

# Chapter 18

## Displaying Chemical Information on a Biological Network Using Cytoscape

Iain M. Wallace, Gary D. Bader, Guri Giaever, and Corey Nislow

### Abstract

Cytoscape is an open-source software package that is widely used to integrate and visualize diverse data sets in biology. This chapter explains how to use Cytoscape to integrate open-source chemical information with a biological network. By visualizing information about known compound–target interactions in the context of a biological network of interest, one can rapidly identify novel avenues to perturb the system with compounds and, for example, potentially identify therapeutically relevant targets. Herein, two different protocols are explained in detail, with no prior knowledge of Cytoscape assumed, which demonstrate how to incorporate data from the ChEMBL database with either a gene–gene or a protein–protein interaction network. ChEMBL is a very large, open-source repository of compound–target information available from the European Molecular Biology Laboratory.

**Key words:** Cytoscape, Network visualization, Druggable targets, Chemical biological networks, Chemical networks

---

### 1. Introduction

Cytoscape is an open-source software package that is used to visualize biological data as an interaction network (1). In such a biological network, nodes represent biological entities (e.g., proteins) and connections or edges between nodes represent a defined relationship between the biological entities (e.g., a direct physical interaction between proteins). This type of network can then be analyzed in a variety of ways to gain insight into the system of interest (2). For example, identifying clusters of highly connected proteins could indicate that they share function(s); alternatively, proteins that are connected to many other proteins could be a central node in the system.

One of the strengths of Cytoscape is its ability to easily allow a researcher to integrate and visualize different types of data, such as data from gene expression experiments or functional annotations, onto a biological network (3). This is accomplished by first loading the data onto a node or an edge attribute (such as gene expression data or interaction confidence) and then mapping these attributes onto visual attributes (such as node color or edge width). In this manner, diverse data sources and types can be visualized simultaneously in a user-defined manner, in a network context. For example, Cytoscape was used to visualize a network based on the correlation of yeast genetic interaction profiles (4) and the clustering of functionally related yeast genes was visualized by coloring the nodes based on gene function annotations.

Another source of information that can be easily added to a biological network in Cytoscape is compound–target information, for example the name or mode of action of a target’s inhibitor. As we describe in the methods, this type of data can be annotated to nodes in a network and visualized by changing the color of nodes if a compound is known to inhibit the protein in question. By visualizing information about known compound–target interactions in the context of a biological network of interest, an experimentalist can rapidly identify novel avenues to perturb the system with compounds to obtain new mechanistic insight and potentially identify therapeutically relevant targets. For example, such a strategy can be employed using a synthetic lethal interaction network, a type of gene–gene network where nodes represent genes and edges indicate that if both of these genes are mutated the cell is sick or dead. Thus, coloring such nodes highlights tractable drug targets for inhibition. The prediction from such a visualization is that a cancer cell harboring a mutation in one of these genes will be hypersensitive to the chemical inhibition of its synthetic lethal partner. For instance, genes in a synthetic lethal network based on known oncogenes can be colored to indicate if there is a known inhibitor for the protein product of this gene. A strategy based on identifying druggable synthetic lethal interactions in cancer was demonstrated to be effective in a recent phase II clinical trial. In this study, Olaparib™, a poly(ADP-ribose) polymerase (PARP) inhibitor that induces synthetic lethality in homozygous BRCA-deficient cells in culture, was found to be effective in the clinic for treating advanced ovarian cancer in patients with *BRCA1* or *BRCA2* mutations (5).

There are a number of freely available databases that contain compound–target information which can be displayed on a biological network using Cytoscape. They range from databases focused on approved drugs and clinical candidates (6) to databases that attempt to extract all available information regarding named compounds and their targets with text mining algorithms available in the literature (7). One of the largest, freely available databases of this type is ChEMBL which is available from the European Molecular Biology Laboratory (EMBL). ChEMBL is a database of bioactive drug-like small molecules with associated bioactivities

that were manually curated from the literature. It contains more than 500,000 compounds and more than 2.5 million bioactivities. We used ChEMBL to demonstrate how to display chemical data and activity information on a biological network because each compound–target interaction is supported by multiple levels of quality control. First, the source publication for each interaction is provided. Second, each interaction is given a confidence level to indicate how likely the named target is the true target based on the type of assay involved. Finally, the concentration of the compound is also provided.

---

## 2. Materials: Software and Databases

### 2.1. Cytoscape

Cytoscape is a freely available, open-source bioinformatics software platform for visualizing interaction networks and integrating these networks with alternative sources of data. The core distribution of Cytoscape provides a basic set of features for network analysis and visualization. There is an active community of developers who expand the features of Cytoscape by creating freely available plug-ins for a wide variety of uses. For example, there is a plug-in available to predict gene function (8) or for searching a network for a subgraph matching a specific topology (9). There are extensive tutorials on how to use Cytoscape in different contexts, such as integrating networks with gene expression data (3).

To install Cytoscape, navigate to the Cytoscape homepage (10), click on the button “Download Cytoscape Now,” and follow the installation instructions for the appropriate operating system. Cytoscape is available for all of the major platforms as a JAVA program.

### 2.2. Chemviz Plug-In

Chemviz is a Cytoscape plug-in that extends Cytoscape’s capabilities into the realm of cheminformatics. It can display 2D chemical structures using “smiles” or InChI strings that are annotated to one or more nodes in a network. Chemviz can also perform calculations based on these attributes, such as creating a compound similarity network for a selection of nodes. More details are available on the Chemviz home page (11).

To install the plug-in:

1. In the Cytoscape menu bar, click on “Plugins” → “Manage Plugins.”
2. Click on the “Analysis” tab in the “Available for Install” section.
3. Select the plug-in called “ChemViz” and press “Install.” A status bar appears indicating the installation progress. When installed, the plug-in is immediately available from the menu bar under “Plugins” → “Cheminformatic tools.” Default settings, such as which attributes contain the structure details, can be modified in the “Settings” option of the plug-in.

## 2.3. Sample Files

Two sample files, a network file and an attribute file, are required to run the first protocol described in Subheading 3.1. Both are available from our supplementary Web page (12).

### 2.3.1. Network File

“KRAS\_SL\_genes\_from\_publications.txt” is a tab-delimited file that describes a synthetic lethal interaction network between KRAS and 116 genes that was derived from two publications (13, 14). A synthetic lethal interaction is a type of genetic interaction, where cell death results from the combined mutations in two genes but not from mutations in each gene individually. The file contains five columns:

*Screened Entrez Gene:* The KRAS Entrez Gene number.

*SL Entrez Gene:* The Entrez Gene number for the gene that was reported to be synthetic lethal with the KRAS gene.

*Screened Gene:* Gene symbol for KRAS.

*SL Gene:* Gene symbol for the gene that was reported to be synthetically lethal with the KRAS gene.

*PubMed Link:* Link to the publication that reported this interaction.

### 2.3.2. Attribute File

“ChEMBL\_targets\_with\_named\_compounds.txt” is a tab-delimited file describing 2,104 proteins from a variety of species that are reported in the ChEMBL database (v6) to have at least one published, high-confidence ( $\geq 5$  ChEMBL confidence level), potent interaction (reported potency in ChEMBL  $< 1 \mu\text{M}$ ) with a compound. This file is meant to be used as a guide, and the interested reader is encouraged to generate similar files to display different types of chemical data, such as compound–target interactions from different databases (see Note 1).

This sample file contains 14 columns:

*UniProt identifier (ID):* The UniProt ID for the protein, as listed in ChEMBL.

*Entrez Gene ID:* The Entrez Gene ID for this protein based on the ID mapping service provided by the UniProt Web site (see Note 2).

*Protein name from ChEMBL:* The common name for this protein as listed by ChEMBL.

*Number of different compounds reported in ChEMBL:* The number of distinct compounds that have a value of  $< 1 \mu\text{M}$  associated with this protein at a ChEMBL confidence level of 5 or greater. Because commercial availability of particular compounds can vary over time, this number could be used as an estimate of how easy it is to source a compound to inhibit this protein. Additional availability information can be obtained from the ZINC database (15), a free database of commercially available compounds.

*Number of publications reporting a compound–protein interaction:* Based on the same criteria as above, this is the number of distinct

publications that report at least one compound with an interaction with this protein. This number could be used as a proxy for how often different binding studies on this protein are carried out.

*Top Named Compound:* Based on the criteria above, this is the compound in ChEMBL that has the most publications demonstrating an interaction with this protein out of all the compounds in ChEMBL that have a name. Only ~4,000 compounds in ChEMBL (version 6) have a name associated with them. Names in ChEMBL are derived from a variety of sources, including commonly used names, research codes, trade names, and United States Adopted Names (USANs, unique nonproprietary names assigned to pharmaceuticals marketed in the USA).

*SMILES for example compound:* The SMILES string in this column represents the chemical structure for an example compound that has been reported in ChEMBL to interact with this protein. The example compound was chosen because it has the most publications demonstrating an interaction with this protein out of all the compounds in ChEMBL. Note: This compound is not necessarily a compound with a name associated with it in ChEMBL.

*SMILES for named compound:* Same as above, but for the named compound.

*InChIKey for example compound:* InChIKey is an alternative chemical representation of compounds that is optimized for searching for chemical information using text-based search engines, such as Internet searches and database queries.

*InChIKey for named compound:* Same as above, but for the named compound.

*#Publications reporting interaction for example compound:* This column contains the number of publications in ChEMBL in which the example compound has been shown to interact with this protein. It could be used as a measure of how reliable this particular compound interaction is based on the assumption that the more often the interaction is tested and reported the more reliable it is.

*#Publications reporting interaction for named compound:* Same as above, but for the named compound.

*ChEMBL link:* A link to the ChEMBL Web page for this protein. The Web page contains a wide variety of additional information regarding the reported interaction of this protein with different compounds.

*Compound:* A flag to identify which proteins/genes have been annotated in ChEMBL. This is useful for changing the visualization of the network in Cytoscape based on the presence or absence of compound information for a particular node.

### 3. Methods

In this section, we outline two different approaches to annotating biological networks with compound information available from the ChEMBL database. In the first protocol, the visualization of a node representing a gene in a gene–gene interaction network will be modified if an inhibitor is known for this gene product (Fig. 1). This protocol can be easily adapted to other types of networks, such as protein–protein interactions, provided an annotation file exists that can be used to map what is being visualized to an interaction with a compound. It can also be adapted to display compound information from any database that reports compound–target interaction, such as Drugbank (6) or the Therapeutic Target Database (16). All that is required is a Cytoscape attribute file, similar to the one used here for ChEMBL, which enumerates these interactions.

In the second protocol, a protein–protein interaction network is augmented with additional nodes that represent compounds known to inhibit a particular protein in the network. This type of network can be used to explore drug poly-pharmacology, i.e., to identify compounds that bind multiple proteins, as well as for identifying which such protein inhibitors have been published. It uses the PSICQUIC Web service (17), which provides a standard interface to access a variety of molecular interaction databases. In this

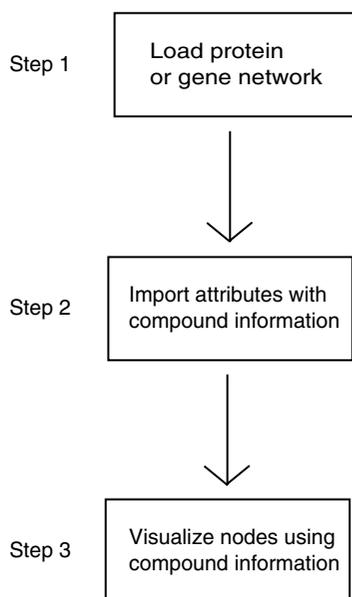


Fig. 1. A flowchart outlining the basic steps involved in annotating a biological network with chemical information in Cytoscape.

protocol, we focus on integrating compound data from ChEMBL with the protein–protein interaction database, MINT (18). The protocol also serves as a general guide on how to annotate any of the interaction networks that are available through the PSICQUIC service with compounds from the ChEMBL database. An updated list of available networks is available from the PSICQUIC Web page (19). The advantage of using this protocol over the first is that the network is always current because it is downloaded directly from the source repository and does not require an attribute file that would have to be updated with each new source repository update. A disadvantage is that the initial network visualization might get too crowded particularly if the protein of interest has an abundance of inhibitors present in ChEMBL.

### **3.1. Annotating a Gene–Gene Interaction Network with Compounds**

In this protocol, a gene–gene interaction network is created and annotated with known compound interaction data. The gene–gene interaction network represents a synthetic lethal interaction network centered on the KRAS gene. Mutations in the KRAS gene, a member of the Ras family of small GTPases, are frequently found in pancreatic, thyroid, colon, lung, and liver cancers and are correlated with poor prognosis (14). In this network, nodes represent genes and edges represent a synthetic lethal interaction based on published interactions (13, 14). By overlaying compound information onto this network, we can rapidly identify proteins that could be of therapeutic relevance for cancers expressing the KRAS oncogene. Compounds that target a gene product that is synthetically lethal with KRAS should, in principle, affect only cancer cells and not normal cells.

#### **3.1.1. Read In and Layout the Network**

1. From the Cytoscape menu bar, click “File” → “Import” → “Network from Table/MS Excel.” This opens a new screen called “Import Network and Edge Attributes from Table.”
2. Click on “Select File(s)” and open the file “KRAS\_SL\_genes\_from\_publications.txt.” A preview of the contents of this file is displayed at the bottom of the “Import Network and Edge Attributes” screen. Note that the default importing options do not work for this file, as it appears that more than 14 attributes are being imported (only 14 attributes are defined in the file).
3. Click on the “Show Text File Import Options.” This provides a number of options to allow Cytoscape to read in the data correctly. Deselect the box on the left-hand side of the “Space” delimiter option. Next, click on “Transfer first line as attribute names.” Note how the preview of the data changes as each option is chosen. It is possible to check that all the text is being imported properly by using the scrollbar in the data preview window. Now, there are 14 attributes being imported which is the same number as in the original file.

4. In the interaction definition area of the import screen, choose “Column 1” from the “Select Source node Column” drop-down menu. Next, choose “Column 2” from the “Select Target node column.” Then, in the data preview window, click on the column entitled “PubMed Link” and click on the “Import” button. Click “Close” once the network has been successfully loaded. The nodes in this network represent genes, which are labeled with Entrez Gene IDs, and edges represent synthetic lethal interactions.
5. From the menu bar, click “Layout” → “Y Files” → “Organic.” This layout displays the nodes and edges in the network using a force-directed algorithm (3). Nodes are modeled with mutually exclusive forces while edges induce an attractive force between nodes that they connect. The algorithm places the nodes in such a manner as to minimize these forces. This layout facilitates the identification of “hub” nodes that are involved in many interactions and the identification of clusters of tightly connected nodes, which could indicate functional modules. In this example, the KRAS gene is connected to all of the other genes and is laid out in the middle of a circle.

### 3.1.2. Import Node Attributes with Compound Information

1. From the menu bar, click “File” → “Import” → “Attributes from Table (Text/MS Excel).” This opens a new screen called “Import Annotation File.”
2. Click on the “Select Files” button and locate the file “ChEMBL\_targets\_with\_named\_compounds.txt.” A preview of the contents of the file appears in the data preview box in a similar manner as when the network was imported at Subheading 3.1.1.
3. Click on the “Show Text File Import Options” check box and then select the “Transfer first line as attribute names” check box. The “UniProt ID” column is highlighted in blue because the default option to map attributes to a network is to use column 1 as the key ID. However, in the sample network, the key ID is Entrez Gene ID. This must be modified to read in the attributes correctly so that they match with the nodes, also identified using Entrez Gene IDs (see Note 2).
4. Click on the “Show Mapping Options” check box. Chose “Entrez Gene ID” from the “Primary Key” drop-down box. Click “Import” to annotate the nodes with these attributes.

### 3.1.3. Change the Visualization of Nodes that Contain Compound Information

1. Click the “VizMapper” tab. Next, click the “Node Color” label in the Visual Mapping Browser to highlight it. Then, double click on the text that reads “Double-Click to create.” This moves the Node Color object from the “Unused Properties” to the “Node Visual Mapping” properties section.

2. Click the “Please select a value” beside the “Node Color” label and select the “compound” attribute created earlier. Next, click “Please select a mapping type!” below the Node Color label and select “Discrete Mapper.” Choose a color to assign to the nodes that represent compounds.
3. Repeat the above two steps for the “Edge Color” attribute, choosing a color based on the “PubMed Link” attribute.
4. Repeat steps 1 and 2 for the “Node Size” attribute, choosing the size of the node based on the attribute “Number of different compounds reported in ChEMBL.” In this example, we suggest binning the values such that a node with 0–10 compounds gets a size of 40, 11–100 compounds, a size of 60, 101–500, a size of 80, and if more than 500 compounds are associated with this node, it gets a size of 100.
5. Next, assign a visual mapping to the “Node Label” attribute based on the attribute “Top named compound.” In this case, choose the “Passthrough Mapping” option. This labels each of the nodes with a compound name. It is possible to change the position of this label and font size used by clicking on the box labeled “Defaults.” This box is located just above the label “Visual Mapping Browser.” It is possible to override a default setting for any of the visual mapping attributes by right clicking on the node and selecting “Visual Mapping Bypass.” This is particularly useful if labels overlap with other objects, such as other labels or edges.

### 3.1.4. Exploring Compounds Annotated to Nodes

In the final network (Fig. 2), there are 24 genes that are colored due to their having an inhibitor annotated in ChEMBL. Each of these genes has nine compound-related attributes annotated to them which are described in Subheading 2.3.2. To display them in the “Data Panel,” click on the “Select Attributes” button, which is the left-most icon in the “Data Panel,” and select each of them. Selecting a colored node on the network displays the attributes for that gene in the “Data Panel.”

One intriguing observation about this network is that four nodes are annotated to the compound “Alvocidib.” The largest node with this label is the node corresponding to Gene ID “983.” It is one of the three largest nodes in the network, reflecting that compound–protein interactions are often reported for this protein. To select this node, just click on it. This highlights the node and displays its attributes in the data panel. This gene encodes the protein “Cyclin-dependent kinase 1.” There are 77 publications in which 558 compounds are associated with this protein in ChEMBL. To go to the ChEMBL page for this protein, click on the “Chembl Link” attribute in the data panel. To view the chemical structure of the example compound, right click on this node and select the option “CheminformaticsTools” → “Depict2DStructure” → “Show

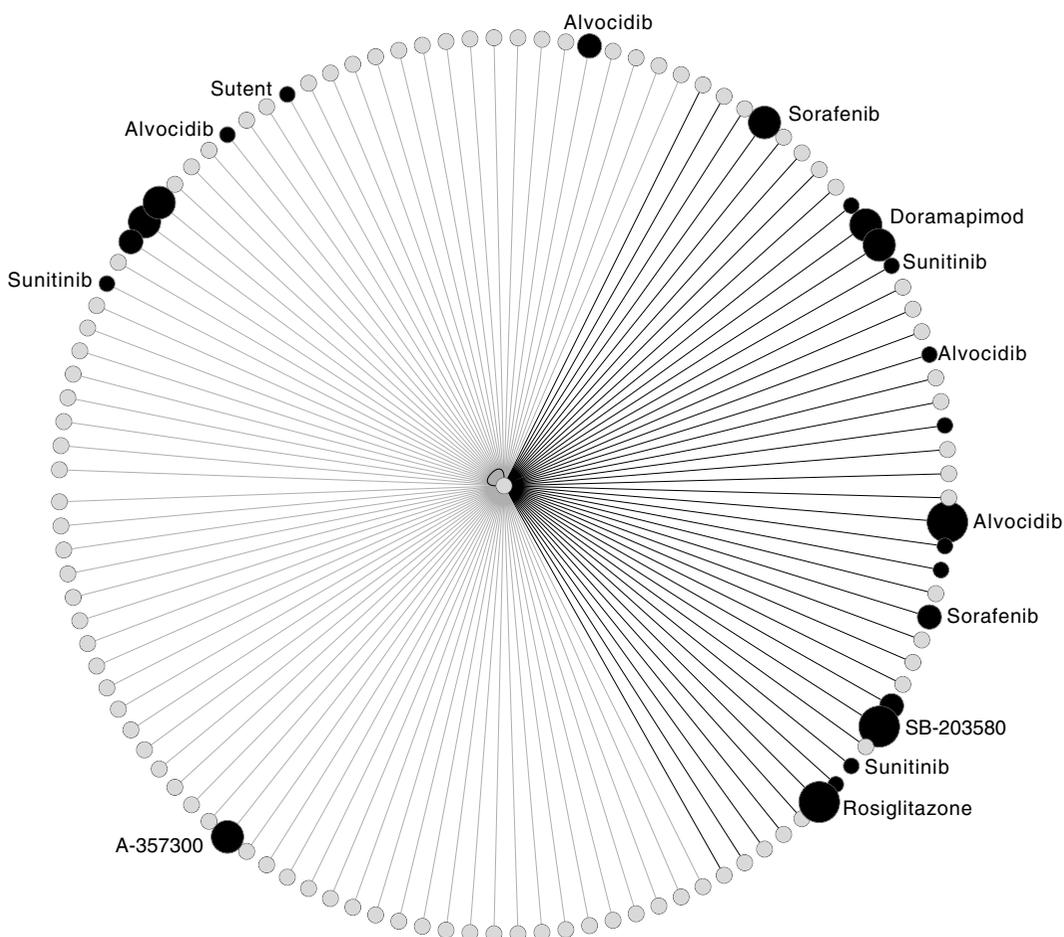


Fig.2. A synthetic lethal network centered on the KRAS oncogene that is produced by following the steps outlined in Subheading 3.1. Nodes represent genes, and edges represent a synthetic lethal interaction between the two connected nodes. Nodes are colored *black* if an inhibitor is known for the gene product of this node. Node size represents the number of compounds that have been associated with this gene in ChEMBL. Nodes are labeled with a compound name if a suitable name exists in ChEMBL. Edges are colored depending on which one of two publications the synthetic lethal interaction is derived from.

2D structures from this Node.” By right clicking on the InChIKey for this compound, one can search for it on the Web using the Google search engine. The top hit is from Drugbank and shows that an alternative name for this compound is flavopiridol and that it is a known cyclin-dependent kinase inhibitor. The other three proteins associated with this compound are cyclin A2, dual-specificity tyrosine-phosphorylation-regulated kinase 1B, and mitogen-activated protein kinase 7. Perhaps, the fact that it induces numerous synthetic lethal interactions with KRAS contributes to the effectiveness of alvocidib for treating certain cancer types. It is in a phase I trial (Clinical Trials ID: NCT00101231) for treating patients with relapsed or refractory acute myeloid leukemia (AML), acute lymphoblastic leukemia, (ALL), or chronic myelogenous leukemia.

AML and ALL are cancers associated with mutations in the KRAS oncogenes (20).

Clicking on “File → Save as” from the Cytoscape menu bar can save the network session. The network visualization can be exported to a variety of file formats, including PDF, by clicking on “File” → “Export” → “Network view as graphics.”

### **3.2. Annotating a Protein–Protein Interaction Network with Compounds**

This protocol allows one to annotate a protein–protein interaction network derived from the MINT database with compounds from the ChEMBL database. In this network, nodes represent both compounds and proteins. Edges between two protein nodes represent an experimentally verified protein–protein interaction that has been curated from the literature. Edges between a protein and compound nodes represent a high-confidence protein–compound interaction from ChEMBL as defined by the ChEMBL curators (see Note 1). A key step in this protocol is coloring the compound nodes differently from protein nodes. The advantage of this protocol over the one described in Subheading 3.1 is that the most up-to-date networks are downloaded dynamically from the relevant groups. It does not rely on a static attribute file, which has to be updated with every new release of ChEMBL. The disadvantage is that the final network might be cluttered, as some proteins have many known inhibitors annotated in ChEMBL.

#### *3.2.1. Load Networks*

1. From the Cytoscape menu bar, click on “File” → “Import” → “Import network from webservices.” A new window appears entitled “Import network from database.” This window allows Cytoscape to connect to a variety of resources on the Internet to download networks. Different data sources can be chosen from the drop-down list located at “Data Source,” and information describing the selected data source can be obtained by clicking the “About” button.
2. Select the “PSICQUIC Universal Web Service Client” from the “Data Source” drop-down list.
3. In the search box, type “P19438 or Q13546” and press the Search button. P19438 and Q13546 are UniProt protein identifiers for the input proteins: tumor necrosis factor receptor 1 (P19438) and receptor-interacting serine/threonine-protein kinase 1 or (Q13546), both involved in the apoptosis pathway. Click “yes” when prompted to generate a new network from all the interactions that were found. Click “OK” when presented with the PSICQUIC data import result message box. This searches all of the molecular interaction databases available to PSICQUIC and identifies any of the interactions that involve any of these three proteins. The data import box displays all of the databases from which interactions were obtained. Clicking “OK” creates a separate network

for each of the different databases. Of particular note for this protocol are the ChEMBL and the MINT networks. The ChEMBL network contains compounds that have been shown to interact with the proteins that were input into the PSICQUIC search box and the input proteins themselves if they are present in ChEMBL. The MINT network contains protein–protein interactions involving the input proteins.

### 3.2.2. Color the Compound Nodes by Compound Property

1. Select the ChEMBL network from the network panel on the left.
2. From the menu bar, click “Layout” → “Y Files” → “Organic.” In this example, the hub nodes are the input proteins and the rest of the nodes are the compounds retrieved from ChEMBL.
3. Open the “Create New Integer Attribute” by clicking on the middle icon in the top left-hand corner of the data panel and selecting “Integer Attribute.”
4. Enter “compound” in the “Create New Integer Attribute” box, and select OK. This creates a new attribute named “compound” which is used to identify and color nodes that are compounds in a similar manner for protocol 1.
5. From the menu bar, click on “Select” → “Nodes” → “Select All Nodes.”
6. Click on each of the hub genes while holding shift key. This step deselects each of the original input proteins and leaves all of the compounds selected. An alternative approach to deselect the original input proteins would be sorting the ID column in the data panel. The compound IDs start with either “chebi” or “ChEMBL” while the protein IDs start with “uni-protkb.” Clicking on the protein IDs (holding shift key while clicking allows multiple IDs to be selected) in the data panel highlights the corresponding nodes in green on the network. Clicking on each of these nodes in turn deselects them.
7. Open the Node Attribute Batch Editor by clicking on the left-most icon on the top right-hand corner of the data panel. Select the “compound” attribute from the blank drop-down box, enter “1” into the text box, and press “GO.” Close the node attribute batch editor by pressing the “x.” This sets the color attribute to “1” for all the compounds.
8. Use the VizMapper protocol as outlined in protocol 1 to color the nodes based on the compound property. Coloring nodes in this manner means that compound nodes can be easily distinguished from protein nodes when the ChEBML network is merged with the protein–protein network.

### 3.2.3. Merge Networks

1. From the menu bar, click on “Plugins” → “Advanced Network Merge.” This launches the “Advanced Network Merge”

plug-in that allows the user to combine two or more networks into one network. In this protocol, we merge the ChEMBL and MINT networks. Any of the networks downloaded can also be merged and annotated with compounds in a similar fashion.

2. In the “Networks to merge box,” select the ChEMBL network and then click on the arrow pointing to the selected “networks box.” Repeat this for the MINT network, and then click on the Merge button. This creates a new network called “Union” which contains all of the nodes and edges in both of the original networks. It also retains the node coloring that was generated based on the compound annotation. Compound nodes connecting multiple protein nodes indicate poly-pharmacology for compounds in the network. The user can rerun this protocol to expand the initial list of input proteins to annotate interesting regions of the protein–protein network with compound–protein interactions.

---

## 4. Notes

1. Care must be taken when creating an attribute file to use to map compound–target information onto a biological network from any compound–target database. While these databases contain vast amounts of useful data, they also contain unexpected data and potential experimental artifacts. For instance, a protein–compound network based on PubChem Web site, without careful filtering, contains many false positive interactions due to the presence of promiscuous compounds and other factors (21). In another database, Drugbank (6), there are two references that are annotated as providing evidence for over 7,000 compound–target interactions each. This is unlikely to be correct and will seriously bias downstream analysis if not filtered out.
2. It is important to take care when integrating multiple sources of data due to ambiguous naming conventions. In the case of mapping chemical information onto gene–gene networks, we have found that it is best to map the data using Entrez Gene IDs as they are unique and unambiguous. In contrast, gene names are not a useful for mapping information onto a network because multiple gene names can match to many genes. In the case of protein–protein networks, we have found it best to map compound interactions using UniProt IDs as they are the most widely used identifiers in the compound databases. To convert between various protein and gene identifiers, we have found the mapping service provided at the UniProt Web site (22), as well as the service provided by the EBI (23) to be particularly useful.

## Acknowledgments

The second protocol discussed here is inspired by and adapted from a presentation by Anna Gaulton at the “Small molecule bioactivity resources at the EBI” course in January 2010. This work was supported by grants from the NHGRI to CN and GG, the CIHR to CN (MOPS-84305) and to GG (MOPS-81340), by NSERC support to GB and IMW, and from a Marie Curie fellowship to IMW. GG is a Canada Research Chair in Chemical Biology.

## References

- Shannon P, et al., *Cytoscape: a software environment for integrated models of biomolecular interaction networks*. *Genome Res*, 2003. **13**(11): p. 2498–504.
- Merico D, D Gfeller, and GD Bader, *How to visually interpret biological data using networks*. *Nat Biotechnol*, 2009. **27**(10): p. 921–4.
- Cline MS, et al., *Integration of biological networks and gene expression data using Cytoscape*. *Nature protocols*, 2007. **2**(10): p. 2366–82.
- Costanzo M, et al., *The genetic landscape of a cell*. *Science*, 2010. **327**(5964): p. 425–31.
- Audeh MW, et al., *Oral poly(ADP-ribose) polymerase inhibitor olaparib in patients with BRCA1 or BRCA2 mutations and recurrent ovarian cancer: a proof-of-concept trial*. *Lancet*, 2010. **376**(9737): p. 245–51.
- Wishart DS, et al., *DrugBank: a comprehensive resource for in silico drug discovery and exploration*. *Nucleic Acids Research*, 2006. **34**(Database issue): p. D668–72.
- Kuhn M, et al., *STITCH: interaction networks of chemicals and proteins*. *Nucleic Acids Res*, 2008. **36**(Database issue): p. D684–8.
- Montojo J, et al., *GeneMANIA Cytoscape Plugin: Fast gene function predictions on the desktop*. *Bioinformatics*, 2010.
- Ferro A, et al., *NetMatch: a Cytoscape plugin for searching biological networks*. *Bioinformatics*, 2007. **23**(7): p. 910–2.
- Cytoscape Home Page*. <http://www.Cytoscape.org>.
- ChemViz Plugin Home page*. <http://www.cgl.ucsf.edu/Cytoscape/chemViz/>.
- Supplementary Web Page*. <http://baderlab.org/VisualizingChemicalInformation>.
- Barbie DA, et al., *Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1*. *Nature*, 2009. **462**(7269): p. 108–12.
- Luo J, et al., *A genome-wide RNAi screen identifies multiple synthetic lethal interactions with the Ras oncogene*. *Cell*, 2009. **137**(5): p. 835–48.
- Irwin JJ and BK Shoichet, *ZINC—a free database of commercially available compounds for virtual screening*. *Journal of chemical information and modeling*, 2005. **45**(1): p. 177–82.
- Zhu F, et al., *Update of TTD: Therapeutic Target Database*. *Nucleic Acids Research*, 2010. **38**(Database issue): p. D787–91.
- Orchard S, et al., *implementing data standards: a report on the HUPOPSI workshop September 2009, Toronto, Canada*. *Proteomics*, 2010. **10**(10): p. 1895–8.
- Ceol A, et al., *MINT, the molecular interaction database: 2009 update*. *Nucleic Acids Res*, 2010. **38**(Database issue): p. D532–9.
- PSICQUIC databases*. <http://www.ebi.ac.uk/Tools/webservices/psicquic/registry/registry?action=STATUS>.
- Liang D-C, et al., *K-Ras mutations and N-Ras mutations in childhood acute leukemias with or without mixed-lineage leukemia gene rearrangements*. *Cancer*, 2006. **106**(4): p. 950–6.
- Han L, Y Wang, and SH Bryant, *A survey of across-target bioactivity results of small molecules in PubChem*. *Bioinformatics*, 2009. **25**(17): p. 2251–5.
- UniProt*. [www.uniprot.org](http://www.uniprot.org).
- Côté RG, et al., *The Protein Identifier Cross-Referencing (PICR) service: reconciling protein identifiers across multiple source databases*. *BMC Bioinformatics*, 2007. **8**: p. 401.