

TROVE (Tool for hallmark annotation, visualization and characterization)

USER GUIDE (Version 1.0) Linux

Chua Huey Eng
3/13/2017

Contents

1.	TROVE Installation.....	2
1.1	PostgreSQL 9.5	2
1.2	network_lib_TROVE folder	5
1.3	TROVE.jar	6
2.	Launching TROVE	7
3.	TROVE Features.....	8
3.1	Loading the human signalling network base file	8
3.2	Loading and Visualizing Specific Cancer Signaling Network (Existing)	9
3.3	Adding New Cancer Subtypes	11
3.4	Editing of Graph.....	17
3.5	Editing of Node Annotation.....	19
3.6	Editing of Hallmark Annotation.....	20
3.7	Editing of Hallmark-GO Mapping Used for Hallmark Annotation	20
3.8	Visualization of Expression Fold Change and Mutation Frequency.....	21
3.9	Visualization of Node Neighbourhood.....	23
3.10	Visualization of Pathway(s) Between a Pair of Nodes	24
3.11	Visualization of Hallmarks.....	25
3.12	Network Topology-based Characterization of Hallmarks.....	26
3.13	Trouble-shooting.....	28

1. TROVE Installation

TROVE uses several other publicly available tools and software for pre-processing files and storing data. In particular, the PostgreSQL database management system is used for storing data. PostgreSQL has to be installed in order for TROVE to function. Please adhere to the recommended version as it has been tested. Note that TROVE has been tested for installation on Ubuntu version 16.04 LTS.

1.1 PostgreSQL 9.5

Installing PostgreSQL 9.5 (instructions are available at the URL: <http://postgresguide.com/setup/install.html>)

Step 1: Install PostgreSQL by following the commands relevant to your linux system.

APT systems (Ubuntu, Debian, Mint, etc):

- sudo apt-get install postgresql

Arch linux:

- sudo pacman -S postgresql

YUM installs (Fedora, Red Hat, CentOS, Scientific linux) [assuming PostgreSQL 9.2 on CentOS 6.4 x64]:

- Go to PostgreSQL Yum Repository to select the version of PostgreSQL that you want to install and then your OS, version and architecture.
- Download the RPM for your platform
 - curl -O http://yum.postgresql.org/9.2/redhat/rhel-6-x86_64/pgdg-centos92-9.2-6.noarch.rpm
- Install the RPM
 - rpm -ivh pgdg-centos92-9.2-6.noarch.rpm
- Do a quick search to list the available packages for postgres
 - yum list postgres*
- Install packages as needed
 - yum install postgresql92-server-x86_64 postgresql92-contrib.x86_64 postgresql92-devel.x86_64

Installing pgAdmin 3 (GUI for PostgreSQL)

Step 2: Install pgAdmin 3

APT systems (Ubuntu, Debian, Mint, etc):

- sudo apt-get install pgadmin3

Arch linux:

- sudo pacman -S pgadmin3

YUM install:

- yum install pgadmin3_91

Verifying the installation

Step 3: Start pgAdmin 3.

- pgadmin3

Step 4: Create the PostgreSQL 9.5 server (Fig. 1)

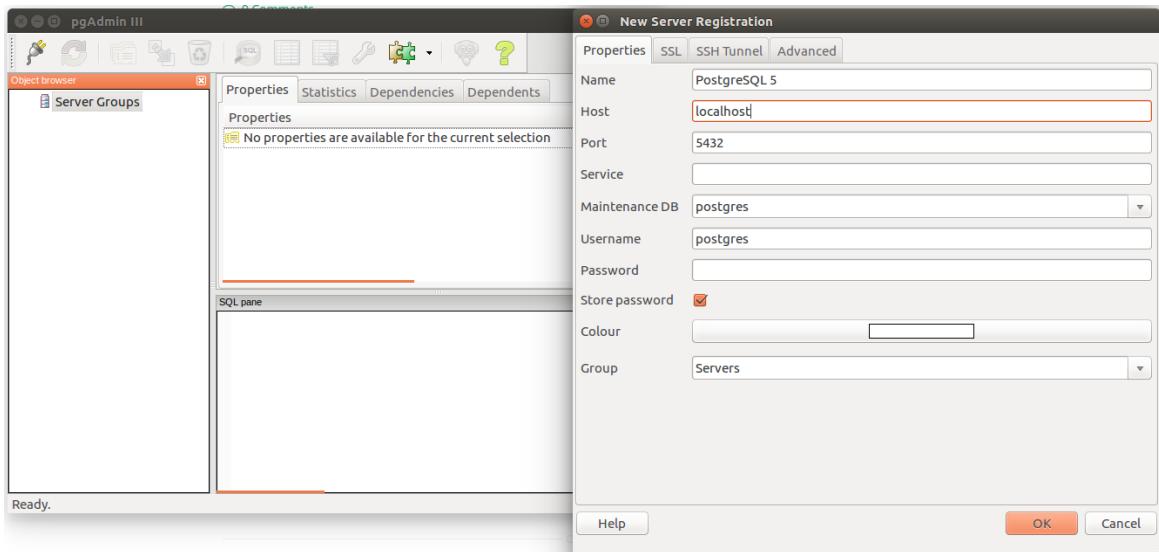


Fig. 1: Creating PostgreSQL server

Setting or resetting the password to ‘trove’

Step 5: Start a terminal.

- sudo –u postgres psql
- ALTER USER postgres WITH ENCRYPTED PASSWORD ‘trove’;

Creating the TROVE database

The TROVE database is used by TROVE for storing data and has to be created before running TROVE. The database creation has to be done only once during the installation phase and will be available subsequently when TROVE is run.

Step 6: Right-click on “Databases” on the object browser and select “New Database...” (Fig. 2)

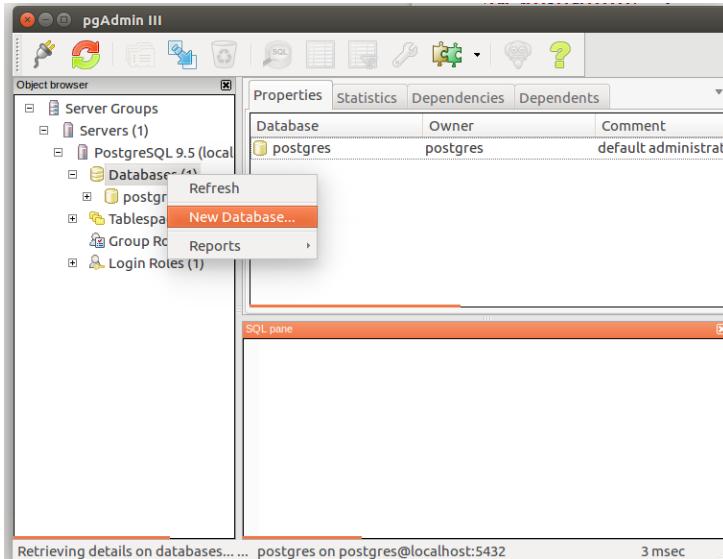


Fig. 2: Creating new database

Step 7: Type in “TROVE” as the name of the database (Fig. 3) and select “OK”.

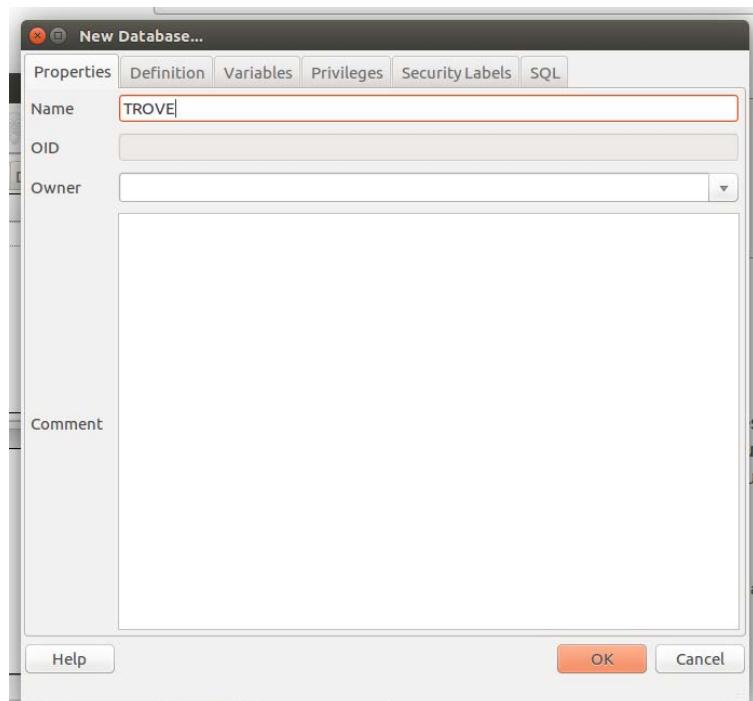


Fig. 3: Setting the name of the new database

Step 8: Verify that the “TROVE” database has been created

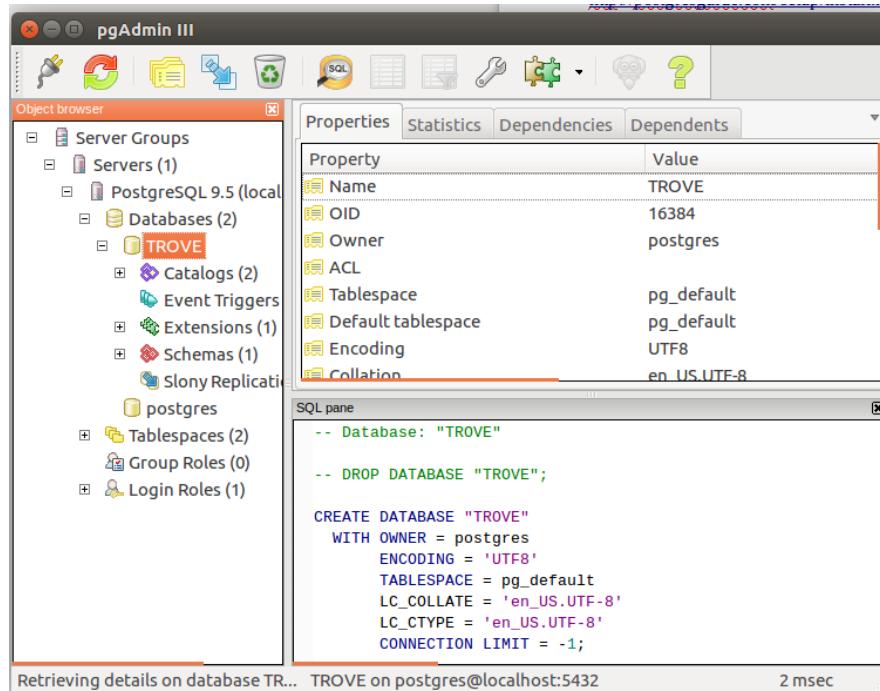
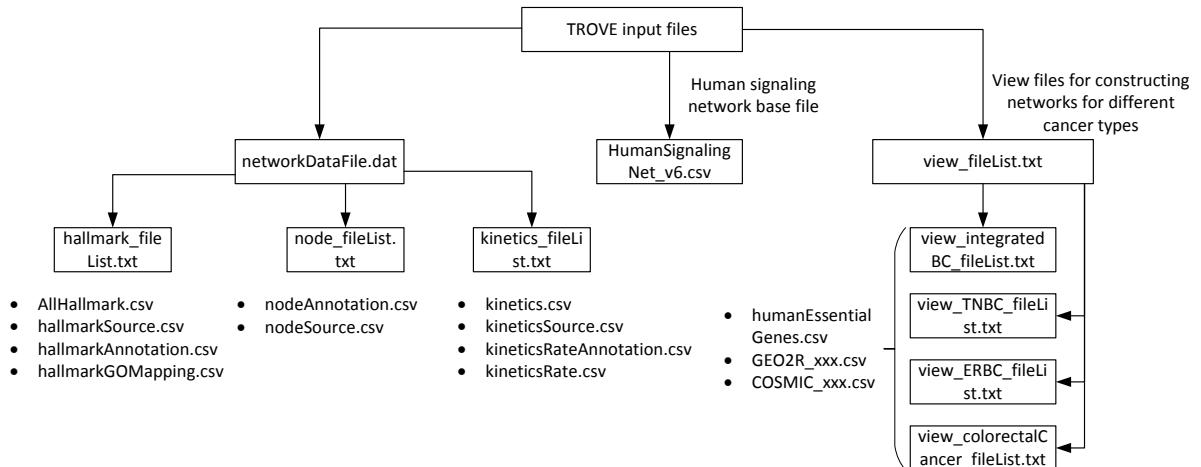


Fig. 4: Verify that the new database has been created

1.2 network_lib_TROVE folder

This folder contains the example files for running TROVE. TROVE uses SVM to perform hallmark characterization and intermediate files generated while running the SVM are stored in the “svmTraining” folder within the “network_lib_TROVE” folder.

TROVE requires several files from this folder for proper execution. The structure of the files is as follows:



1. HumanSignalingNet_v6.csv – the base signalling network file that the cancer type-specific view is built on
2. view_fileList.txt – contains various views for cancer type. Users can expand this list to include new types. Each new view should contain three files, namely, humanEssentialGenes.csv, GEO2R_<specificView>.csv and COSMIC_<specificView>.csv. The GEO2R_<specificView>.csv is extracted from the GEO Omnibus dataset using the GEO2R tool and shall contain three columns of data (Gene, Entrez ID and Fold change). The COSMIC_<specificView>.csv is extracted from the COSMIC database and shall contain 6 columns (index in numeric order, gene name, entrez ID, number of mutated samples, number of samples tested, percentage). Note that percentage=number of mutated samples/number of samples tested*100
3. networkDataFile.dat – this file contains mainly annotations related files such as hallmark annotations (hallmark_fileList.txt), node annotations (node_fileList.txt) and kinetics annotations (kinetics_fileList.txt). For hallmark annotations, the GO-assisted hallmark annotations uses the hallmarkGOMapping.csv as reference for assigning hallmarks to nodes based on selected GO terms. The users can modify the .csv to update the GO-hallmark mapping as required.

Step 10: Copy the “network_lib_TROVE” folder to the desktop (Fig. 5)

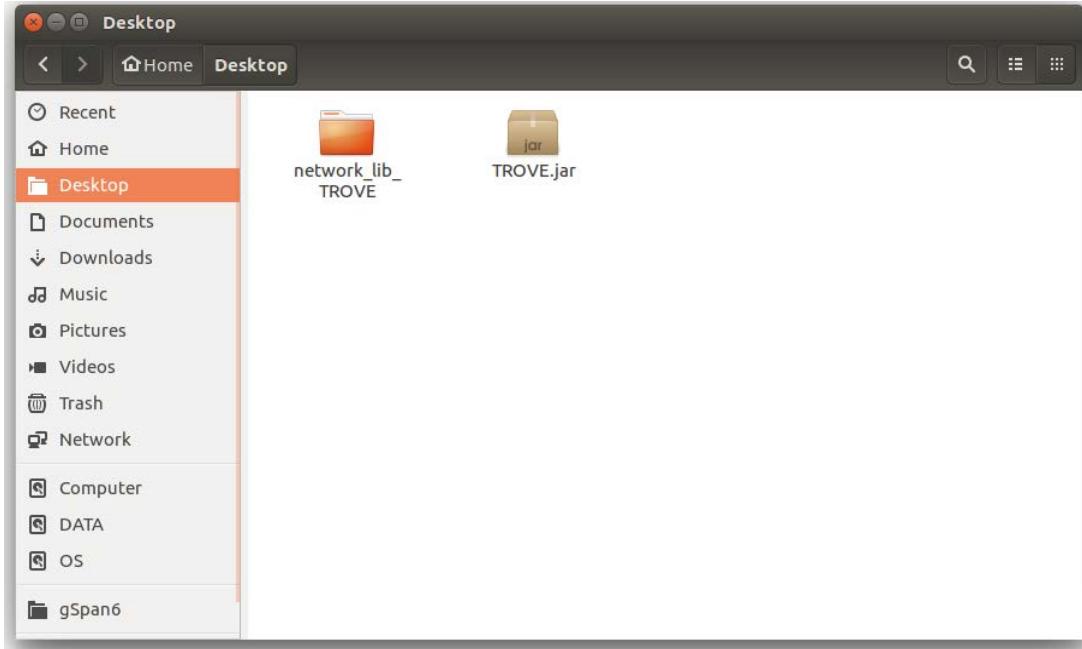


Fig. 5: Copy “network_lib_TROVE” folder to Desktop

1.3 TROVE.jar

Step 11: Check that JDK has been pre-installed using the command “javac –version” in the terminal window. TROVE runs on Java 1.8

Note: If a JDK version number is returned, then JDK has been installed. Check that the version is 1.8 or later. If the JDK version is prior to 1.8 or JDK has not been installed, proceed to install JDK by following Step 11a.

Step 11a (assuming it's APT system):

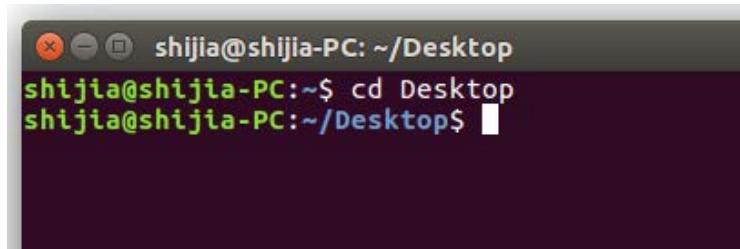
- sudo apt-get install openjdk-8-jdk
- apt-cache search jdk
- export JAVA_HOME=/usr/lib/jvm/java-8-openjdk
- export PATH=\$PATH:/usr/lib/jvm/java-8-openjdk/bin
- Validate installation: javac –version

Note: /usr/lib/jvm/java-8-openjdk is used for demonstration here. You should use your path as per your installation.

Step 12: Copy TROVE.jar to a desired location (e.g., “Desktop”).

2. Launching TROVE

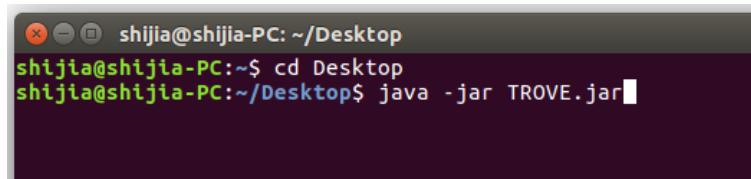
Step 1: Change the directory to the one that contains TROVE.jar (Fig. 6).



```
shijia@shijia-PC: ~/Desktop
shijia@shijia-PC:~$ cd Desktop
shijia@shijia-PC:~/Desktop$
```

Fig. 6: Set to the directory containing TROVE.jar

Step 3: Run TROVE by using the command “java -jar TROVE.jar” (Fig. 7).



```
shijia@shijia-PC: ~/Desktop
shijia@shijia-PC:~$ cd Desktop
shijia@shijia-PC:~/Desktop$ java -jar TROVE.jar
```

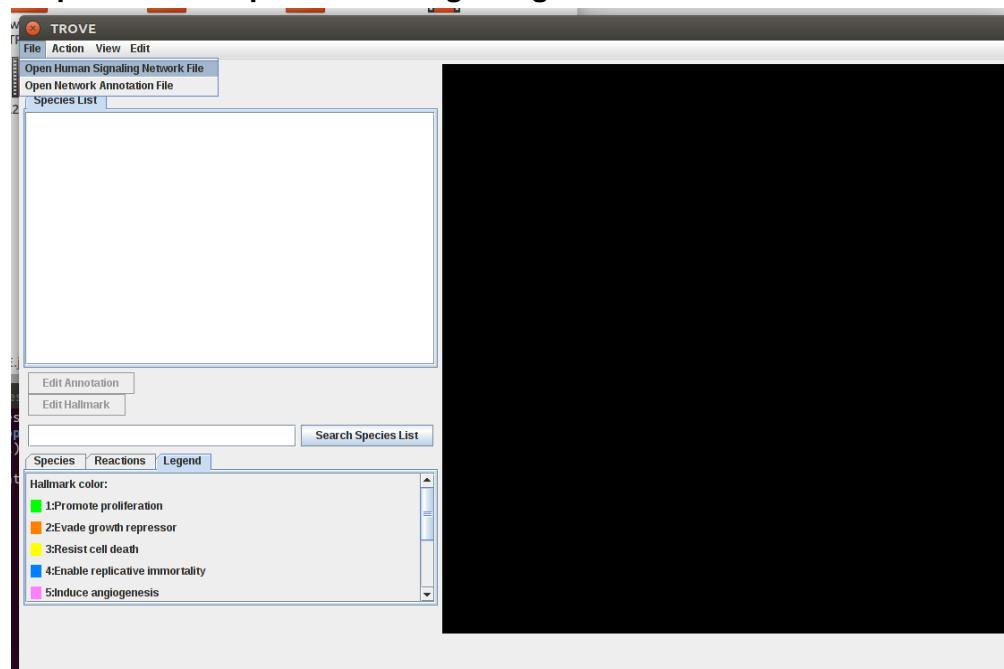
Fig. 7: Run TROVE

3. TROVE Features

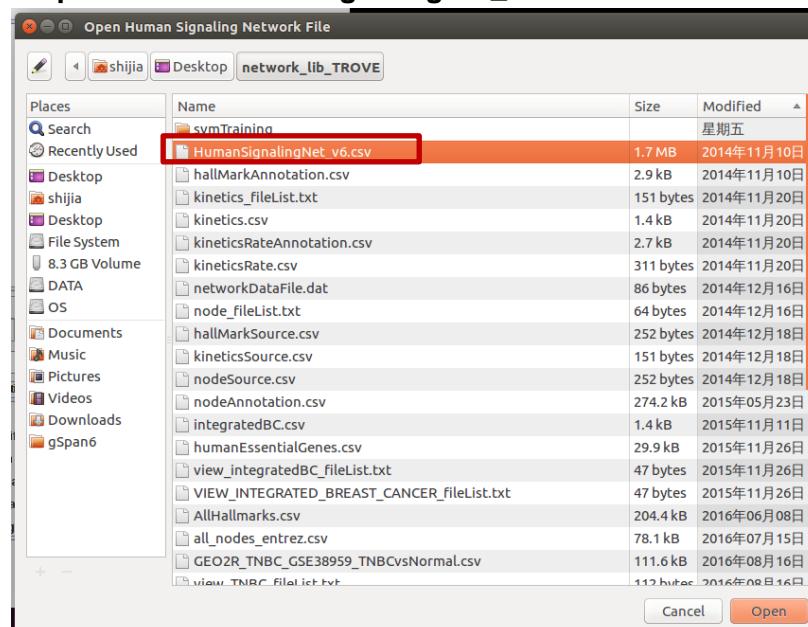
3.1 Loading the human signalling network base file

The human signalling network is the base signalling network that is used to generate the cancer (sub)type-specific signalling network and must be uploaded first. The human signalling network is obtained from <http://www.cancer-systemsbiology.org/data-software> and currently uses version 6. The human signalling network contains binary interaction of human genes. The annotations for these genes are contained in other files and is loaded by using the “Open Network Annotation File” option.

Step 1: Select “Open Human Signaling Network File” from “File” menu item



Step 2: Select HumanSignalingNet_v6.csv



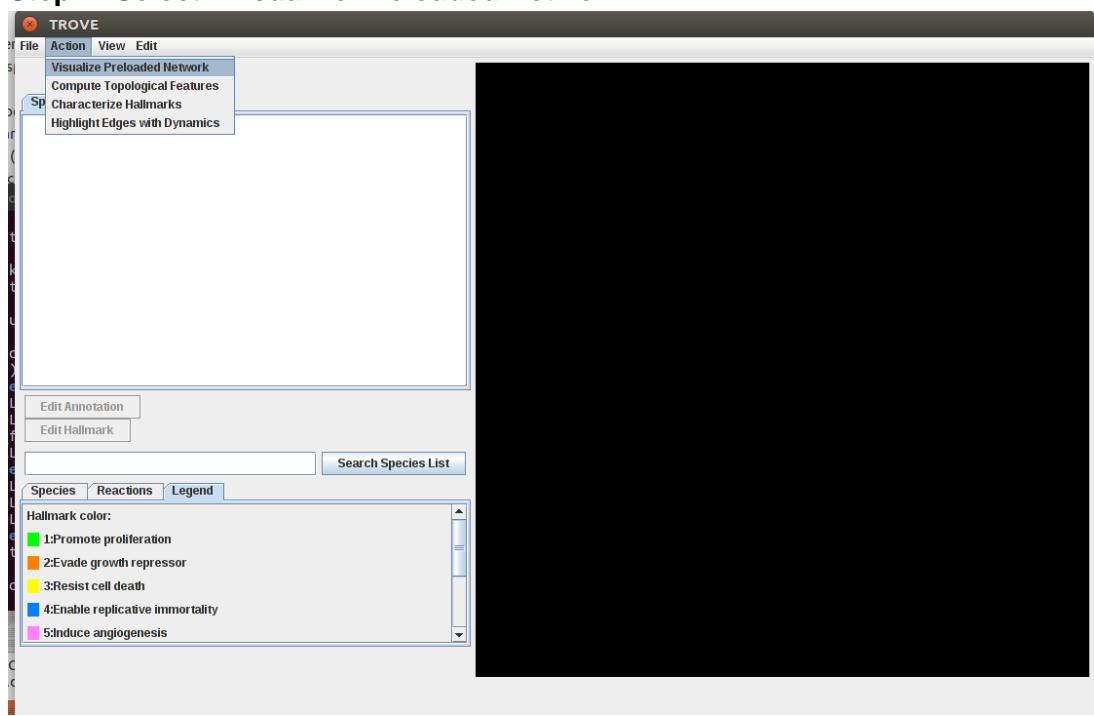
Step 3: Select “Open Network Annotation File” from “File” menu item

Step 4: Select networkDataFile.dat

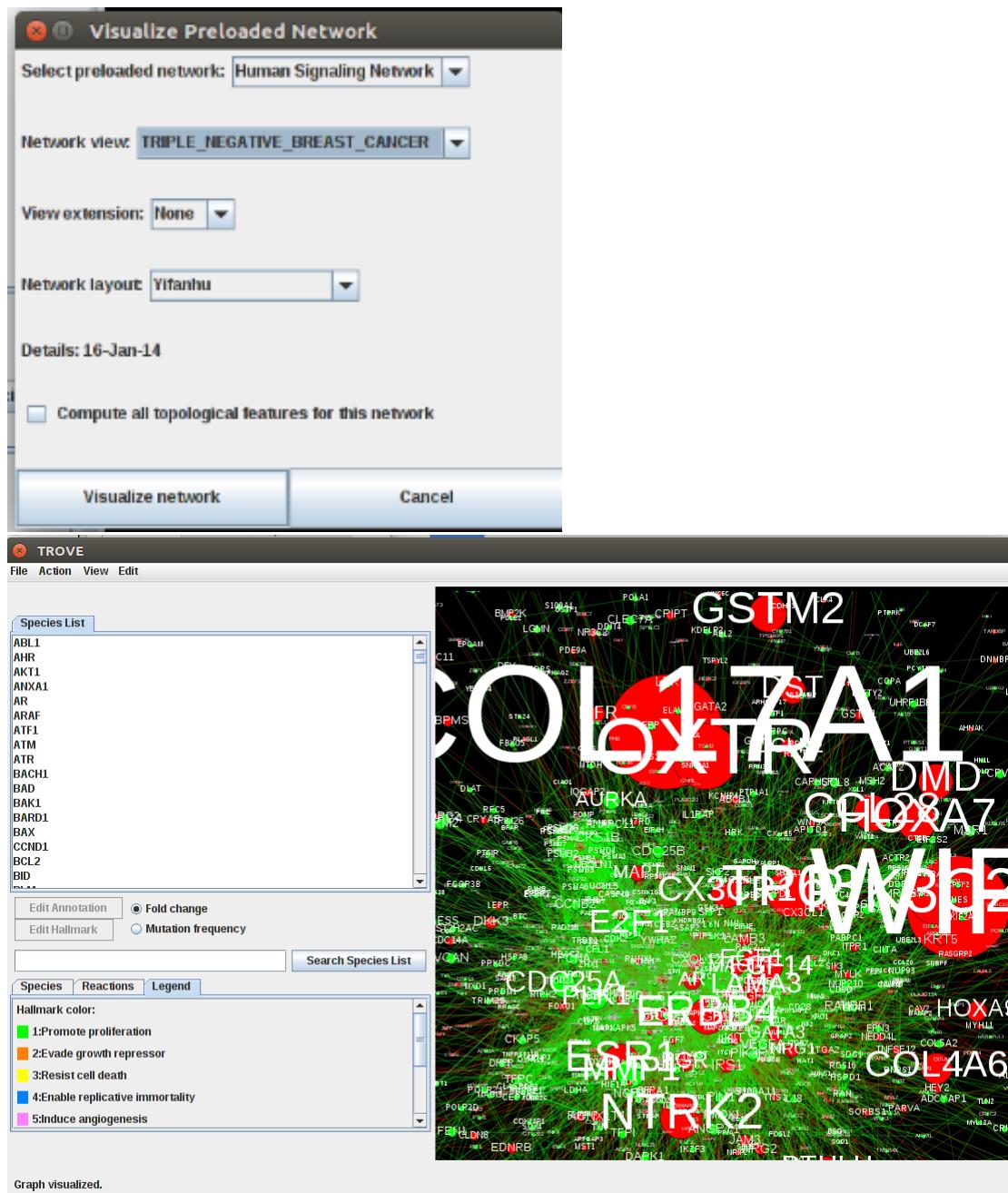
3.2 Loading and Visualizing Specific Cancer Signaling Network (Existing)

TROVE supports the analysis of specific cancer signalling networks and uses dataset from GEO Omnibus (<https://www.ncbi.nlm.nih.gov/geo/>) and COSMIC database (<http://cancer.sanger.ac.uk/cosmic>) to generate these networks.

Step 1: Select “Visualize Preloaded Network”



Step 2: Select the network view, view extension and network layout. Then, click “Visualize network”. The network view currently contains triple negative breast cancer, ER positive breast cancer and colorectal cancer. The view extension is in terms of hops and extend the network further to avoid missing out relevant nodes. The network layout allows configuration of how the visualized network should be displayed.



Using user-generated gene expression data files instead of GEO data files

As mentioned in Section 1.2, the GEO2R_<specificView>.csv is extracted from the GEO Omnibus dataset using the GEO2R tool and shall contain three columns of data (Gene, Entrez ID and Fold change) where fold change is the log2-fold change of expression levels of the gene under two experimental conditions (normal physiological condition and cancerous condition). Hence, TROVE requires the GEO data files to be in a comma-delimited files containing three columns. Users can choose to replace the GEO expression data with their own experimentally-derived data by creating a comma delimited file (.csv) containing the three columns: gene, entrez ID and fold change.

3.3 Adding New Cancer Subtypes

The new cancer subtypes can be inserted using TROVE GUI. First, the user have to generate the COSMIC and GEO2R dataset relevant to the new cancer subtypes and save them in the network_lib_TROVE directory. The steps for generating these files are given below.

COSMIC dataset

The COSMIC dataset reports the frequency of gene mutation for particular cancer types. To generate a new COSMIC dataset, search for the specific cancer type you are interested in (e.g., breast cancer) and follow the steps below.

Step 1: Select cancer type

The screenshot shows the COSMIC homepage at cancer.sanger.ac.uk/cosmic. The search bar contains 'breast cancer' and the 'SEARCH' button is highlighted with a red box. To the right is a circular genome map showing mutation data across chromosomes 1 through 22 and X. Below the search bar are sections for 'Resources' and 'Tools'.

Step 2: Select COSMIC classification

The screenshot shows the COSMIC search results page for the query 'breast cancer'. The search bar at the top also has 'breast+cancer' entered. The results table is shown with two entries:

COSMIC classification	Paper description	Tested samples	Mutations
breast,carcinoma	breast,cancer	44057	190514
breast,NS,carcinoma,NS	breast,NS,cancer,NS	28635	122622

Showing 1 to 2 of 2 entries.

Step 3: Select tissue, sub-tissue, histology and sub-histology

The screenshot shows the COSMIC Cancer Browser interface. The top navigation bar includes links for Home, Resources, Curation, Tools, Data, News, Help, About, and a search bar. A red box highlights the 'Tissue selection' section, which lists various tissues like Adrenal gland, Autonomic ganglia, Biliary tract, Bone, Breast, Central nervous system, Cervix, Endometrium, Eye, and Fallopian tube. Below this is a link to 'Breast (9919 / 47259)'. Another red box highlights the 'Subtissue selection' section, which includes 'Include all' and categories like Extramammary, Nipple, and NS. A third red box highlights the 'Histology selection' section, which includes 'Include all' and categories like Carcinoma, Carcinoma in situ, Hyperplasia, NS, and Other. A fourth red box highlights the 'Subhistology selection' section, which includes 'Include all' and detailed categories for different carcinoma types. Below these sections is a 'Filters' section with a 'Screen Type' dropdown and a 'Go' button.

Step 4: Click on “Genes with Mutations” tab in the Cosmic>> Cancer Browser >> Breast panel and select the “CSV” export option. The exported file is the COSMIC dataset that should be saved to the network_lib_TROVE folder

The screenshot shows the COSMIC Cancer Browser interface with the 'Genes with Mutations' tab selected. A red box highlights this tab. At the top, there are four filter boxes: 'Breast (9919 / 47259)', 'Include all', 'Carcinoma (8733)', and 'Basal (triple-negative) carcinoma (256)'. Below these is a 'Filters' section with a 'Screen Type' dropdown and a 'Go' button. The main content area displays a table of genes with mutations, showing columns for Gene, Mutated samples, and Samples tested. At the bottom right of the table, there is an 'Export' button with options for 'CSV' (highlighted with a red box) and 'TSV'. The table shows 10 entries out of 25,876 total.

GEO Omnibus dataset

The GEO Omnibus dataset contains the gene expression data related to experiments. To generate a new GEO2R dataset, search for the specific GEO dataset for the cancer type you are interested in (e.g., GSE38959) and follow the steps below.

Step 1: Select GEO dataset (e.g., GSE38959)

The screenshot shows the main interface of the GEO Expression Omnibus website. At the top, there's a navigation bar with links for 'Secure', 'How To', 'GEO Home', 'Documentation', 'Query & Browse', and 'Email GEO'. On the right side of the header, there's a 'Sign in to NCBI' link. Below the header, the 'Gene Expression Omnibus' logo is displayed. A search bar at the top right contains the accession number 'GSE38959', which is highlighted with a red box. To the right of the search bar is a magnifying glass icon and a 'Search' button. The main content area is divided into three columns: 'Getting Started' (with links to Overview, FAQ, About GEO DataSets, and About GEO Profiles), 'Tools' (with links to Search for Studies at GEO DataSets, Search for Gene Expression at GEO Profiles, Search GEO Documentation, and Analyze a Study with GEO2R), and 'Browse Content' (with links to Repository Browser, DataSets: 4348, Series: 83469, and Platforms: 17105). A brief description of GEO's mission is also present in the center column.

Step 2: Select “Analyze with GEO2R”

The screenshot shows the 'Accession Display' page for the dataset GSE38959. At the top, the NCBI logo and a 'GEO Expression Omnibus' logo are visible. The URL in the address bar is 'NCBI > GEO > Accession Display [?]' and the page title is 'Not logged in | Login [?]'.

On the left, there's a sidebar with a 'BioProject' section showing 'PRJNA169423'. The main content area starts with a search bar containing 'Scope: Self', 'Format: HTML', 'Amount: Quick', 'GEO accession: GSE38959', and a 'GO' button. Below the search bar, the dataset details are listed: Status (Public on Dec 21, 2012), Title (Gene expression profiling of triple negative breast cancer, normal ductal cells, and normal tissues), and Organism (Homo sapiens). There are also links for 'Download family', 'Format', 'SOFT', 'MINIML', and 'TXT'.

A large red box highlights the 'Analyze with GEO2R' button. Below it, there's a table for 'Supplementary file' with one entry: 'GSE38959_RAW.tar' (Size: 96.9 Mb, Format: (http)(custom), File type/resource: TAR (of TXT)). Below the table, there are three lines of text: 'Raw data provided as supplementary file', 'Processed data included within Sample table', and 'Processed data provided as supplementary file'.

Step 3: Define groups in dataset. In general, a dataset would consist of two groups: cancer data and normal data

The screenshot shows the GEO2R interface for dataset GSE38959. The 'Samples' section has a 'Define groups' dropdown open, with the input field containing 'TNBC'. A list below shows 'TNBC' and 'Normal' assigned to their respective groups. The main table lists 47 samples, all of which are currently unselected.

Group	Accession	Title	Disease state	Cell type	Gender	Characteristics	Tissue	Sample ID
-	GSM952671	triple negative breast cancer cells T253	triple negative breast cancer cells	triple negative breast cancer	triple negative breast cancer cells	female	age (y): 49	
-	GSM952672	triple negative breast cancer cells T359	triple negative breast cancer cells	triple negative breast cancer	triple negative breast cancer cells	female	age (y): 55	
-	GSM952673	triple negative breast cancer cells T362	triple negative breast cancer cells	triple negative breast cancer	triple negative breast cancer cells	female	age (y): 37	
... 44 more rows ...								

Step 4: Assign relevant rows to the group by first selecting the rows and then clicking on the group to assign to (e.g., TNBC). Successful assignment is reflected when the rows are coloured with the group colour (e.g., blue)

The screenshot shows the same dataset after rows have been assigned to the 'TNBC' group. The rows for samples GSM952692, GSM952694, GSM952695, GSM952696, and GSM952697 are now colored blue, matching the group color. The main table lists 47 samples, with 43 selected and assigned to the 'TNBC' group.

Group	Accession	Title	Disease state	Cell type	Gender	Characteristics	Tissue	Sample ID
TNBC	GSM952692	triple negati	triple negative breast cancer cells	triple negative breast cancer	triple negative breast cancer cells	female	age (y): 64	
TNBC	GSM952694	triple negati	triple negative breast cancer cells	triple negative breast cancer	triple negative breast cancer cells	female	age (y): 52	
TNBC	GSM952695	triple negati	triple negative breast cancer cells	triple negative breast cancer	triple negative breast cancer cells	female	age (y): 42	
TNBC	GSM952696	triple negati	triple negative breast cancer cells	triple negative breast cancer	triple negative breast cancer cells	female	age (y): 63	
TNBC	GSM952697	triple negative breast cancer cells T445	triple negative breast cancer cells	triple negative breast cancer	triple negative breast cancer cells	female	age (y): 39	
TNBC	GSM952698	triple negative breast cancer cells T453	triple negative breast cancer cells	triple negative breast cancer	triple negative breast cancer cells	female	age (y): 50	
TNBC	GSM952699	triple negative breast cancer cells T392	triple negative breast cancer cells	triple negative breast cancer	triple negative breast cancer cells	female	age (y): 46	
TNBC	GSM952700	triple negative breast cancer cells T78	triple negative breast cancer cells	triple negative breast cancer	triple negative breast cancer cells	female	age (y): 45	
Normal	GSM952901	normal mammary gland ductal cells N19	normal mammary gland ductal cells	normal	mammary gland ductal cells	female	age (y): 55	
Normal	GSM952902	normal mammary gland ductal cells N21	normal mammary gland ductal cells	normal	mammary gland ductal cells	female	age (y): 52	
Normal	GSM952903	normal mammary gland ductal cells N40	normal mammary gland ductal cells	normal	mammary gland ductal cells	female	age (y): 35	
Normal	GSM952904	normal mammary gland ductal cells N40	normal mammary gland ductal cells	normal	mammary gland ductal cells	female	age (y): 42	
Normal	GSM952905	normal mammary gland ductal cells N59	normal mammary gland ductal cells	normal	mammary gland ductal cells	female	age (y): 42	
Normal	GSM952906	normal mammary gland ductal cells N43	normal mammary gland ductal cells	normal	mammary gland ductal cells	female	age (y): 42	
... 33 more rows ...								

Step 5: Generate GEO2R analysis by clicking on “Top 250” in GEO2R tab

The screenshot shows the GEO2R interface with the 'Samples' table displayed. The table lists various samples with their metadata. Below the table, there is a navigation bar with tabs: GEO2R, Value distribution, Options, Profile graph, and R script. A 'Quick start' section provides instructions for using the tool. In the center, there is a green button labeled 'Top 250' which is highlighted with a red box. Below this button is another button labeled 'Save all results'.

Step 6: Export the results. (i) Select the relevant columns using “Select Columns”. Check the following columns: logFC, ID, Gene symbol. Confirm the selection by clicking “Set” button and (ii) Select “Save all results”

The screenshot shows the 'Select columns' dialog box overlaid on the GEO2R interface. The dialog box has two main sections: 'Data columns' and 'Annotation columns'. Under 'Data columns', the 'logFC' checkbox is checked and highlighted with a red box. Under 'Annotation columns', the 'ID' and 'Gene symbol' checkboxes are checked and highlighted with a red box. At the bottom of the dialog box, there is a 'Set' button which is also highlighted with a red box.

The screenshot shows the GEO2R interface with the 'Selected 43 out of 47 samples' table displayed. The table lists genes with their logFC values. Below the table, there is a navigation bar with tabs: GEO2R, Value distribution, Options, Profile graph, and R script. A 'Quick start' section provides instructions for using the tool. In the center, there is a green button labeled 'Save all results' which is highlighted with a red box.

Step 7: The results are returned as a webpage in the format as shown. Save the result in a .csv file. The order of the columns of the .csv file should be (1) Gene, (2) Entrez and (3) Fold change. These columns correspond to Gene.symbol, ID and logFC in the GEO2R returned webpage. Note that the columns in the .csv are comma delimited and the values in the column are without quotation marks.

GEO2R webpage format

ID	logFC	Gene.symbol
"19472"	"-4.4554962"	"ANLN"
"8710"	"-4.7379997"	"TPX2"
"8432"	"-4.8295826"	"NEK2"
"25240"	"-3.142165"	"HIST1H2AG"
"43277"	"-3.8850869"	"UHRF1"
"18972"	"-4.2896407"	"KIF20A"
"16543"	"-5.2292227"	"KNL1"
"27754"	"-4.25102"	"TTK"
"34008"	"-4.5610271"	"BIRC5"
"11421"	"-4.2516085"	"KIF20A"
"36067"	"-4.5160093"	"BIRC5"

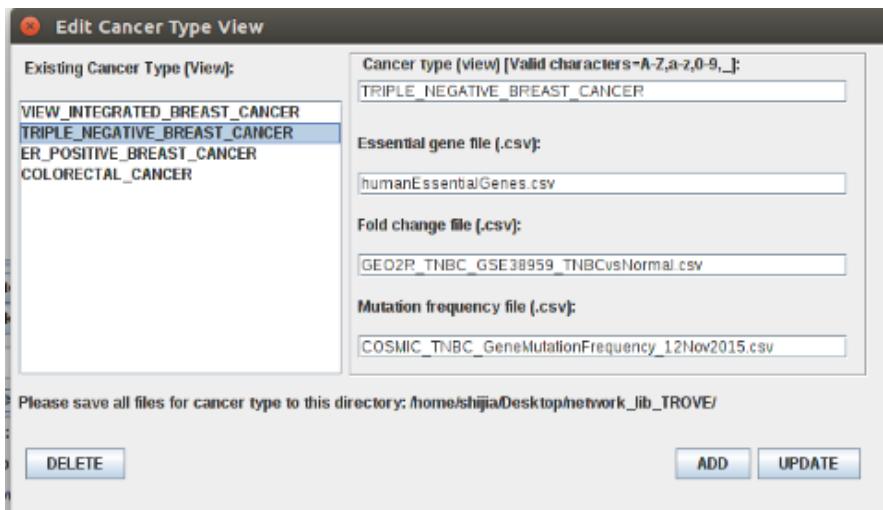
.csv file format

Gene	Entrez	Fold change
SCGB1D1	10648	-6.6087739
PIP	5304	-6.1815279
SCGB2A2	4250	-6.162622
COL17A1	1308	-5.5276956
WIF1	11197	-5.3716743
PI15	51050	-5.1684623
EDN3	1908	-4.8708207

TROVE GUI

Step 1: Select “Edit Cancer Type View” from the “Edit” menu

Step 2: Add the new cancer type by filling up the new cancer type (view), essential gene file, fold change file and mutation frequency file. The cancer type (view) is the name of the view. For essential gene file, the user can use the default humanEssentialGenes.csv provided.

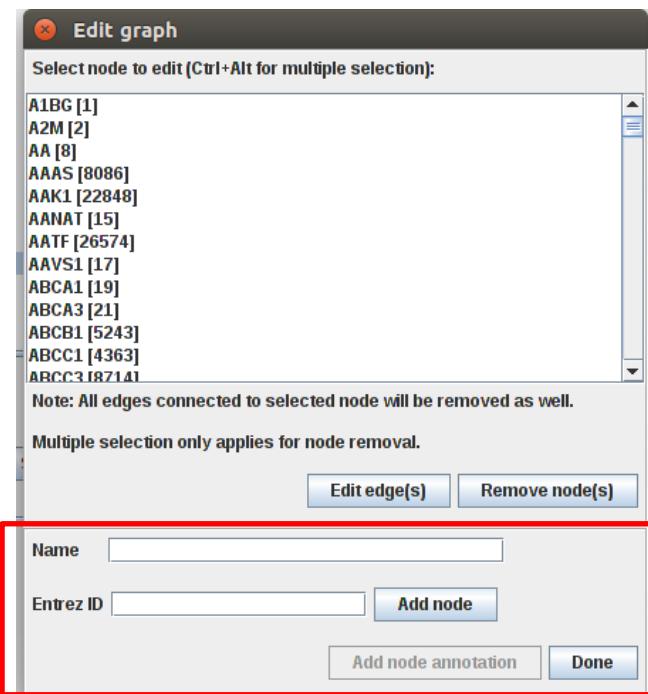


3.4 Editing of Graph

Users can modify the graph by selecting “Edit Graph” from the “Edit” menu

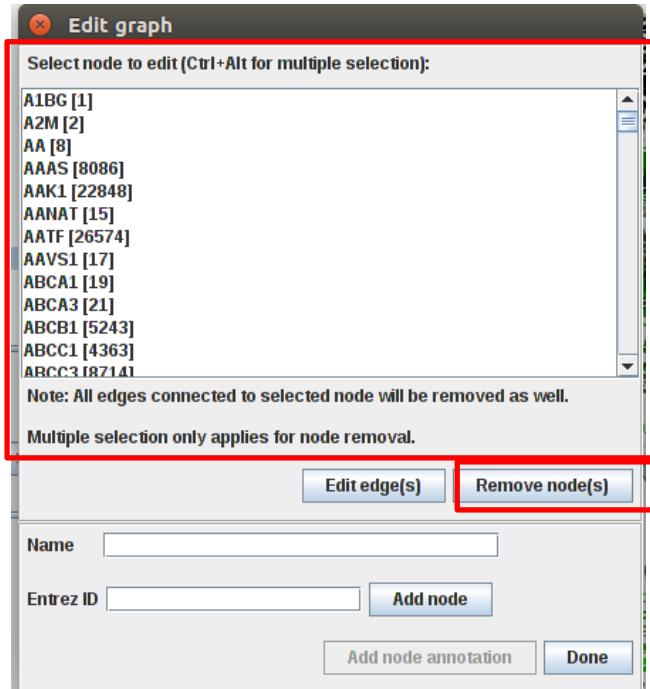
Add node

Enter the “Name” and “Entrez ID” for the new node and click “Add node”. The Entrez ID for the node can be found from <https://www.ncbi.nlm.nih.gov/gene/>



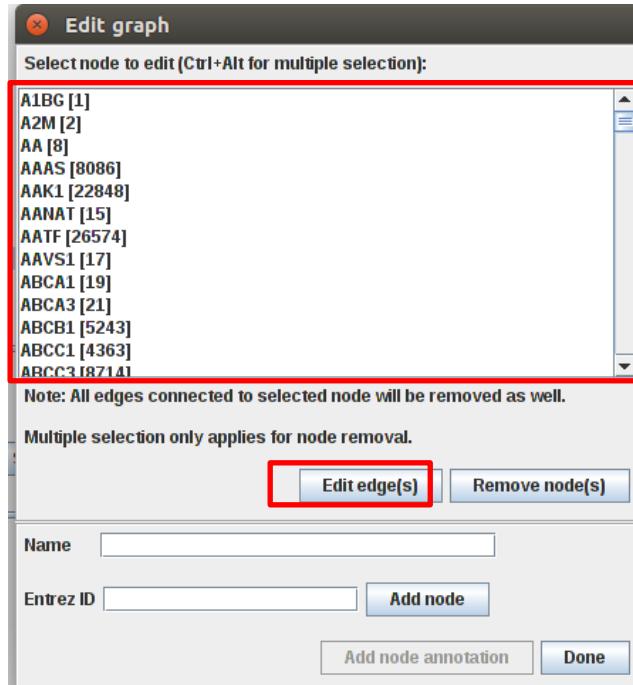
Delete node

Select the node to be removed and click “Remove node(s)”

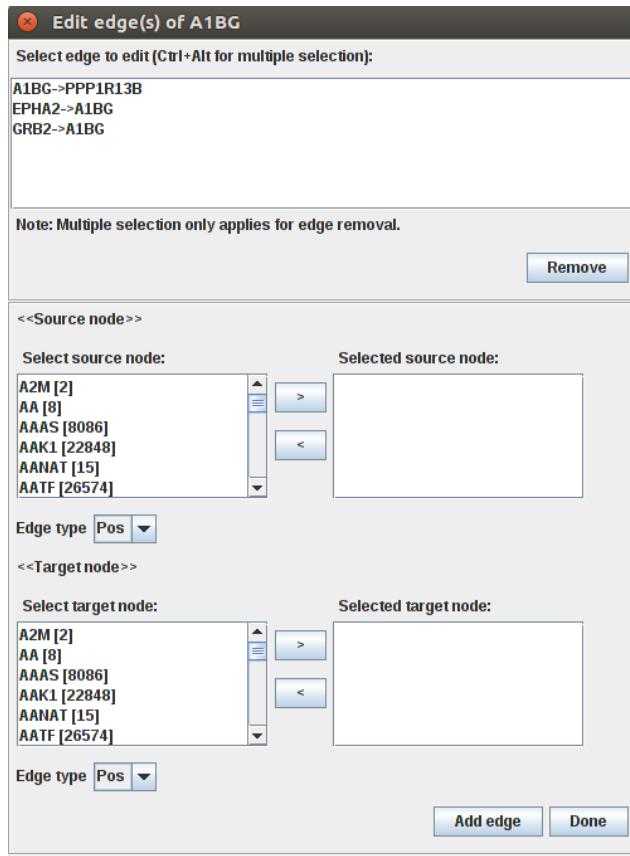


Add edge

Step 1: Select the node whose edge requires editing then click “Edit edge(s)”

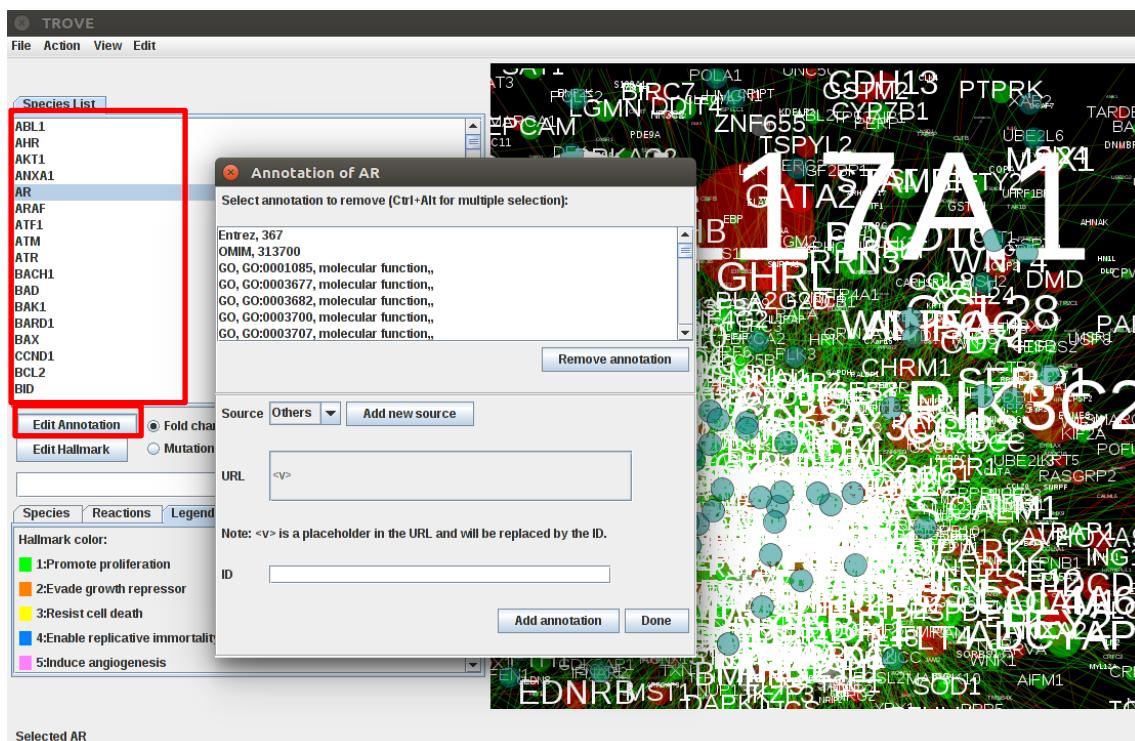


Step 2: Select the edge to edit. Users can either remove the edge, or add new incoming and outgoing edges of the selected node (e.g., A1BG).



3.5 Editing of Node Annotation

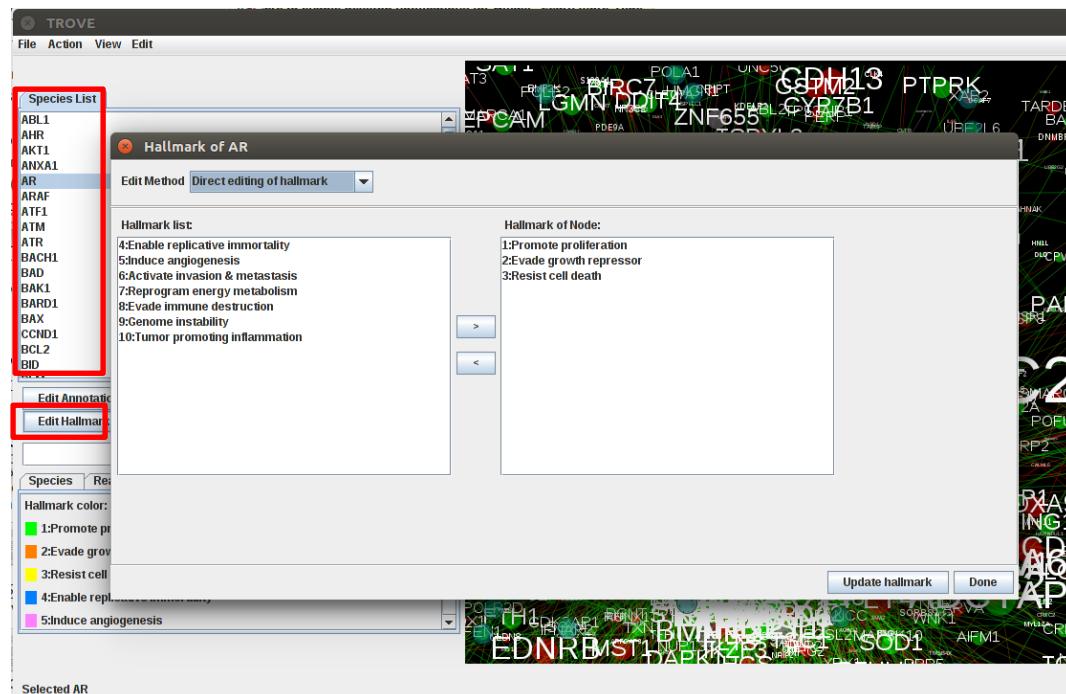
Users can modify annotation of nodes by selecting the node from the “Species List” tab, then clicking the “Edit Annotation” button. Users can remove existing annotation or add new ones.



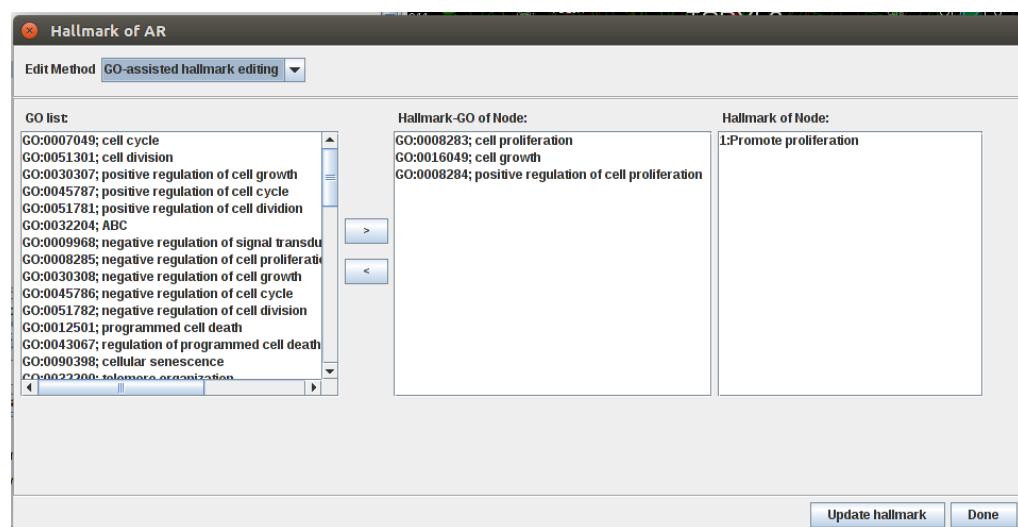
3.6 Editing of Hallmark Annotation

Users can modify hallmark annotation by selecting the node from the “Species List” tab, then clicking the “Edit Hallmark” button.

Direct Annotation: User selects the hallmarks directly from a given list



GO-assisted Annotation: User selects GO-terms from a given list. The hallmarks are automatically assigned.



3.7 Editing of Hallmark-GO Mapping Used for Hallmark Annotation

Users can modify the hallmark-GO mapping that is used for hallmark annotation by selecting “Edit Hallmark GO Mapping” from the “Edit” menu.

Existing hallmark-GO mapping can be deleted and modified. New mappings can also be added. When a hallmark is deleted, all hallmark-GO mappings related to the hallmark will be deleted. Alternatively, users can delete specific mappings of a particular hallmark by using the “Delete GO Term” button.

“Update” button does the following:

Hallmark A = hallmark selected in combobox

Hallmark B = hallmark selected in Hallmark List

GO term C= GO term selected in Textfield

GO term D = GO term selected in GO List

When hallmark A is the same as hallmark B AND GO term C is the same as GO term D, the existing mapping (Hallmark B-GO term D) will be updated with the new GO description and type

When hallmark A is the same as hallmark B BUT GO term C is different from GO term D, the existing mapping (Hallmark B-GO term D) will be updated with new GO ID, description and type

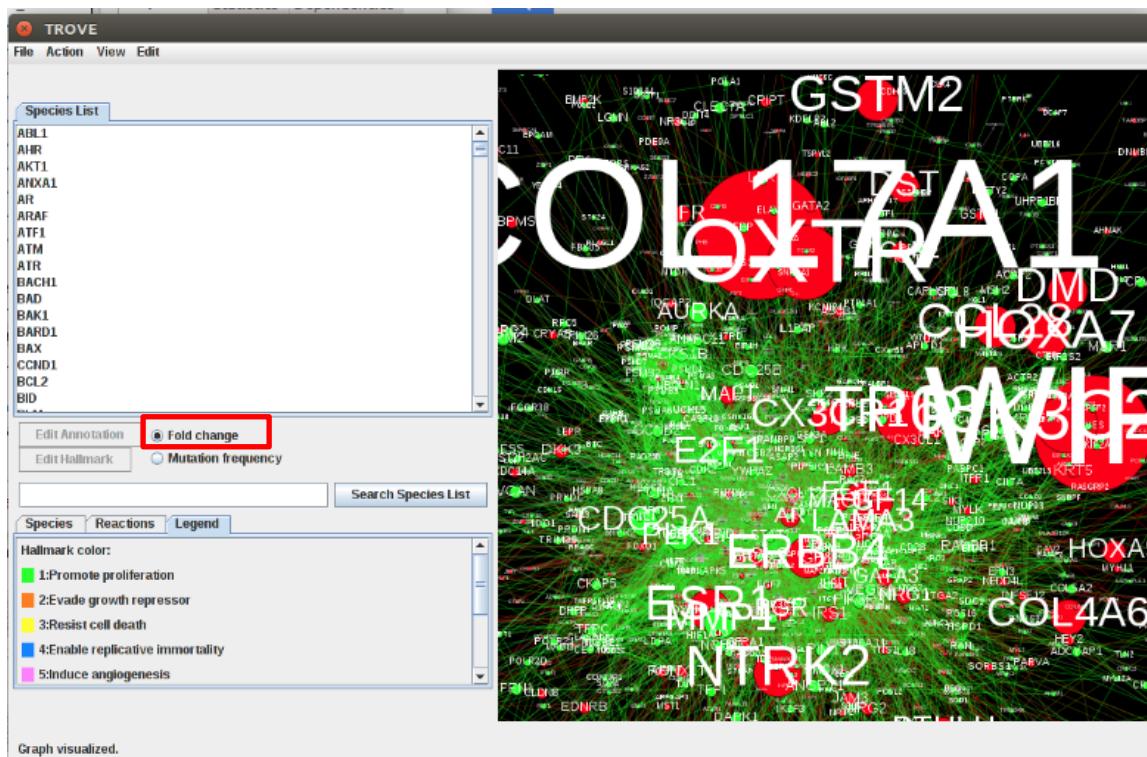
When hallmark A is different from hallmark B, the existing mapping (Hallmark B-GO term D) will be removed and a new mapping (Hallmark A-GO term C) will be added if it does not already exists. If the new mapping (Hallmark A-GO term C) exists, it will be updated with the new GO description and type.

The screenshot shows the 'Edit Cancer Hallmark Mapping' dialog. On the left, under 'Hallmark List', there is a list of 10 items: 1:Promote proliferation, 2:Evade growth repressor, 3:Resist cell death, 4:Enable replicative immortality, 5:Induce angiogenesis, 6:Activate invasion & metastasis, 7:Reprogram energy metabolism, 8:Evade immune destruction, 9:Genome instability, 10:Tumor promoting inflammation. Under 'GO List', there is a placeholder text field. On the right, there are four input fields: 'Hallmark' (containing '1:Promote proliferation'), 'GO ID' (empty), 'GO Description' (empty), and 'GO Type' (containing 'biological process'). At the bottom, there are four buttons: 'Delete Hallmark', 'Delete GO Term', 'Update', and 'Add'.

3.8 Visualization of Expression Fold Change and Mutation Frequency

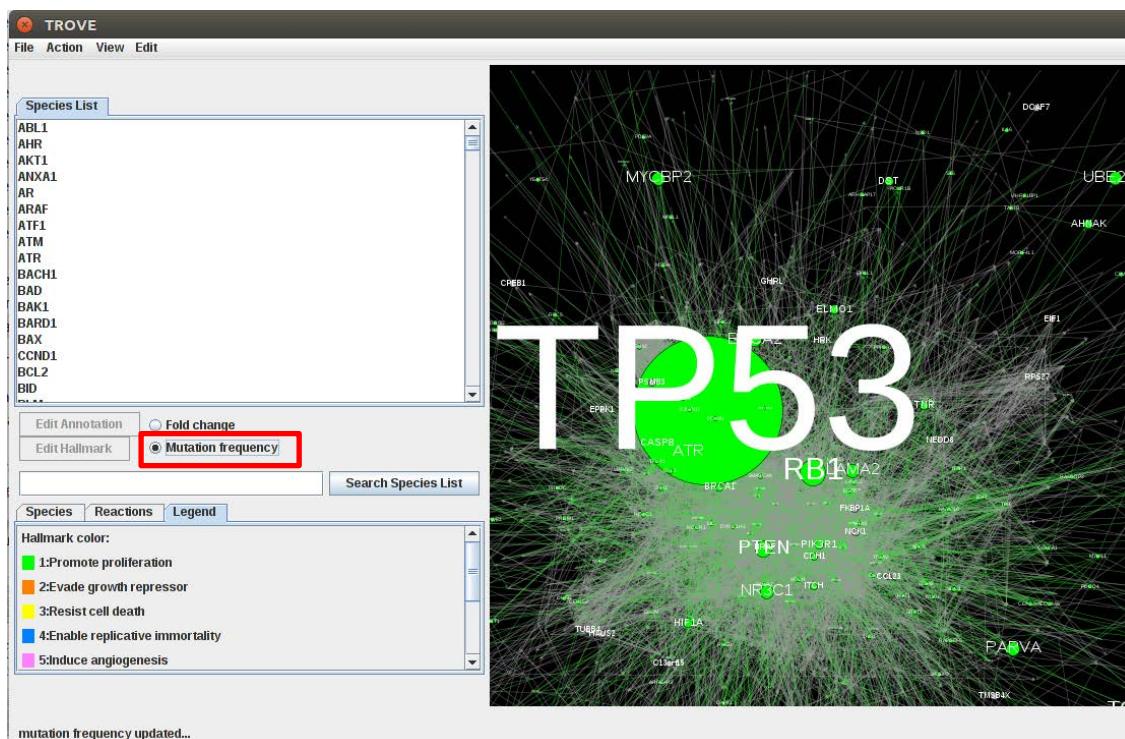
Expression Fold Change: Select the “Fold change” radio button

The size of the node of indicates the extent of fold change. Larger sized nodes indicate large fold change. Red implies under-expression and green implies over-expression.



Mutation Frequency: Select the “Mutation frequency” radio button

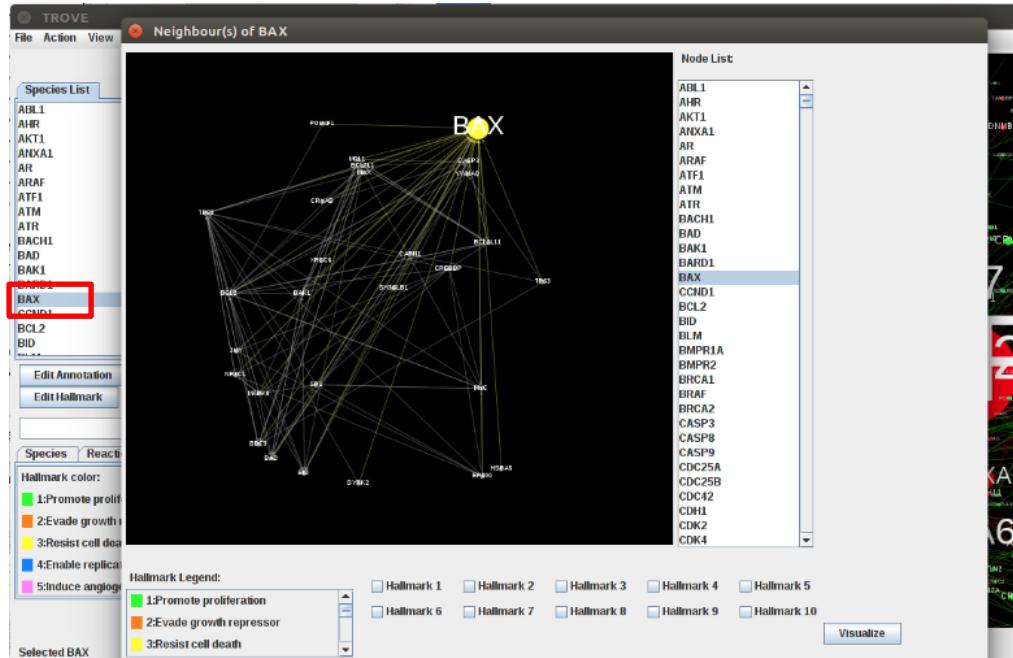
The size of the node of indicates the extent of mutation. Larger sized nodes indicate greater extent of mutation.



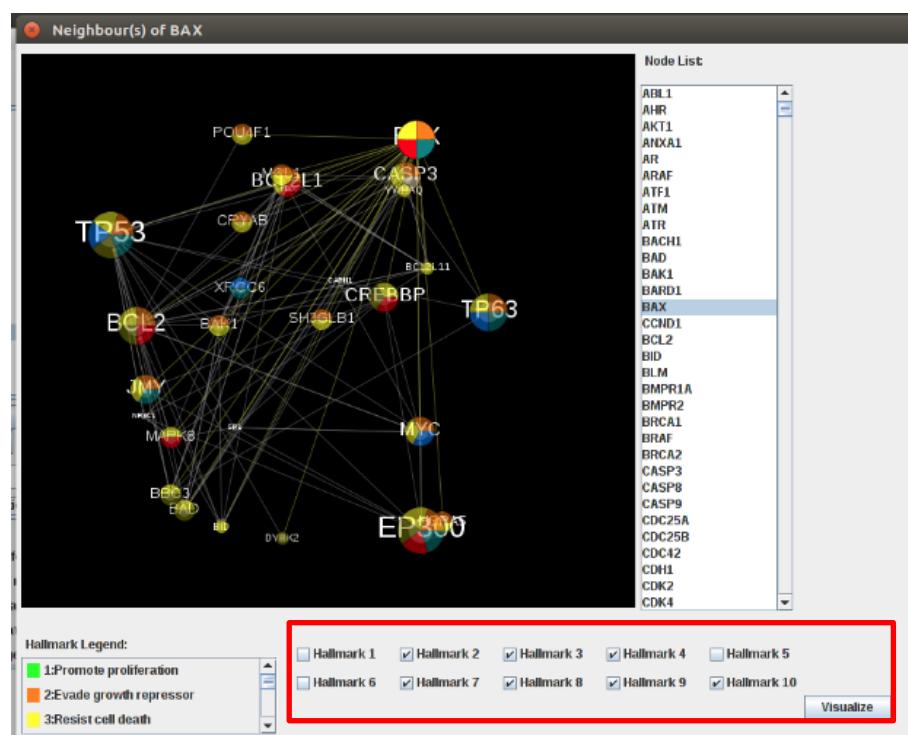
3.9 Visualization of Node Neighbourhood

Select a node from the “Species List” and perform a right-mouse button click.

The user can view the immediate neighbours of other nodes by selecting (left-mouse click) from the node list in the Neighbour panel.



The user can view the hallmarks of the neighbouring nodes by selecting the desired hallmarks and clicking on “Visualize” button. The hallmarks are shown as multi-coloured pie-charts where each colour represent one selected hallmark. The colour associated with each hallmark can be found in the Hallmark Legend.



3.10 Visualization of Pathway(s) Between a Pair of Nodes

Select “View Pathway Between Node Pair” from “View” menu

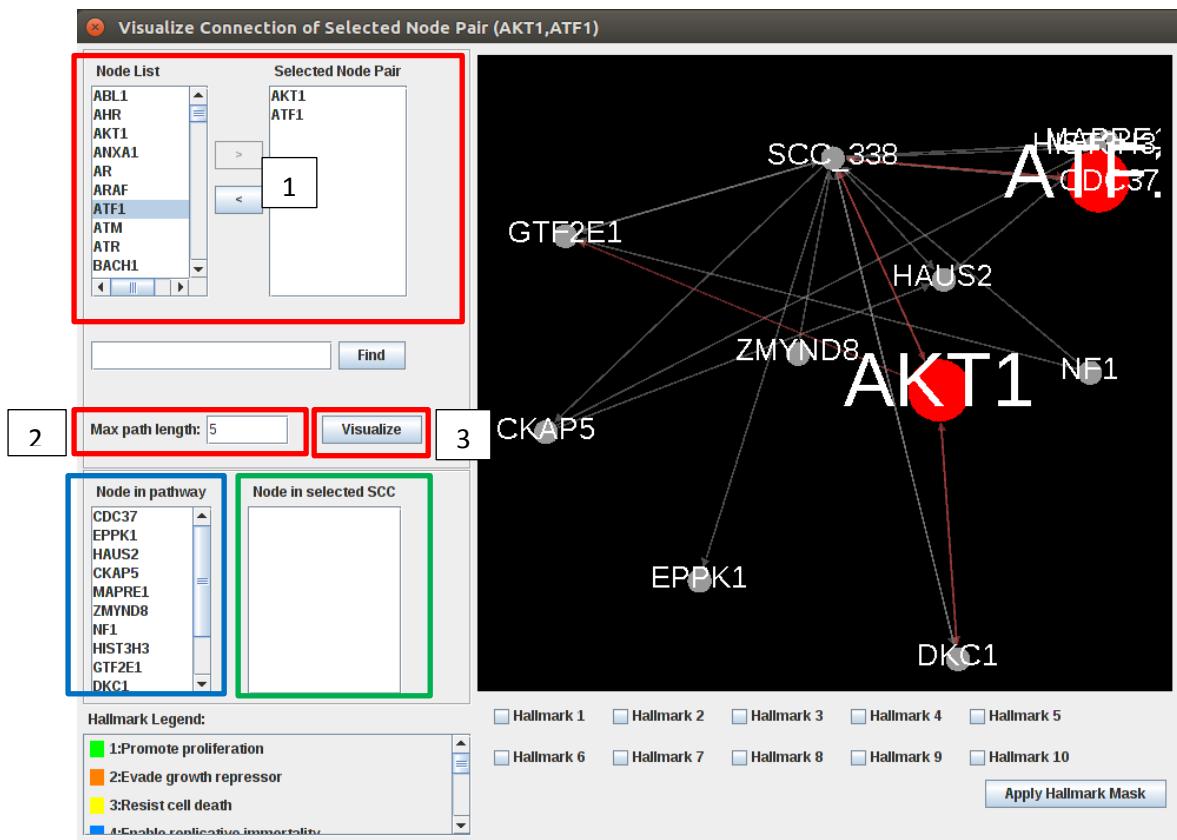
Step 1: Select the node pairs from the node list

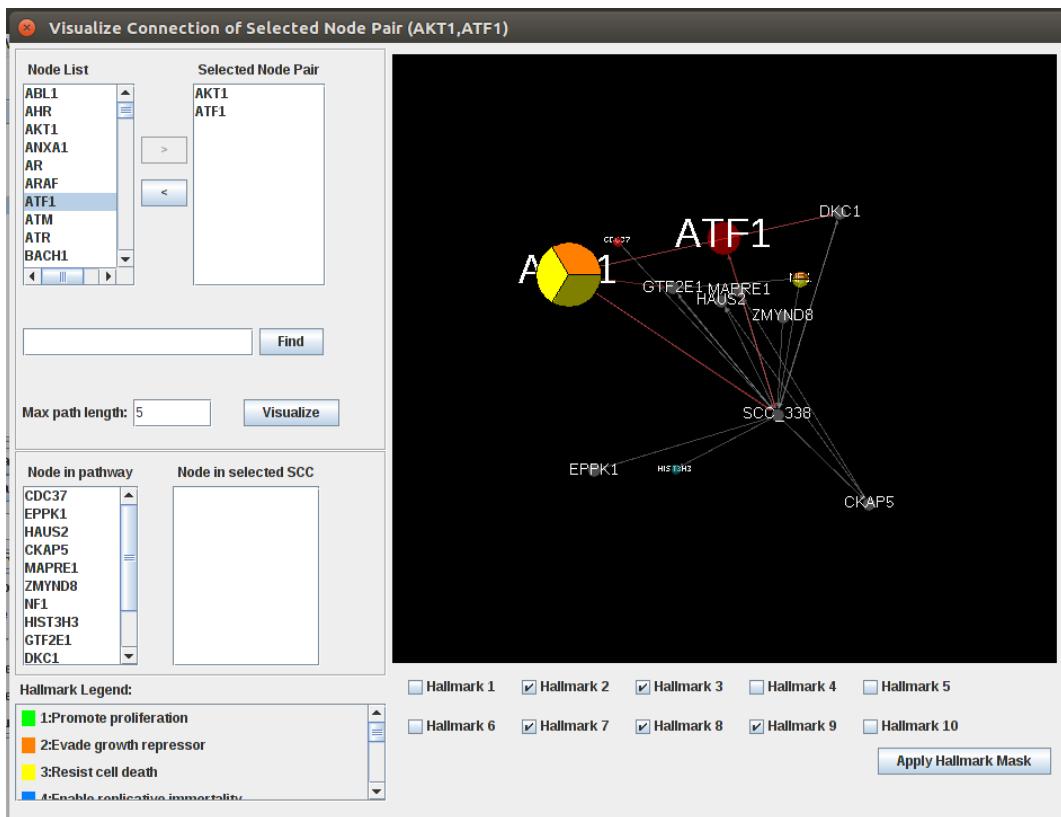
Step 2: Specify the maximum path length

Step 3: Click on “Visualize” button

The graphical result is shown in the black panel and the node pairs are highlighted in red.

The nodes in the pathway are listed in “Node in pathway” list (blue highlighted box). If the nodes are part of a strongly-connected component (SCC), then, the SCC is collapsed in the graph. Users can view the nodes in the SCC by selecting the SCC from the “Node in pathway” list. The nodes in the selected SCC will be displayed in “Node in selected SCC” (green highlighted box).





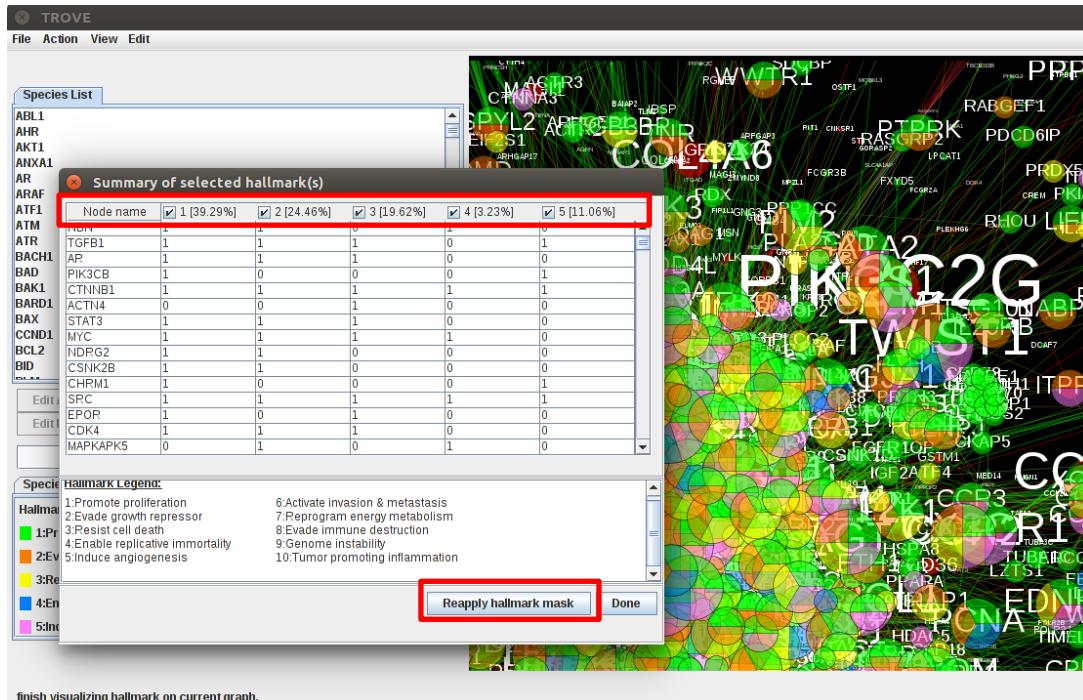
The user can view the hallmarks of the pathway nodes by selecting the desired hallmarks and clicking on “Visualize” button. The hallmarks are shown as multi-coloured pie-charts where each colour represent one selected hallmark. The colour associated with each hallmark can be found in the Hallmark Legend.

3.11 Visualization of Hallmarks

Select “View Hallmark” from “View” menu

The user can select the hallmarks that he/she wishes to visualize. The node of the selected hallmark can also be exported for further online analysis.

A summary table of the nodes with the selected hallmarks is provided, together with the percentage of nodes in the network annotated with the hallmarks. The hallmarks are shown as multi-coloured pie-charts where each colour represent one selected hallmark. The colour associated with each hallmark can be found in the Hallmark Legend.



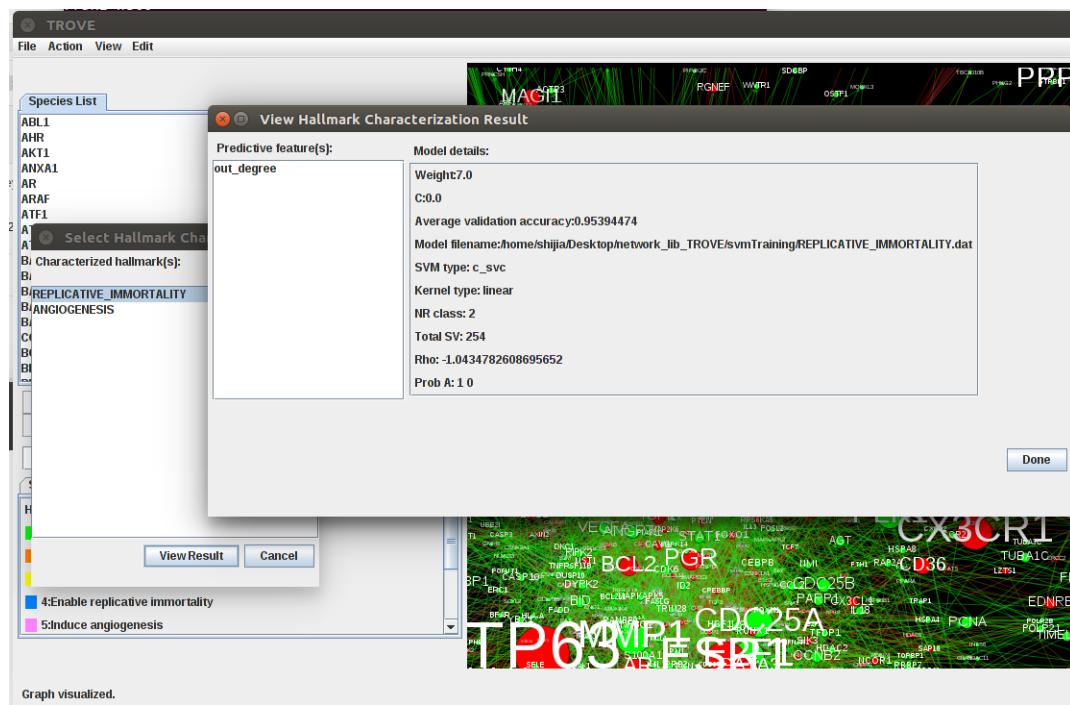
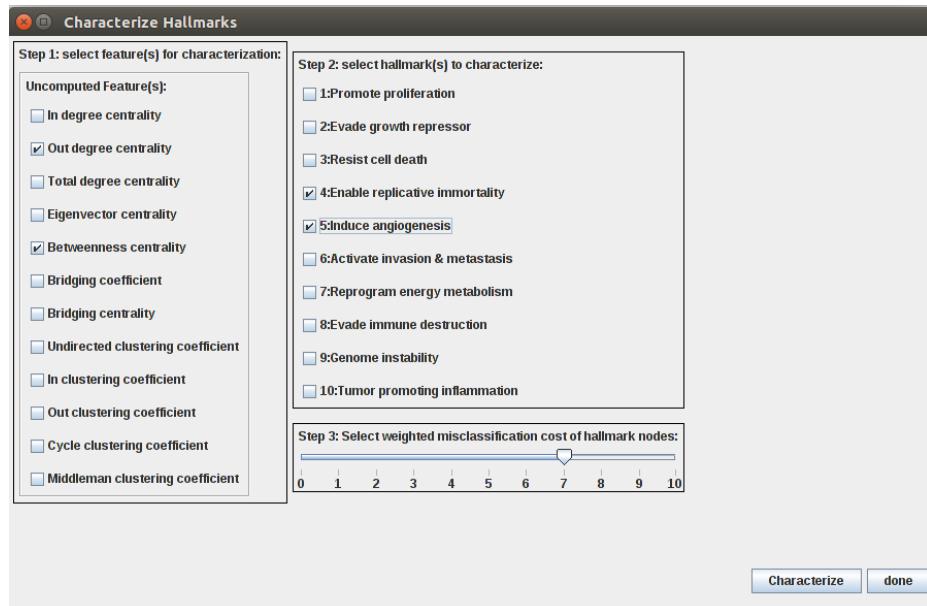
The user can modify the selected hallmarks by checking or unchecking the check boxes in the “Summary of selected hallmark(s)” dialog box and click “Reapply hallmark mask” to visualize a subset of chosen hallmarks.

3.12 Network Topology-based Characterization of Hallmarks

Select “Characterize Hallmarks” from “Action” menu

The characterization is performed using support vector machine (SVM). The SVM selects features using backward stepwise elimination and uses weighted misclassification cost (WMC) to address the issue of noisy labels and imbalanced data set. WMC proportionates the misclassification cost of the training data according to class such that misclassification of hallmark nodes (nodes that are annotated with the selected hallmark) incur greater cost. The hallmark characterization results summarize the configuration setting that was used to perform hallmark characterization, a list of predictive topological features used for constructing the final SVM model and prediction accuracies for the training data (cross-validation). Users can select the topological features used for characterization, the hallmarks to be characterized and the weighted misclassification cost (default value is 7).

The characterization results can be viewed by selecting “View Characterization Results” from “View” menu. The results are a set of predictive features and details of the SVM model.



3.13 Trouble-shooting

- Problem with PostgreSQL installation on Linux.

The following URL: <http://postgresguide.com/setup/install.html>

provides some helpful instructions on how to use homebrew to install PostgreSQL on Linux.

- Problem with language and locale setting of Linux.

TROVE assumes the language of the user's Linux system is in English and expects to find the files stored in the directory "Desktop". Note that in other languages, "Desktop" may be named as something else. To mitigate the problem, users can rename their Desktop directory as "Desktop" to facilitate TROVE to identify the right path for the input files.

- Problem with visualizing hallmark.

Sometimes the graph may take a while to load or refresh. Zooming in and out at the graph display panel can help to get this right. In the event that it still does not visualize properly, either reload the network again by selecting "Action" -> "Visualize Preloaded Network" or restart TROVE.

- Problem with input files.

Some of the input data files seems to be corrupted when I copied it using a USB thumbdrive from the Windows system to the Linux system. The users may want to try downloading the files directly from the website onto the Linux system.