genetic network uses the Entrez gene identifier, whereas the physical network uses the Ensembl gene identifier), the CyThesaurus plug-in can be used to map between gene identifiers. As input, the plug-in requires a tab-delimited file that provides a mapping between the two identifiers. The first line in this file provides the names of the two different gene identifiers. Each subsequent line consists of two tokens separated by a tab character. The first token represents a gene described using the first gene identifier, whereas the second token represents its corresponding identity in the second identifier.

An example file might look like this:

```
Ensembl Gene ID    UniProt Gene ID
ENSG00000211890   IGHA2_HUMAN
ENSG00000211891   IGHE_HUMAN
```

The first line specifies the name of the two gene identifiers, Ensembl Gene ID and UniProt Gene ID. Each subsequent line provides a mapping between these two gene identifiers for a single gene.

18| *Optional step:* Choose the two different gene identifier names used in the genetic and physical network in the Source ID Type(s) selection box. In the Target ID Type selection box choose the target gene identifier you wish to map to. Finally, in the selection box titled 'All target ID(s) or first only?' select the option to keep the first target ID only. Next, click OK. A message will pop up indicating how many gene identifiers were successfully mapped.

Generating a module map using the PanGIA plug-in: selecting the physical and genetic network

19| In the uppermost panel in the PanGIA console (Physical Network panel, see **Fig. 3**), select the physical network to be used in the Network selection box. The name of the physical network will correspond to the name of the file from which the network was imported.

20| Select the genetic network to be used in the Network selection box located in the Genetic Network panel. Again, the name of the network will correspond to the name of the file from which it was imported.

21| *Optional step:* Use this step if quantitative interaction data are being used. In the Attribute drop-down menu located in the Physical Network panel, select the appropriate attribute name (i.e., the name assigned to the quantitative attribute for physical interactions from Step 11). Similarly, select the appropriate attribute name for genetic interactions in the Attribute drop-down menu located in the Genetic Network panel.

22| *Optional step:* Use this step if quantitative interaction data are being used and no biological annotation data are present. Even without a set of known complexes or pathways, PanGIA can leverage the confidence values assigned to each interaction (physical or genetic) to identify modules and intermodule links that contain highly confident interactions. However, it is necessary to let PanGIA know how the quantitative information is scaled. In the Scale selection menu located in both the Physical Network and Genetic Network subpanels (**Fig. 3**), choose one of the following options: 'lower'—this option indicates that smaller quantitative values (both positive and negative) represent more confident interactions; 'upper'—this option indicates that larger quantitative values (both positive and negative) represent more confident interactions; or 'none (prescaled)'—this option should only be chosen if the quantitative attribute attached to either the physical or genetic interactions already represents the likelihood that a given interaction falls within a known biological module. This option enables the user to perform the training procedure outside of PanGIA and use the subsequent results in the module search process. If the example files are being used, simply choose 'none'. During the training process, PanGIA will automatically scale the score attached to each interaction to reflect how likely that interaction is to fall either within a module or between two modules.

23| *Optional step:* Use this step if the gene identifiers in either the physical or genetic network were mapped to a new gene identifier. In the Advanced Options panel, select the target gene identifier to which genes in both networks were mapped to under the Node Identifiers subpanel. If no gene identifier mapping was performed or if the user is following this protocol with the sample data, skip to Step 24.

Generating a module map using the PanGIA plug-in: setting the module size and edge reporting parameters (optional)

24| *Optional step:* PanGIA features a number of advanced options for tuning the search process. The size and number of modules returned by the search process can be controlled by changing the Module Size parameter (located in the Advanced Options panel). This can be done using the graphical slider in the Search Parameters panel. Dragging the slider to the right

will result in fewer modules with larger average size, while dragging the slider to the left will result in more modules with a smaller average size (**Fig. 1b**). The value of the Module Size parameter will be displayed in a text box to the right of the slider. It is recommended to leave the slider in its default position for the first run and to adjust it later if the results are unsatisfactory. For the sample data provided, set the Module Size parameter to -1.6 by moving the slider to the left.

25| Optional step: Often, the physical network being used covers a much larger set of proteins than those examined in the genetic interaction screen. In such a case, it is often useful to trim the physical network to include only proteins that are either present in the genetic network or are neighbors of such proteins within the physical network. This trimming is controlled by setting the 'network filter degree' parameter (located in the Advanced Options panel). A value of 0 will trim the physical network to only include nodes from the genetic network. Higher values represent the acceptable distance (through edges) separating a protein in the physical network from a node in the genetic network. If no trimming is desired, leave the box blank to prevent PanGIA from filtering any nodes. If the sample data file is being used, leave the network filter degree parameter at its default value of two.

▲ CRITICAL STEP The network filter degree parameter provided should be a positive integer (e.g., numbers such as 1, 2 or 10).

26| Optional step: Use this step only if quantitative interaction data are present. Every intermodule link found by PanGIA can be assigned a *P* value, after which insignificant edges are filtered from the resulting module map. The significance threshold can be set by changing the position of the slider in the Edge Reporting subpanel. Dragging the slider to the left (toward 'Less') will result in a higher significance threshold and less intermodule links in the final map (**Fig. 1c**). The *P* value cutoff will be displayed in a text box immediately to the right of the slider. If the example files are being used, move the slider to the left and set the threshold to 0.05.

Generating a module map using the PanGIA plug-in: training PanGIA (optional)

27| Optional step: Steps 27–29 should be used only if an annotation set is present. The training and module labeling steps require a list of annotations to be imported into Cytoscape. Assemble your list of annotations into the node attribute file format as described in **Box 2**. Import this file into Cytoscape by navigating to File → Import → Node Attribute.... Navigate to the appropriate file and click Open. If using the sample data, the file *CYC2008_yeast_complexes.txt* (**Supplementary Data 3**) should be used in this step.

28| Optional step: In the Annotation subpanel under Advanced Options, select the annotation attribute that will be used during the training and labeling process. The name of the annotation set is specified in the node attribute file, which was uploaded in the previous step (see **Box 2** for more details). If the sample data have been used, the attribute name will be CYC2008. Select the annotation set name in the selection box titled Annotation attribute.

29| PanGIA can be trained to better identify module and intermodule links by examining actual examples of biological modules provided in the annotation set. To train PanGIA, simply check the box titled 'Train PanGIA' in the Annotation subpanel. If the sample data are being used, make sure this box is checked.

Generating a module map using the PanGIA plug-in—labeling modules (optional)

30| Optional step: This step should only be used if an annotation set is present. PanGIA can label individual modules with the name of an annotation, if their member genes overlap with the genes belonging to that annotation (**Fig. 1c**). To have PanGIA label modules, check the Label modules box in the Annotation subpanel (**Fig. 3**). Next, specify the overlap threshold (defined here as the Jaccard index) in the Labeling Threshold text box. If the sample data are being used, set the Labeling Threshold to 0.2.

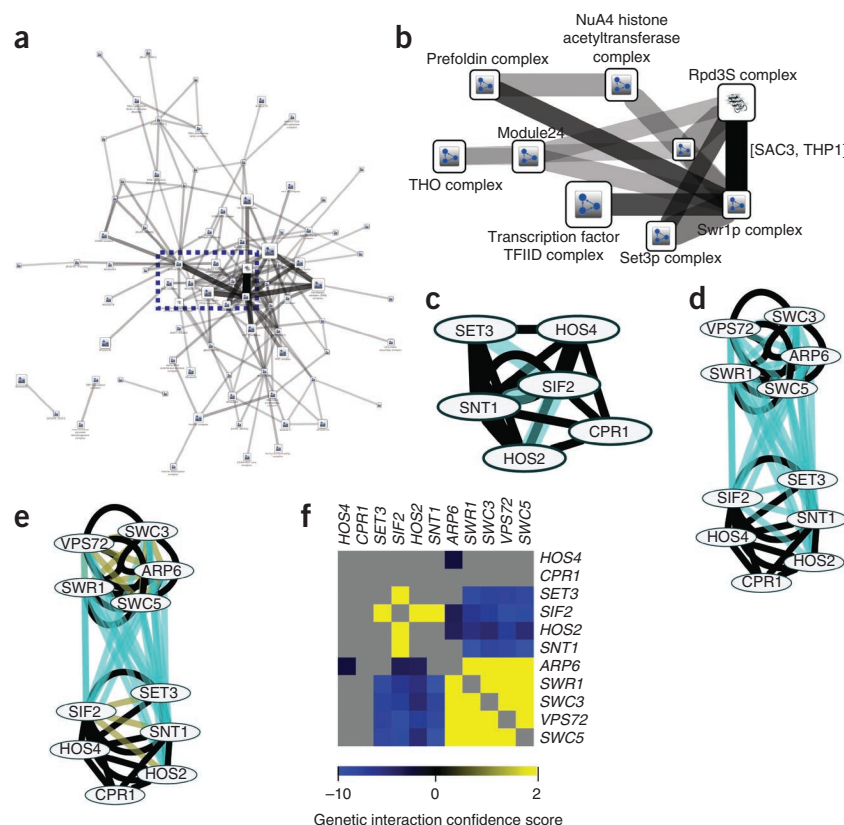
? TROUBLESHOOTING

31| Optional step: If desired, PanGIA can output a report containing a summary of the module-finding process. This includes a summary of the networks used by PanGIA, the results of the training process and a summary of the resulting module map. To have PanGIA output a report, specify an output file in the Report subpanel. After a successful search, an HTML file will be created, which can be viewed using any Internet browser.

32| At this point, PanGIA is fully configured. The module search process can be initiated by clicking the Search button located at the bottom-right corner of the PanGIA console. Depending on the size of the network and the computer hardware, the module-finding process should take anywhere from 1 to 10 min. If the sample data are being used, the search process should take less than 1 min.

? TROUBLESHOOTING

Figure 4 | PanGIA output. **(a)** The module map returned by PanGIA. Each node is a separate module or complex and the area of the node reflects the number of genes contained within the module. **(b)** A zoomed-in portion (blue box) of the module map shown in **a**. If an annotation set was provided and the labeling option was chosen, modules which overlap substantially with an annotation are labeled as such (e.g., Rpd3S complex). Modules not overlapping with any of the provided annotations are either given a generic name (e.g., Module 24) or labeled with a gene name (e.g., [SAC3,THP1]) if the module contains only one or two genes. **(c)** A detailed view of a single module. Each node represents a single gene that was assigned to this module. Physical interactions are colored black, whereas genetic interactions are colored turquoise. **(d)** A detailed view for two modules. Edges are colored similarly to **c**. The layout algorithm seeks to physically separate each module. **(e)** The same detailed view of two modules as shown in **d**, except that positive genetic interactions are colored yellow, whereas negative genetic interactions are colored turquoise. **(f)** The same network as shown in **e**, but visualized as a hierarchically clustered heatmap using MeV³².



Visualization of the module map using nested networks: navigating the module map

33 | Once the search process is complete, a window titled ‘Module Overview Network’ will appear in the Cytoscape Canvas panel (**Fig. 4a**). This network is the resulting global module map. Each node represents an individual module composed of a set of genes densely interconnected by genetic and physical interactions. The area of a module scales according to the number of genes that it contains. Links between modules are composed of genetic interactions; the thickness of the interactions corresponds to the number of genetic interactions spanning the two modules. If the labeling option was chosen, modules that overlap with one of the annotations provided will be labeled as such (**Fig. 4a,b**).

34 | You can zoom into the module map using the Zoom In button on the toolbar. This icon is displayed as a magnifying glass with a ‘+’ symbol in the middle. You can zoom out by clicking on the Zoom Out button (magnifying glass with a ‘-’ symbol in the middle). Alternatively, you can zoom in and out using the scroll wheel on the mouse. Scrolling up zooms into the area centered on the mouse pointer. Scrolling down zooms out on the area centered on the mouse pointer.

35 | To pan around the module map, two options are available—using the mouse (option A) or using the network browser (option B):

(A) Using the mouse

- (i) Click the middle button on the mouse (or the scroll wheel, if present) anywhere in the active network being viewed in the Cytoscape canvas and drag the mouse in the desired direction.

(B) Using the network browser

- (i) Navigate to the ‘Network Browser’ by clicking on the Network tab (**Fig. 3**) located to the left of the PanGIA tab. In the bottom half of the Network Browser is a bird’s-eye view of the active network being viewed in the Cytoscape canvas; a blue selection box highlights the particular region of the network currently being viewed. To pan around the network, click and hold the blue selection box and move it in the desired direction.

Visualization of the module map using nested networks—identifying modules of interest

36 | To further investigate modules of interest (i.e., function enrichment or detailed visualization), the module or modules of interest must be selected. We describe three different options for doing so: direct selection of modules (option A), direct selection of intermodule links (option B) and search-based selection of modules (option C).

PROTOCOL

(A) Direct selection of modules

- (i) Select any single module by clicking on it with the left mouse button. The selected module will turn yellow. Several modules can be selected by holding down and dragging the left mouse button to define a rectangular selection region. Alternatively, multiple modules may be selected by holding down the shift button and left-clicking on multiple modules.

(B) Direct selection of intermodule links

- (i) To select any edge, click on the edge with the left mouse button. The selected edge will turn red. Several edges can be selected by holding down and dragging the left mouse button to define a rectangular selection region.

(C) Search-based selection of modules

- (i) To find and highlight modules in the map that contain a gene of interest, enter the name of the gene into the Enhanced Search plug-in search box located in the command toolbar (**Fig. 3**). If your gene of interest falls within a module, that module and its intermodule links will be highlighted yellow.

Visualization of the module map using nested networks—exploring modules of interest

37 | PanGIA returns numerous useful statistics or attributes regarding the modules identified, including module size, number of physical/genetic interactions among the genes in this module and so on. A complete list of attributes returned by PanGIA is provided in **Table 3**. The Data Panel (**Fig. 3**) can display any/all of the attributes listed in **Table 3**. Select a module(s) of interest from the module map displayed in the Cytoscape Canvas as described in Step 36. When a single module or groups of modules have been selected in the Cytoscape Canvas, the selected modules will be listed in the Data Panel (**Fig. 3**). Next, click on the Select Attributes button located in the upper left corner of the Data Panel. This will cause a list of attributes to appear; select which attributes you wish to view by clicking on their name. Exit this menu by clicking anywhere else.

38 | The Data Panel can also display detailed information regarding intermodule links in the map. Select one or more intermodule links of interest in the map as described in Step 36. In the Data Panel, click on the tab labeled Edge Attribute Browser. The panel will display the edges that have been selected. Similar to the modules, intermodule links identified by PanGIA also have several informative attributes as outlined in **Table 3**. These attributes can be viewed by selecting them through the Select Attributes menu (see Step 37).

39 | To visually inspect a single module or a group of modules in greater detail, select the module(s) of interest as outlined in Step 36. Next, right-click any of the selected module(s) and choose PanGIA → Create Detailed View. A new window will appear in the Cytoscape Canvas area containing the module (**Fig. 4c**) or modules (**Fig. 4d**) of interest. In this detailed view, each node represents a single gene. Edges represent either physical interactions (colored black) or genetic interactions (colored turquoise). If quantitative genetic interaction data are used, positive genetic interactions will be colored yellow, whereas negative genetic interactions will be colored turquoise (**Fig. 4e**).

TABLE 3 | Description of module-level attributes returned by PanGIA.

Attribute name	Attribute type (node or edge)	Description
PanGIA Member Count	Node	Number of genes present in module
PanGIA Module Physical Interaction Count	Node	Number of physical interactions present in this module
PanGIA Module Genetic Interaction Count	Node	Number of genetic interactions present in this module
PanGIA Source Size	Edge	Member count of the source module
PanGIA Target Size	Edge	Member count of the target module
PanGIA Genetic Interaction Count	Edge	Number of genetic interactions spanning the two modules connected by this edge
PanGIA Physical Interaction Count	Edge	Number of physical interactions spanning the two modules connected by this edge
PanGIA <i>P</i> value	Edge	Significance of the intermodule link
PanGIA Edge Score	Edge	The total score of genetic interactions spanning two modules minus the score of the physical interactions
PanGIA Genetic Interaction Density	Edge	Represents the Edge Score divided by the Genetic Interaction Count

40 | The network displayed in the detailed view can be laid out and manipulated similarly to the module map as described in Steps 33–35. Individual genes and interactions between genes can be selected similarly to the way in which modules are selected in the module map as described in Step 36.

41 | *Optional step:* Steps 41–44 should be followed if quantitative interaction data are present. An alternate means of visualizing a single module or a set of connected modules is via a hierarchically clustered heat map (**Fig. 4f**). In this view, each row or column represents a single gene. Each cell in the matrix is colored to represent the quantitative value attached to the interaction between those two genes. For example, **Figure 4f** is a hierarchically clustered representation of the between-cluster model shown in **Figure 4e**. The colors in the heat map represent the genetic interaction confidence scores between the genes. PanGIA can output a matrix containing either the genetic interaction confidence scores or physical interaction confidence scores between individual genes (option A), between all genes in a module or set of modules (option B):

(A) Output interaction matrix for a select number of genes

- (i) Select the genes of interest from a detailed view as described in Step 36. Right-click on any of the selected genes and select PanGIA → Save Selected Nodes to Matrix File.
- (ii) Next, choose the desired quantitative attribute to be outputted (i.e., physical interaction confidence or genetic interaction confidence). The names of these quantitative attributes will be the ones assigned by the user in Step 11.
- (iii) A dialog box will appear prompting to you enter the output file name. Enter the file name and click Save.

(B) Output interaction matrix for all genes in a module or set of modules

- (i) Select a module(s) of interest as outlined in Step 36. Right-click on any of the selected modules and select PanGIA → Save Selected Nodes to Matrix File.
- (ii) Choose the desired attribute to be outputted. Enter the output filename and click Save. If you are using the Sample data, select the modules labeled 'Swr1p complex' and 'Set3p complex'. Right-click on one of these two modules and select PanGIA → Save Selected Nodes to Matrix File → GScore.
- (iii) Provide an appropriate file name and click Save.

42 | *Optional step:* Start the MeV program. The Multiple Array Viewer window should pop up. Load the interaction matrix generated in the previous step by navigating to File → Load Data. The Expression File Loader dialog window will appear. Click the 'Browse' button and specify the file containing the interaction matrix. A preview of the interaction matrix should appear in the Expression Table panel. Click the upper-leftmost interaction confidence score and then click Load. A heat map of the interaction matrix will appear in the Multiple Array Viewer window.

43 | *Optional step:* To hierarchically cluster the heat map, click on the Clustering tab located near the top of the window and then select Hierarchical Clustering. In the HCL: Hierarchical Clustering window that will open, check the boxes to Optimize Gene Leaf Order and Optimize Sample Leaf Order. This will ensure that genes with similar interaction profiles will be placed close to one another. Finally, click OK.

44 | *Optional step:* In the rightmost panel of the Multiple Array Viewer navigate to Analysis Results → HCL (1) → HCL Tree. A hierarchically clustered version of the heat map will appear. This image can be saved by clicking on File → Save Image. Multiple output formats are available. If using the example data, the heat map should look similar to **Figure 4f**.

45 | In cases in which a module may contain one or more genes with an unknown function, it is useful to be able to query an external web-based database such as Ensembl or Entrez. Cytoscape features the ability to automatically connect to and query external web databases. Right-click on a gene of interest within the Detailed View and navigate to the LinkOut menu. Numerous databases will be listed including Ensembl, KEGG, UniProt and Entrez. Select one of these databases. An Internet browser window will open automatically displaying any information the selected database has on the gene of interest. This feature provides an effective way to interrogate the function of unannotated genes.

? TROUBLESHOOTING

Functional enrichment of the modules

46 | Start the BiNGO plug-in by selecting Plug-ins → Start BiNGO. The BiNGO Settings window will appear.

47 | Select the module or modules of interest that will be examined for an enriched function. Create a Detailed View as outlined in Step 39. Select the genes contained in the module(s) that will be screened for an enriched GO function. To select all genes, simply press Ctrl + A simultaneously (or Command + A, if using Mac OS X).



PROTOCOL

48| Type in a meaningful name for the set of genes being examined in the box titled ‘Cluster name’. Under the Select Organism/Annotation menu, choose the appropriate organism (for the sample data choose *Saccharomyces cerevisiae*). For the remaining options, the default values will typically suffice. Click Start BiNGO. Depending on the number of genes selected and the computer hardware, this process will take 5–10 min.

? TROUBLESHOOTING

49| BiNGO will return an output window containing a list of GO terms that were found to be enriched along with their respective *P* values. BiNGO will also return a network of GO terms showing the inter-relationships between the various GO terms that were found to be enriched. The color of each term represents its significance of enrichment.

Exporting your results

50| Cytoscape enables multiple ways to export individual modules as well as the global module map. For a thorough explanation of each of these export methods, please refer to the online tutorial (<http://www.cytoscape.org/documentation/users.html>). *Note:* for general troubleshooting and timing advice, please refer to **Tables 4** and **5**.

(A) Export network as a graphics object

- (i) The module map, as well as individual modules, can be exported as a graphics file. Numerous output formats are supported including PDF, JPEG, SVG, PNG and BMP.
- (ii) To export a network as a graphics object, make sure it is the active window and then select File → Export → Network View as Graphics....
- (iii) In the Export Network View as Graphics dialog box, select the output file name and choose the desired output format. Click OK.
- (iv) If the graphics object will be further manipulated in a graphics software package, such as Adobe Illustrator, we recommend exporting the network as a PDF file. Make sure to also check the box titled ‘Export text as font’, which will enable the manipulation of the text labels in the network image.

(B) Export modules as a tab-delimited file

- (i) Each of the individual modules can be exported in a tab-delimited file, where each line consists of two parts separated by a tab character: the name of the module and the genes comprising the module. If multiple genes have been assigned to a module, each gene will be separated by the ‘|’ character.

TABLE 4 | Troubleshooting table.

Step	Problem	Possible reason	Solution
1	Executing cytoscape.bat (Windows) or Cytoscape.sh (Mac OS X, Linux) does not open Cytoscape	Java is not installed properly	Make sure Java version 1.6.014 or higher is installed. Java can be downloaded at http://www.java.com/
30	PanGIA does not label any of the modules in the final module map	Threshold for labeling may be set too high	Set the labeling threshold slightly lower to allow more modules to be labeled
32	The module search process is taking a very long time	Insufficient memory and/or processing power	Very large physical or genetic networks (>500,000 interactions) require a larger amount of memory than specified in the EQUIPMENT SETUP section. See the TIMING section for recommendations on the amount of memory and processing power required for larger networks
45	The queried database does not return any information on the selected gene(s) of interest	Mismatched gene identifiers	When querying an external database, the identifier of the selected gene(s) must be identical to the identifier used by the external database. For example, if querying the Ensembl database, selected genes need to use Ensembl identifiers in order to have any information returned.
48	BiNGO supplies an error message asking to ‘Please select one or more nodes.’	No genes were selected for examining functional enrichment	Visualize the module(s) of interest as outlined in Step 39. In the detailed view, select one or more genes of interest. All nodes (genes) can be selected in a detailed view by pressing Ctrl (or Cmd, if using Mac OS X) + A

- (ii) To export the modules as a tab-delimited file, right-click on any module in the module map (i.e., the Module Overview Network in the Cytoscape Canvas) and select PanGIA → Export → Export Modules to Tab-Delimited file.
- (iii) Specify the output file in the dialog box that pops up and click Save.

(C) Export module map as a tab-delimited file

- (i) The entire module map can be exported as a tab-delimited file, where each single line represents a single interaction between two modules. A single line is split into nine different parts separated by a tab character. The first two parts represent the source and target module. The remaining seven parts represent various attributes describing each interaction as outlined in **Table 3**.

- (ii) To export the module map as a tab-delimited file, right-click on any module in the module map (i.e., the Module Overview Network in the Cytoscape Canvas) and select PanGIA → Export → Export Module Map to Tab-Delimited file.
- (iii) Specify the output file and click Save.

(D) Export the entire PanGIA session as a Cytoscape session file

- (i) The entire PanGIA session can be saved to file. A session file contains all of the results of this entire workflow. This includes all networks that were loaded or generated (physical, genetic, module map, individual modules), any custom visualization styles that were employed and any enrichment results obtained from BiNGO. Saving to a session file will enable the user to continue the analysis at a later point.
- (ii) To save the entire PanGIA session to file, select File → Save As. Type in the name of the output file and click Save.

? TROUBLESHOOTING

Troubleshooting advice for specific steps in the protocol can be found in **Table 4**. In addition, we outline two of the biggest problems a user may face and potential solutions to these problems below:

Module size issues

In some cases PanGIA may fail to return any modules or it may return modules that are either very large or very small (i.e., that consist of a single gene). The problem may be addressed by moving the Module Size slider bar in the Advanced Options panel (see Step 24). Dragging the slider to the right will generally result in fewer but larger modules. Dragging it to the left will have the opposite effect. Once the slider has been set to a new position, make sure the rest of PanGIA is properly configured (Steps 19–31) and hit the Search button located at the bottom of the PanGIA console.

Edge reporting issues

Another common issue is that the module map may contain either too few or too many intermodule links. PanGIA utilizes a sampling-based procedure to assign *P* values to every intermodule link and only those links with a *P* value below a specified threshold are displayed in the final module map. If the threshold is set too high, this may cause a number of spurious interactions to appear in the module. On the other hand, if the threshold is set too low, this may cause PanGIA to filter out intermodule links of biological interest. This problem may be addressed by adjusting the threshold by moving the Edge Reporting slider bar in the Advanced panel (as described in Step 26). Moving the slider to the right will result in a higher threshold and subsequently a larger number of intermodule links in the final map. Moving it to the left will have the opposite effect.

● TIMING

The time required to complete this protocol is almost entirely dependent on the size of the genetic and physical networks being analyzed. **Table 5** charts the amount of time required for the module-search process (under default options) using networks of various sizes as input. For a physical and genetic network containing less than 100,000 interactions each (~200,000 interactions total), PanGIA takes, on average, ~10 min.

TABLE 5 | Time required to run PanGIA on networks of various sizes.

Number of interactions (genetic + physical)	Run time	
	Processor: dual-core, 32-bit (3.2 GHz) Memory: 2 GB Graphics card memory: 256 MB	Processor: 8-core, 64-bit (2.8 GHz) Memory: 8 GB Graphics card memory: 256 MB
10,000	< 1 min	< 30 s
50,000	1 min	< 1 min
100,000	2 min	1.5 min
500,000	15 min	10 min
1,000,000	Insufficient memory	30 min

ANTICIPATED RESULTS

Using the sample physical (**Supplementary Data 1**) and genetic (**Supplementary Data 2**) interaction networks with PanGIA, configured as suggested in this protocol (module size parameter = -1.6 , edge filtering parameter = 0.05 , network filter = 2 , training enabled, labeling threshold = 0.2), will produce a module map containing 82 modules and 164 intermodule links (**Fig. 4a**). Overall, 34 of these modules overlap with known complexes provided in the file CYC2008_yeast_complexes.txt (**Supplementary Data 3**) and will be labeled accordingly.

The resulting module map provides a wealth of hypotheses that can form the basis for follow-up experiments. Because PanGIA has been trained on databases of known complexes and pathways, it is likely that many modules will correspond to known protein complexes in the PanGIA results^{18–20}. Other modules that do not correspond to prior knowledge are prime candidates for novel complexes or pathways. The module map produced using the sample data contains 21 modules (out of 82) with two or more genes that do not overlap with any known *S. cerevisiae* physical complexes. One could test the members of these 21 modules for co-complex membership. An alternate strategy for revealing novel biological functions is to identify modules that are enriched for a common biological function, yet contain some genes that are not yet annotated to that particular function. For example, Module 24 (**Fig. 4b**) is enriched for genes involved in nuclear pore organization ($P < 7.05 \times 10^{-11}$). However, two of the genes in Module 24, SEC31 and SEC16, are not annotated to this function. The logical hypothesis in this case would be that these two genes are involved in nuclear pore organization and that a deletion or knockdown of these genes should have an impact on this function.

Intermodule links, on the other hand, predict functional overlap or synergy between the two connected modules^{18,20}. For example, a large number of genetic interactions span the two modules corresponding to the Rpd3S complex and Swr1p complex (**Fig. 4d,e**). The Swr1p complex has been well established as a chromatin remodeler, which deposits H2A.Z, a histone variant, onto chromatin. The function of the Set3p complex is less well understood. The intermodule link between the two complexes suggests that Set3p may have a role similar to that of the Swr1p complex. Indeed, a recent publication has provided evidence suggesting that this may be the case³⁶.

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

ACKNOWLEDGMENTS We gratefully acknowledge S. Bandyopadhyay and R. Kelley for their role in the development of the framework used in PanGIA. M. Michaut provided useful feedback on the manuscript. C. Doherty and M. Ashkenazi provided helpful beta testing of the PanGIA plug-in. This study was supported by grants from the National Institute of General Medical Sciences (GM070743), the National Science Foundation (IIS0803937) and Microsoft (Computational Challenges in Genome-wide Association Studies).

AUTHOR CONTRIBUTIONS G.H., R.S. and T.I. conceived and led the project. G.H. coded PanGIA with supporting code from R.S., J.R., K.O., P.-L.W. and M.S. R.S., G.H. and T.I. wrote the paper. All authors have contributed to the design of PanGIA and all have read and approved the paper.

COMPETING FINANCIAL INTERESTS The authors declare no competing financial interests.

Published online at <http://www.natureprotocols.com/>.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Boone, C., Bussey, H. & Andrews, B. Exploring genetic interactions and networks with yeast. *Nat. Rev. Genet.* **8**, 437–49 (2007).
- Beyer, A., Bandyopadhyay, S. & Ideker, T. Integrating physical and genetic maps: from genomes to interaction networks. *Nat. Rev. Genet.* **8**, 699–710 (2007).
- Collins, S. *et al.* Functional dissection of protein complexes involved in yeast chromosome biology using a genetic interaction map. *Nature* **446**, 806–810 (2007).
- Schuldiner, M. *et al.* Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell* **123**, 507–519 (2005).
- Fiedler, D. *et al.* Functional organization of the *S. cerevisiae* phosphorylation network. *Cell* **136**, 952–963 (2009).
- Tong, A. *et al.* Global mapping of the yeast genetic interaction network. *Science* **303**, 808–813 (2004).
- Stranger, B.E. *et al.* Population genomics of human gene expression. *Nat. Genet.* **39**, 1217–1224 (2007).
- Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
- Bandyopadhyay, S. *et al.* Rewiring of genetic networks in response to DNA damage. *Science* **330**, 1385–1389.
- Schuldiner, M., Collins, S.R., Weissman, J.S. & Krogan, N.J. Quantitative genetic analysis in *Saccharomyces cerevisiae* using epistatic miniarray profiles (E-MAPs) and its application to chromatin functions. *Methods* **40**, 344–352 (2006).
- Collins, S.R., Schuldiner, M., Krogan, N.J. & Weissman, J.S. A strategy for extracting and analyzing large-scale quantitative epistatic interaction data. *Genome Biol.* **7**, R63 (2006).
- Costanzo, M. *et al.* The genetic landscape of a cell. *Science* **327**, 425–431.
- Pan, X. *et al.* A DNA integrity network in the yeast *Saccharomyces cerevisiae*. *Cell* **124**, 1069–1081 (2006).
- Schlabach, M.R. *et al.* Cancer proliferation gene discovery through functional genomics. *Science* **319**, 620–624 (2008).
- Bakal, C. *et al.* Phosphorylation networks regulating JNK activity in diverse genetic backgrounds. *Science* **322**, 453–456 (2008).
- Breitkreutz, B.J. *et al.* The BioGRID Interaction Database: 2008 update. *Nucleic Acids Res.* **36**, D637–D640 (2008).
- Zhang, L.V. *et al.* Motifs, themes and thematic maps of an integrated *Saccharomyces cerevisiae* interaction network. *J. Biol.* **4**, 6 (2005).
- Kelley, R. & Ideker, T. Systematic interpretation of genetic interactions using protein networks. *Nat. Biotechnol.* **23**, 561–566 (2005).
- Ulitsky, I. & Shamir, R. Pathway redundancy and protein essentiality revealed in the *Saccharomyces cerevisiae* interaction networks. *Mol. Syst. Biol.* **3**, 104 (2007).
- Bandyopadhyay, S., Kelley, R., Krogan, N. & Ideker, T. Functional maps of protein complexes from quantitative genetic interaction data. *PLoS Comput. Biol.* **4**, e1000065 (2008).
- Sharma, V.M., Tomar, R.S., Dempsey, A.E. & Reese, J.C. Histone deacetylases RPD3 and HOS2 regulate the transcriptional activation of DNA damage-inducible genes. *Mol. Cell. Biol.* **27**, 3199–3210 (2007).
- Jaimovich, A., Rinott, R., Schuldiner, M., Margalit, H. & Friedman, N. Modularity and directionality in genetic interaction maps. *Bioinformatics* **26**, i228–i236.
- Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
- Cline, M.S. *et al.* Integration of biological networks and gene expression data using Cytoscape. *Nat. Protoc.* **2**, 2366–2382 (2007).

25. Smoot, M.E., Ono, K., Ruscheinski, J., Wang, P.L. & Ideker, T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* **27**, 431–432.
26. Wilmes, G.M. *et al.* A genetic interaction map of RNA-processing factors reveals links between Sem1/Dss1-containing complexes and mRNA export and splicing. *Mol. Cell.* **32**, 735–746 (2008).
27. Roguev, A. *et al.* Conservation and rewiring of functional modules revealed by an epistasis map in fission yeast. *Science* **322**, 405–410 (2008).
28. Ashburner, M. *et al.* Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* **25**, 25–29 (2000).
29. Maere, S., Heymans, K. & Kuiper, M. BiNGO: a Cytoscape plug-in to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* **21**, 3448–3449 (2005).
30. Ashkenazi, M., Bader, G.D., Kuchinsky, A., Moshelion, M. & States, D.J. Cytoscape ESP: simple search of complex biological networks. *Bioinformatics* **24**, 1465–1466 (2008).
31. van Iersel, M.P. *et al.* The BridgeDb framework: standardized access to gene, protein and metabolite identifier mapping services. *BMC Bioinformatics* **11**, 5 (2010).
32. Saeed, A.I. *et al.* TM4 microarray software suite. *Methods Enzymol.* **411**, 134–193 (2006).
33. Collins, S.R. *et al.* Toward a comprehensive atlas of the physical interactome of *Saccharomyces cerevisiae*. *Mol. Cell Proteomics* **6**, 439–450 (2007).
34. Pu, S., Vlasblom, J., Emili, A., Greenblatt, J. & Wodak, S.J. Identifying functional modules in the physical interactome of *Saccharomyces cerevisiae*. *Proteomics* **7**, 944–960 (2007).
35. Pu, S., Wong, J., Turner, B., Cho, E. & Wodak, S.J. Up-to-date catalogues of yeast protein complexes. *Nucleic Acids Res.* **37**, 825–831 (2009).
36. Hang, M. & Smith, M.M. Genetic analysis implicates the Set3/Hos2 histone deacetylase in the deposition and remodeling of nucleosomes containing H2A.Z. *Genetics*. **187**, 1053–1066 (2011).