Retina Workbench Users Guide

1. Installing Retina Workbench

Minimum requirements for successfully running Retina Workbench include

- 512 Mb of RAM or higher

- 1GHz CPU or higher

- Windows Me or XP

- Screen resolution of 1024x768 or higher (required by Cytoscape)

- Jre 1.4.2 and Cytoscape 2.3.2 installed

- An active internet connection

To install Retina Workbench, follow these steps:

1. Install jre 1.4.2 from http://java.sun.com/j2se/1.4.2/download.html and Cytoscape 2.3.2 from http://www.cytoscape.org/, if not already installed.

2. Add RetinaWorkbench.jar file in the /plugins directory in Cytoscape

Once all of the above steps are executed successfully, and RetinaWorkbench.jar file has been copied into the /plugins directory in Cytoscape, Retina Workbench will be automatically added to the "plugins" menu in Cytoscape and will be ready for use.

2. Launching Retina Workbench

2.1 Starting RetinaWorkbench

Start Cytoscape by double-clicking on cytoscape.bat file or cytoscape.exe file. Once Cytoscape is fully loaded, click on Plugins → Retina Workbench in the top menu. A start-up window where you can register for a new account, connect to the database, and access the Networks window, Expression data windows, MyAccount window, and Help window, is automatically loaded.

2.2 Registering for an account

As a new user, you can easily register for a new account that will enable you to use all RetinaWorkbench functions. To register, click on the "Register new user" button in the main window of RetinaWorkbench. Once the registration window opens, fill out.

Retina Workbench provides an easy automated way of account creation. A new user can register for an account by clicking on "Register new user" and entering his/her first name, last name, email, desired username and password, and a new account will be created immediately using the entered information. Only one account per email address and username is allowed.

2.3 Logging in

To connect to a database, fill out the database server IP address (currently **129.186.93.137**), your user name, and password in the corresponding fields, and click on "Connect to server" button. Once you are connected, the Networks, Expression data, My Account, and Help buttons are automatically enabled. You are now ready to access the Retina database and use RetinaWorkbench functions.
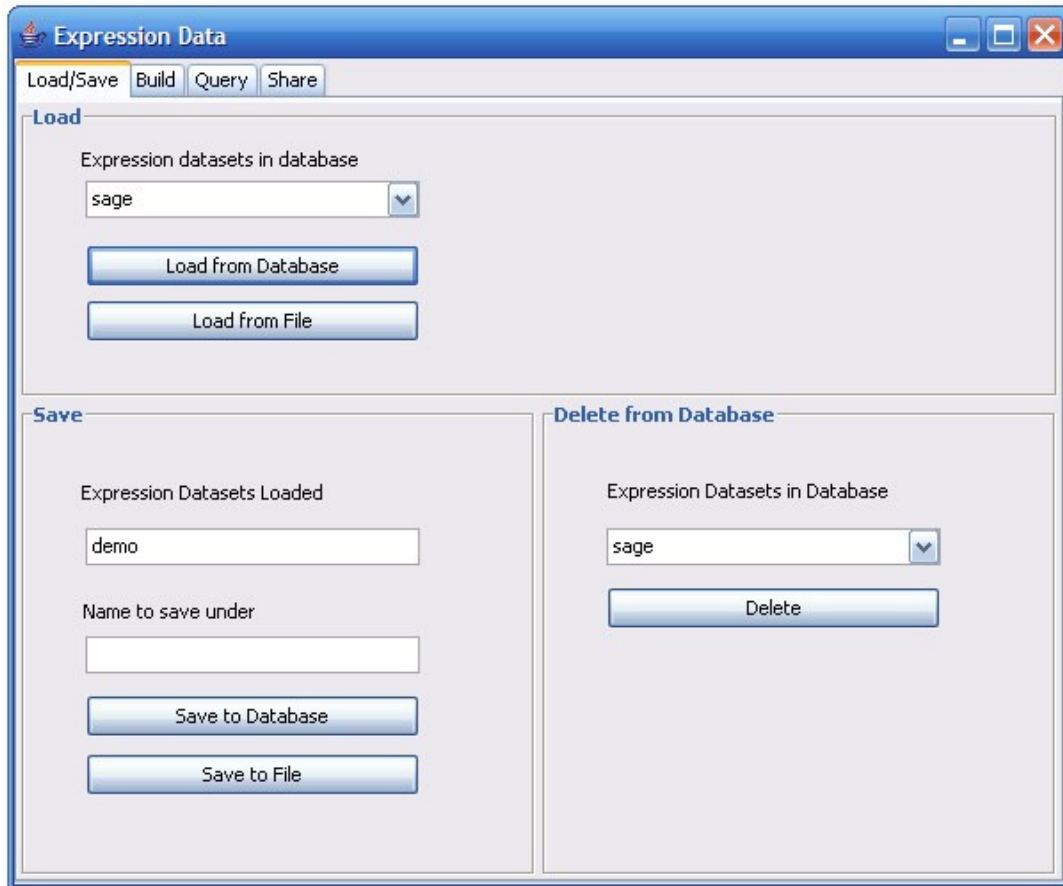

3. Expression data window

The Expression Data window consists of Load/Save, Build, Query, and Share tabs.

3.1 Loading/saving/deleting expression data

The Load / Save panel in the Expression data window was created for managing expression data.

3.1.1 Loading expression data

3.1.1.1 Loading from file

To load a dataset from file, click on the"Load from file" button in the Load/Save panel and select a file you wish to load. A file has to be correctly formatted; the format of expression files in Retina Workbench is consistent with the Cytoscape expression file

format. Once the dataset is loaded, it will be listed among the loaded expression datasets in Expression window.

Expression data file format:

```
#examples
#label_cols     <tab> #data_cols
col1_name       <tab> col2_name       <tab> data_col1   <tab> data_col2 ….
Gene1_label1    <tab> gene1_label2    <tab> value1      <tab> value2 …
```

For Cytoscape, it is required that there are 2 label columns: a *gene name*, and a *description* column. Do not skip or leave the description column blank – it has to contain at least one character. Label column names or values **cannot contain spaces** or **dots**.

3.1.1.2 Loading from database

To load an expression dataset from the Retina database, select the name of the dataset from the list of available datasets in the "Load" section of the Load/Save panel and click on "Load from database" button. Once loaded, the name of the dataset will be displayed in the "Loaded datasets" list.

Loaded datasets can be used for querying or building correlation networks.

3.1.2 Saving expression data

3.1.2.1 Saving to file

To save an expression dataset to a text file, select the name of a loaded dataset from the "Loaded datasets" list and click on "Save to file" button. Select a location and the desired name of the file. A tab-delimited text file following a common Cytoscape format for expression files will be created.

3.1.2.2 Saving to database

To save a loaded expression dataset to database, select the name of the loaded dataset from the list, enter a name for the dataset to store under in the database, and click on "Save to database" button. If the name field is empty or equivalent to the current name of the dataset, it is assumed that the new dataset will replace an old one with the same name.

In order to replace an old dataset, you are required to have an expert or administrator account, to ensure that public data stored in the database is not corrupted, unless you were the original creator of this dataset. If you have an expert user account, you can decide if you would like the newly created dataset to be public (accessible to everyone for viewing and querying), or private (only accessible to you). For general users, all newly created datasets are private. You can later share your private dataset with the users of your choice by using the Share tab in the Expression data window.

Saving to database may take a few minutes or more if the dataset is large.

3.1.3 Deleting expression data

An expression dataset that you have created or to which you have been granted a "delete" right can be deleted from the database. Datasets that can be removed are listed in the "Delete" section of the Load / Save tab. To delete a dataset, select the dataset name and click on "Delete" button.

3.2 Building correlation networks

The Build tab of the Expression Data window allows for an easy construction of correlation networks from expression data, where nodes represent genes/proteins in the dataset and a link between any pair of nodes indicates that the corresponding genes/proteins are correlated with a coefficient larger or equal to the specified cutoff value. Genes that are not correlated with any other genes are excluded from the resulting network.
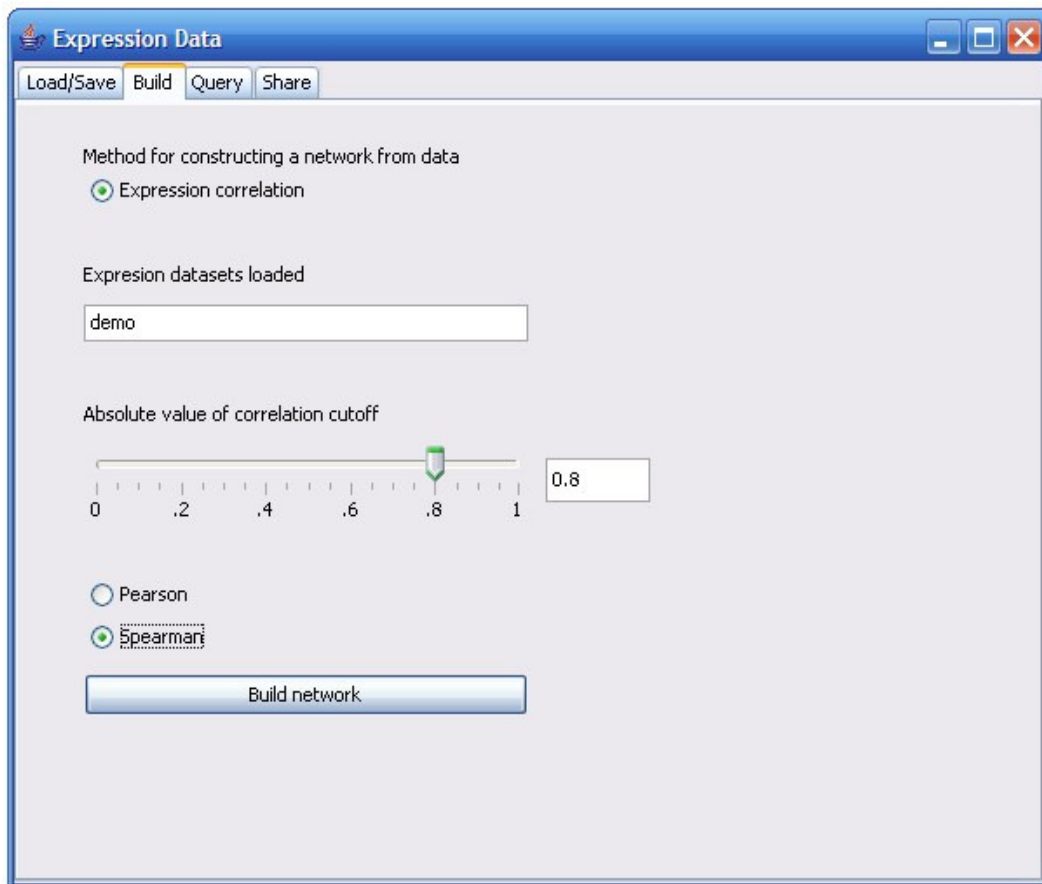
To build a correlation network, an expression dataset from which you wish to construct a network must be already loaded in the Expression data window. To load a dataset, use the Load/Save tab.

To construct a correlation network, select the desired expression dataset from the list of loaded datasets in the Build tab. Move the slider to select a desired cutoff value. Only links between genes for which the absolute value of correlation between their expression profiles is larger or equal to the cutoff value, will be drawn. The cutoff value ranges from 0 (weak or absent correlation) to 1 (perfect correlation). The higher the cutoff value, the fewer links will be included in the resulting network.

Select a correlation measurement you would like to use. Two very common measurements are Pearson correlation coefficient, which assumes linear correlations, and a non-parametric Spearman correlation.

Once the dataset, a cutoff value, and the correlation method are selected, click on the "Build correlation network" button. This operation may take a few minutes for the large datasets.

A resulting network will be automatically loaded in Cytoscape.

3.3 Querying expression data

Expression datasets present in the database or loaded in the Load/Save panel can be queried to extract useful information. Currently, RetinaWorkbench supports querying expression data to find genes correlated with all or any of the genes specified.

To query a loaded dataset, choose the "Query loaded" option and select the dataset among a list of loaded expression datasets. To query the database, choose "Query

database" option. You can either leave the list of datasets present in the database unselected, in which case all of the above datasets will be queried, or select one or more datasets from the list, in which case only the selected datasets will be searched.
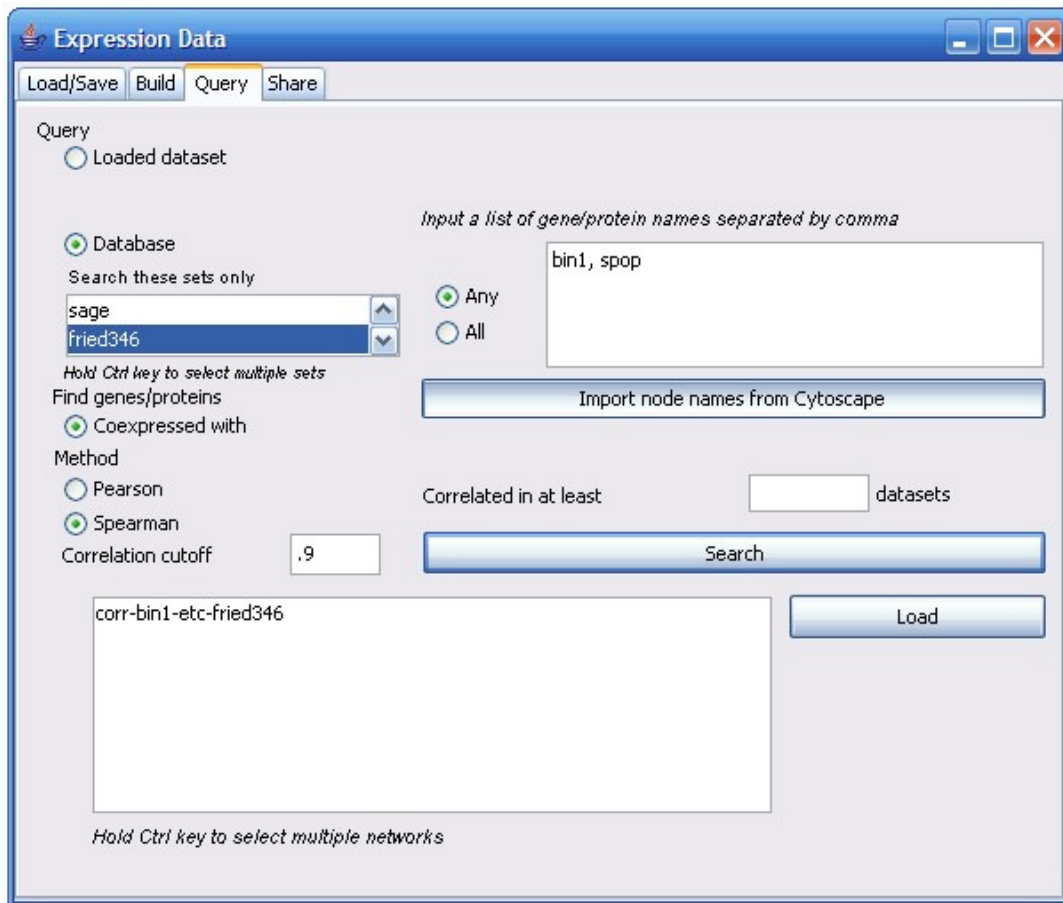
Select the "Find genes correlated with" option and pick a correlation method (Pearson or Spearman). Pick a cutoff value (ranging from 0 to 1, with 0 being the weakest/absent correlation to 1 being the perfect correlation). In the target genes box, enter names of target genes or click "Import selected names from Cytoscape". Importing selected names will transfer names of all the nodes currently selected in Cytoscape from all loaded networks into the target genes box. Select "all" or "any" option. "Any" option will result in retrieving genes correlated with *at least one* gene from the target genes list. "All" option will only retrieve genes that are correlated with *all* the genes on the list.

When "All" option is selected, each dataset is searched separately: for example, if gene "cfh" is co-expressed with a target gene "nsf" in one dataset and also co-expressed with target gene "spop" in a different dataset, but there is no dataset in which "cfh" is co-expressed with both "snf" and "spop", then "cfh" will not be included in the resulting network.

Resulting from this query is/are trimmed correlation network(s) consisting of only the target genes and genes co-expressed with them with (the absolute value of) a correlation coefficient larger or equal to the cutoff value.

When database is searched, the names of resulting networks are listed in the Results section of the Query tab. For each dataset searched, a trimmed correlation network will be built. Only networks that include at least two nodes and at least one link will be listed.

You can load some or all of the resulting networks into Cytoscape by selecting the names of resulting networks and clicking on the "Load" button in the Query tab. You can also build a composite network that would only include genes present in all of the selected result networks by selecting the networks to merge and clicking on the "Merge" button. This operation is equivalent to building an intersection network in the Merge tab of the Networks window.

If you are searching multiple datasets, you can indicate which genes to include in a resulting network specifying the minimum number of datasets in which a gene has to be correlated with any / all of the target genes. In this case, a candidate gene will be included in a resulting network only if it is co-expressed with the target genes in at least K datasets, where K is specified before. Similarly, a link between any pair of nodes (genes) is drawn only if these two genes are correlated in at least K datasets. As a result of this query, a single composite network is built and loaded into Cytoscape.

3.4 Sharing expression data

The Share tab enables you to share your private datasets with other users. As a creator of a dataset, you can grant or rescind a user of your choice a right to view, update, or delete the dataset. To assign the rights to your dataset, select the name of the dataset among the list of datasets you created, select a user whose rights you want to update, and check one or more options among "view", "update", and "delete". Click on the "Grant" button, and the user of your choice will have the right to view/load/query your dataset, a right to update it (save it under the same name, replacing the old dataset), or delete it from the database.
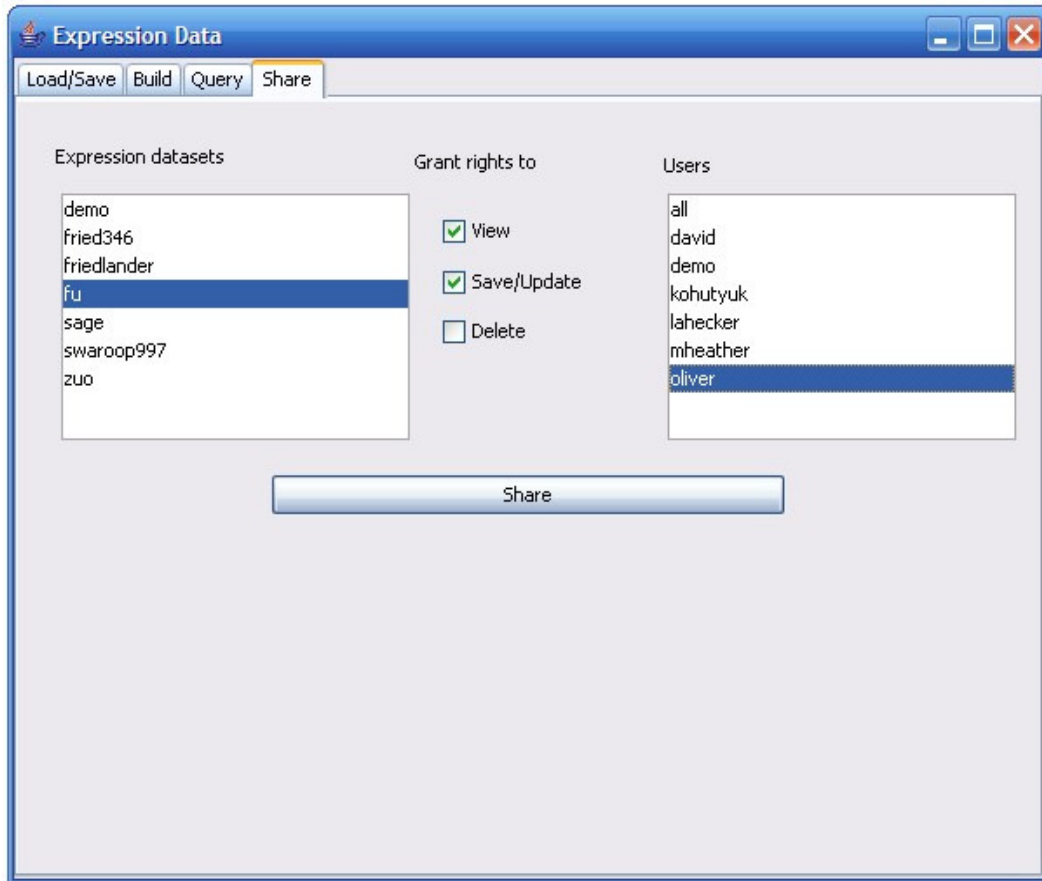
4. Networks window

4.1 Loading/saving/deleting networks

Similarly to expression data, you can load, save, and delete networks to and from the database by using the Load/Save tab of the Networks window.
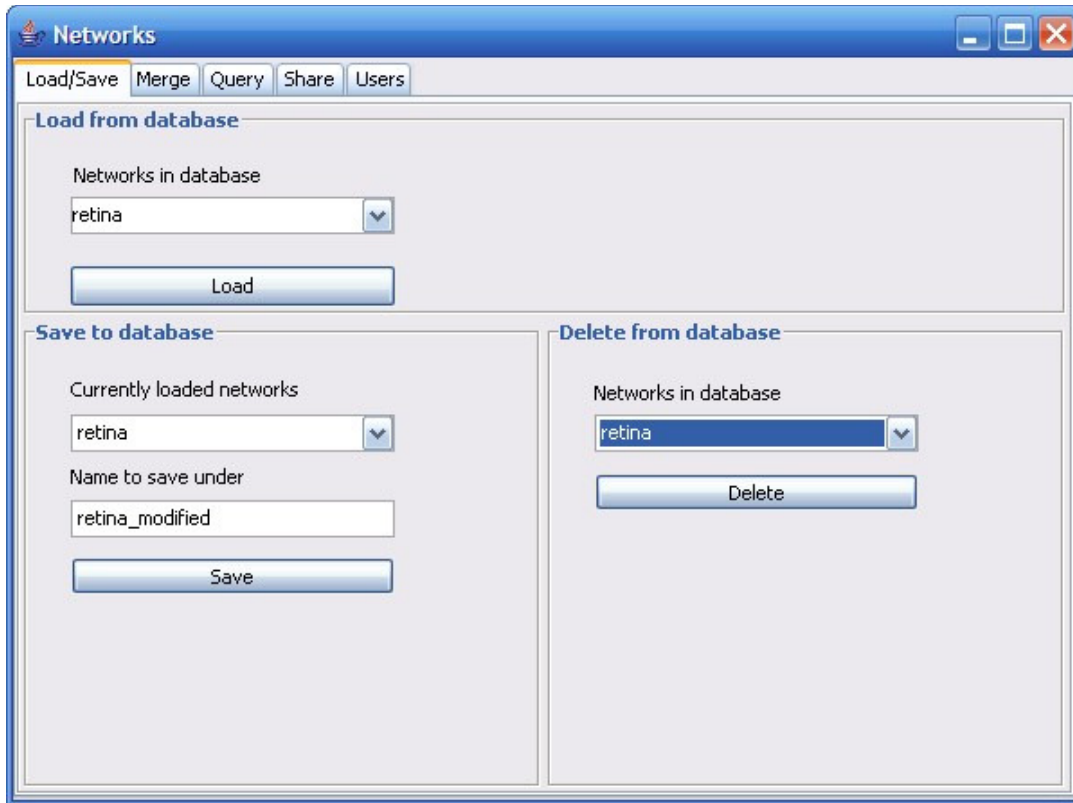
4.1.1 Loading networks

You can load a network from database by selecting a name of the network you would like to load and clicking on the "Load" button. Once loaded, the network will be displayed in Cytoscape.

4.1.2 Saving networks

To save a loaded network to the database, select the network in the list of loaded networks in the "Save" section of the Load/Save tab. Specify the name to save the network under. If no name is specified or if a network under this name already exists, the old network will be replaced by the new one only if you have an expert user account. As

a general user, you cannot modify existing public networks; if you wish to modify and save a network you must choose a name that does not already exist in the database.

Saving a large network to the database may take a few minutes or more.

4.1.3 Deleting networks

A list of networks you have the right to delete from the database is displayed in the "Delete" section of the Load/Save tab. You can always delete a network you created; you can also delete networks on which you have been granted a "delete" privilege. To delete a network, select the network name from the list and click on "Delete" button.

4.2 Merging networks

The Merge tab in the Networks window allows you to easily merge two or more the loaded networks into one network. Select two or more networks from the list of loaded networks in the Merge tab. Choose a "Union" or "Intersection" option and click on "Merge" button. A new network will be automatically loaded in Cytoscape.
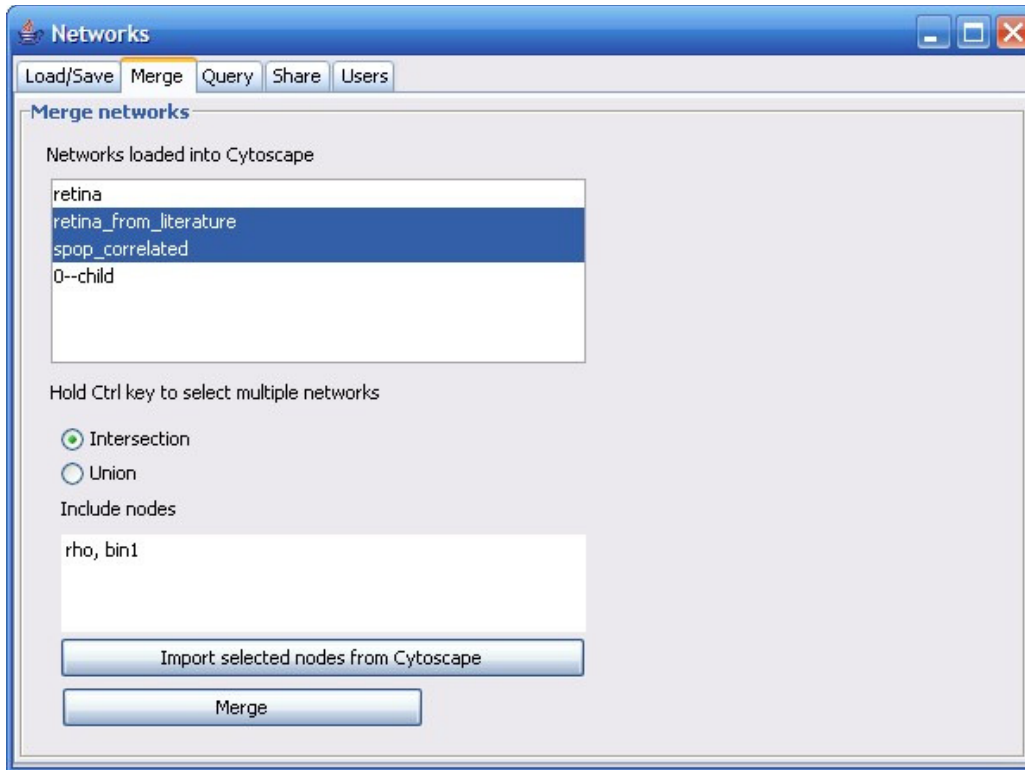
The "Union" option will build a network that includes nodes and links that are present in *at least one* of the original networks. This option is equivalent to a set union.

The "Intersection" option will result in a network that includes only nodes that are present in *all* of the original networks. This option is equivalent to a set intersection. It is useful when you would like to find shared nodes among various networks. You can also list nodes you would like to include in the new network; these nodes will be added to the resulting network even if they do not appear in all of the original networks. For example, if you obtained two trimmed correlation networks from searching for genes co-expressed with gene **mt3** (first network) and gene **spop** (second network), you might want to find genes common to both networks. If you choose the "Intersection" option, only nodes present in both networks will be included, and it is possible that **mt3** and **spop** will not be present in this list. To ensure that the seed proteins / nodes you are building the

network around are included, list them in the "Include nodes" section before clicking on the "Merge" button.
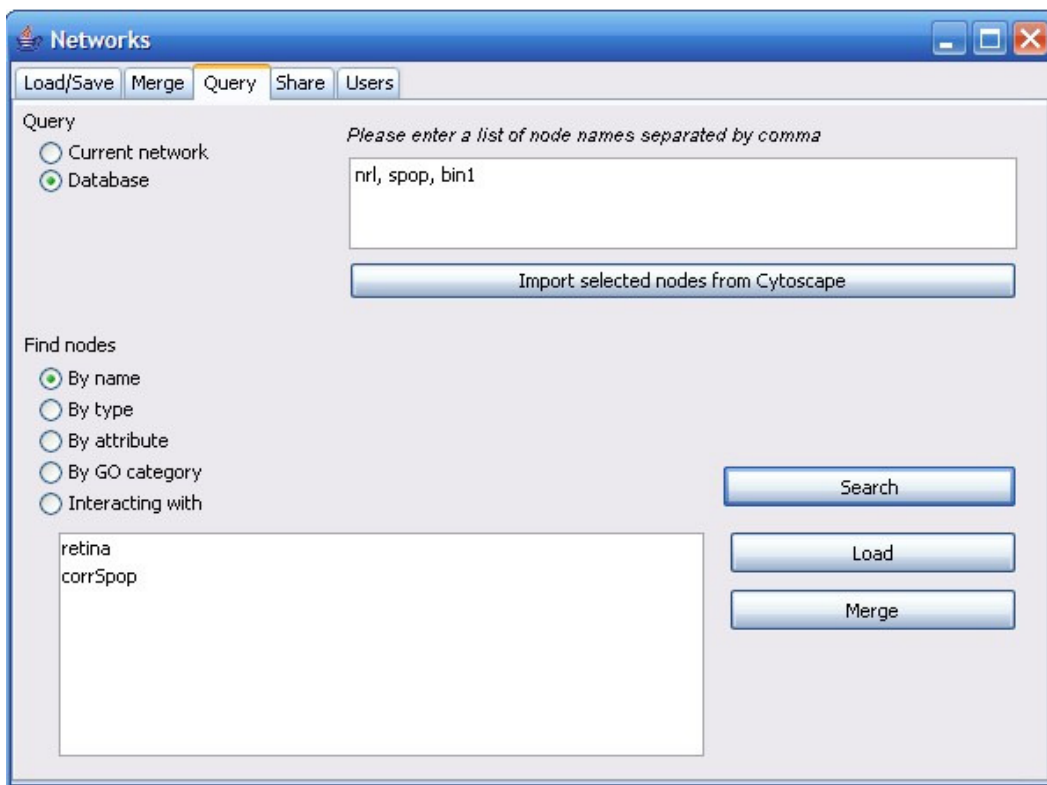
4.3 Querying networks

The Query tab of the Networks window provides multiple ways to query annotated networks and supports extraction of information of various types. Either the database or a current network in Cytoscape can be queried to find genes/proteins with a given name, property, Gene Ontology annotation, or nodes/genes connected to specified list of nodes/genes in a network with the connection holding a certain property. To query a current network, select the "current network" option in the Query tab. A "current network" denotes a network loaded in Cytoscape and selected in the left panel of Cytoscape window. To query networks stored in the database, choose the "query database" option. A list of available networks stored in the database (public networks and your own private networks) will be displayed. If no networks are selected on the list, *all* of the networks listed will be searched; if one or more networks are selected, only *selected* networks will be searched.

Results are displayed by highlighting the nodes fitting the search criteria in the current network (for "query current network" option) or providing a list of networks where nodes with desired property are highlighted (for "query database" option). These results can be used in a variety of ways: creating a new network from highlighted nodes and then further querying the new network, merging resulting networks, and re-using

results for further queries by transferring names of highlighted nodes into a new query in Expression data window or Networks window.

4.3.1 Querying by name

Querying genes by name provides an easy way to locate genes in the loaded network or in the database. Select whether the current network or the database should be queried, select the "find genes by name" option and type a list of gene names to search for in the Gene names field in the Query tab of the Networks window. To quickly transfer names of highlighted nodes in Cytoscape into the Gene names field, click on "Import nodes from Cytoscape" button. This is an easy way to avoid re-typing the names of nodes and to transfer names of nodes from a previous query to a new query.

4.3.2 Querying by node attribute

Selecting a querying by node attribute option will display a lit of all node attributes present in a current network or in the database (depending on whether the query current network or query the database option was chosen.) For each attribute name you select, all values of that attribute occurring in the network / the database will be listed. Once an attribute name and value are selected, clicking on the "Search" button will find all nodes in the current network or the database that are annotated with the given attribute. For example, you can search for nodes (genes) that are annotated with an attribute "organism" and attribute value "human", if such annotation has been created already.
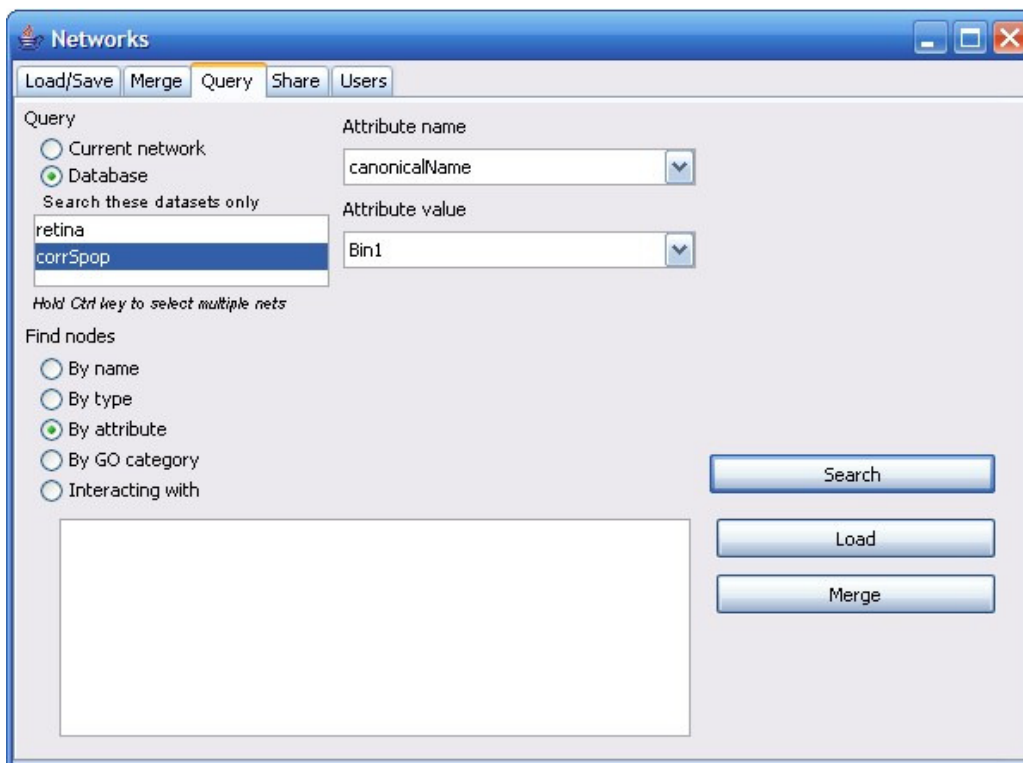
4.3.3 Querying by node type

A node type is an attribute denoting the type of the biological entity this node represents, such as "protein" or "small molecule". Querying by node type is essentially equivalent to querying by attribute, where the attribute is called "type". To query by node type, select a value among the possible node type values listed and click on "Search" button.

### 4.3.4 Querying by Gene Ontology category

Finding genes by their Gene Ontology annotation is a useful feature allowing retrieval of genes that belong to a certain Gene Ontology category without manual look-up of gene annotations and relationships between annotations in the GO graph.
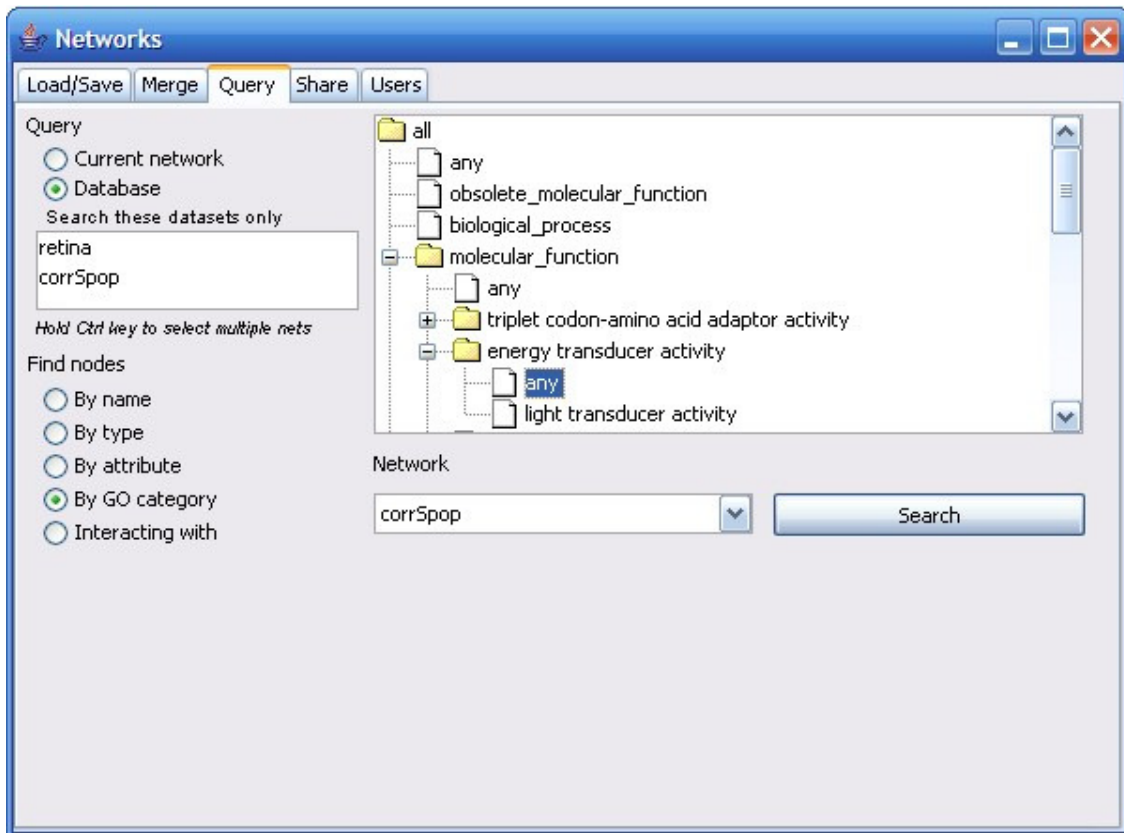
When "search by GO category" option is chosen, a Gene Ontology graph is loaded as a tree listing GO categories in three main areas: biological process, molecular function, and cellular component. Clicking on a category will list the subcategories. A GO category can be chosen at any level of specificity. To select a category, click on the category name, which will result in displaying child categories, and select "any" from the list of subcategories. Select a "query current network" option or "query database" option; if you query the database, you must select a single network in the database you would like to search. Clicking on "Search" button will search the selected network for nodes (genes) that belong to the chosen GO category or any of its subcategories. A resulting network with matching nodes highlighted will be automatically loaded in Cytoscape.

4.3.5 Querying by interaction

To find nodes connected to a list of specified nodes where the interaction is annotated with a certain attribute, use the "Find genes interacting with" option. Select a "current network" or "database" option and type a list of target names in the Names field. To avoid typing node names or to re-use results from previous query/queries (where the matching nodes were highlighted), click on "Import nodes from Cytoscape" button. This action will transfer the names of all highlighted nodes from all networks loaded in Cytoscape into the Names field. To search for nodes connected to target nodes regardless of the type of interaction, leave the interaction attribute name and value at "any". To specify the property of the interaction you are looking for, select the interaction attribute name among the attribute names listed and select a desired attribute value among the values listed. For example, you can search for nodes(genes) interacting with node "apoe" where the "coexpression" attribute is equal to "yes"; you can also search for nodes interacting with "cpna", "ccnd1", or "araf" where links were annotated with attribute "confidence" and attribute value "high".
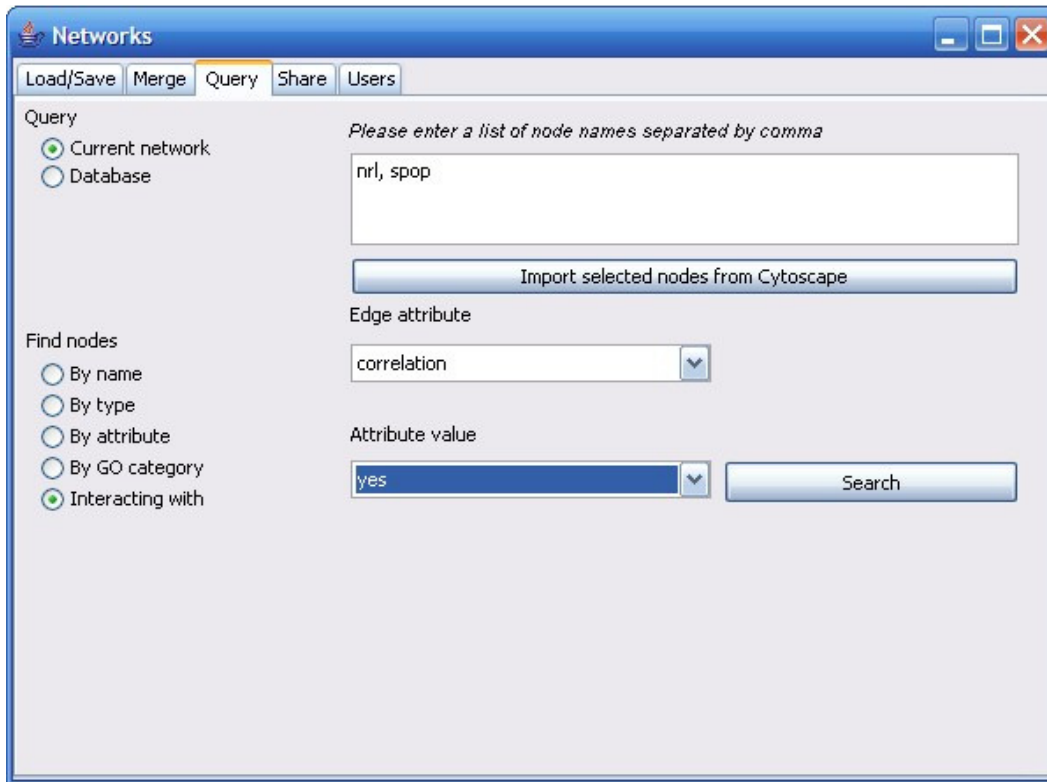
4.1 Sharing networks

To assign a "view/query", "update", or "delete" privilege on a network you created to a user of your choice, use the "Share" tab of the Networks window. In the list of networks you created, select a network you would like to share with another user. Choose a user from the list of all user ids. Check boxes next to "view", "update", and "delete" options if you would like the user to have the corresponding right on your network. "View" option will allow the user to load and query your network (and potentially save it under a different name), "update" will allow the user to make changes to and save your network under the same name, and "delete" option will allow the user to remove your network from the database. Click on "Grant" button to assign selected privileges.
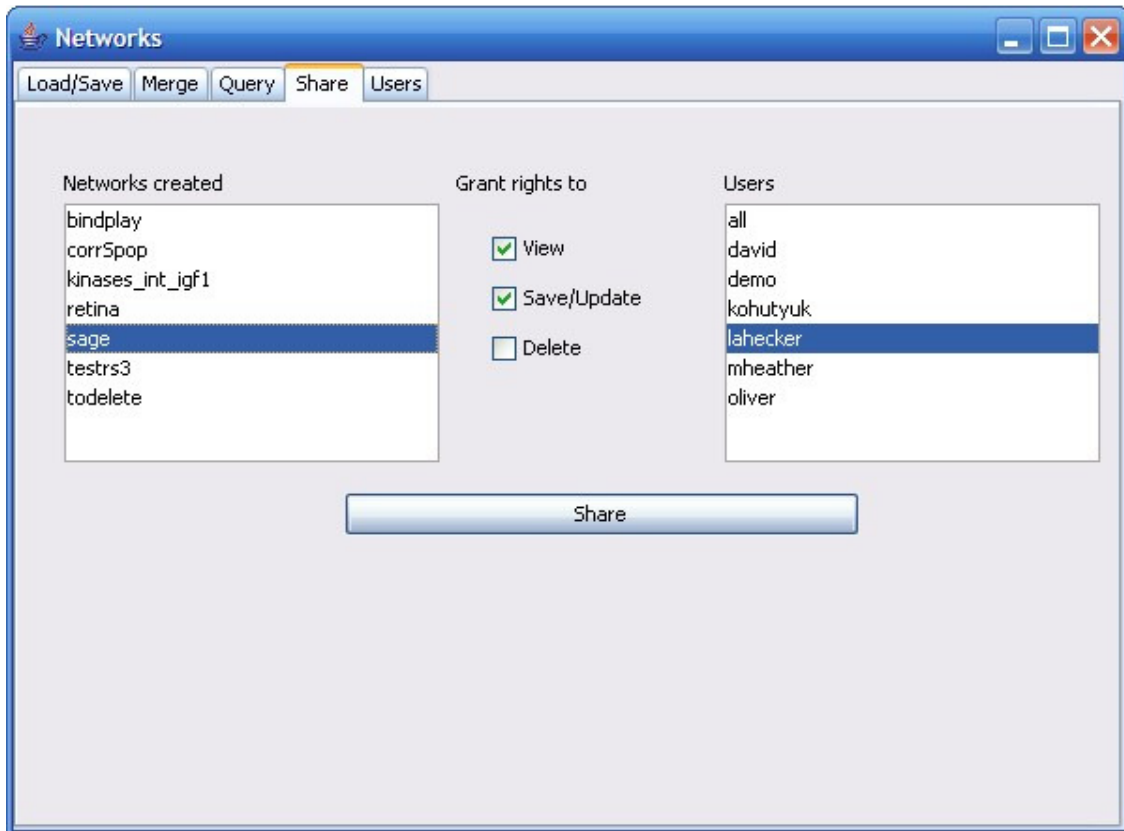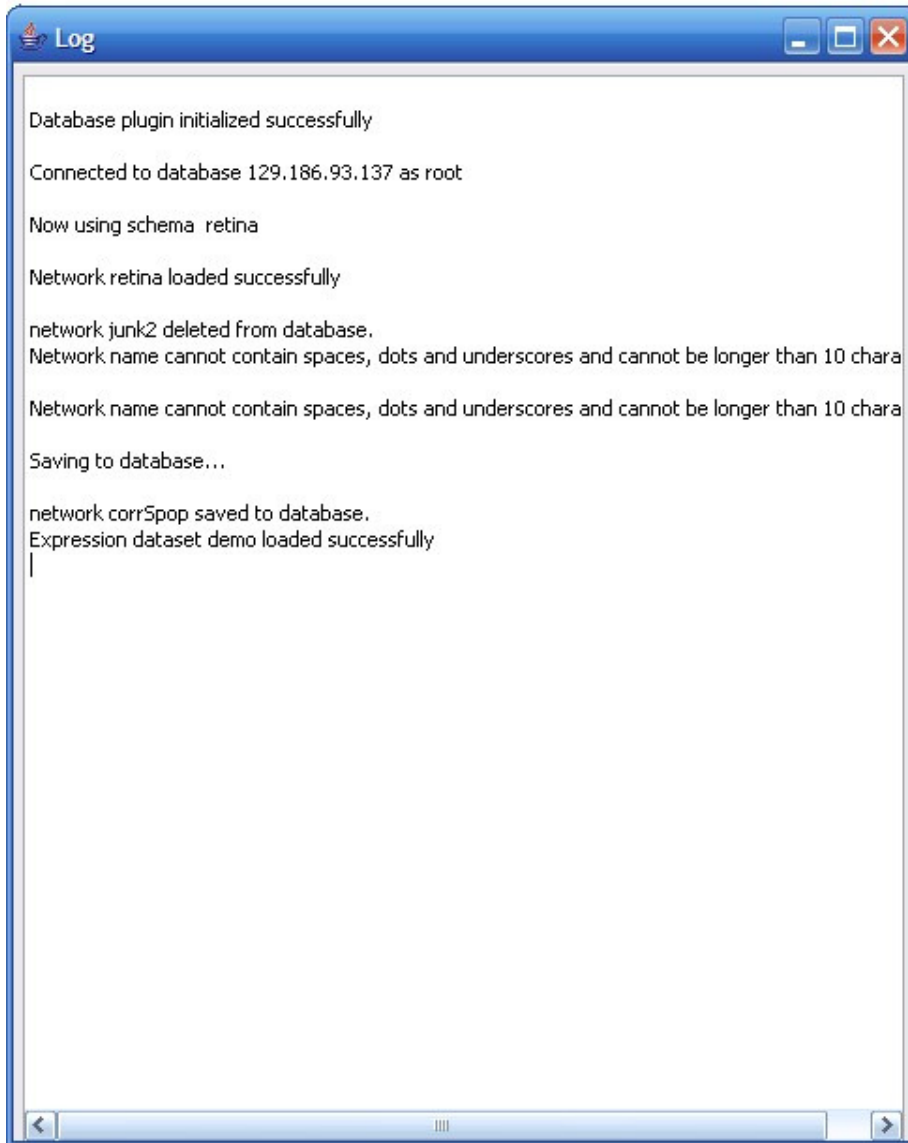
5. Account Window

The Account window allows you to easily update your user information, (first name, last name and email), reset your password, and delete your account.

To update your first name, last name, or email, type in the new information in the appropriate fields and click on the "Update" button in the Account window. To change your password, type in a new password and hit "Reset password" button. To delete your account, click on the Delete button. NOTE: removing your account is non-reversible! After your account is deleted you will not be able to log in using the same login name and your private datasets and networks will be removed from the database.

6. Log window

The Log window displays messages after each operation to inform you of the success of the operation. The Log window can be opened or closed at any time during the use of Retina Workbench. It is not necessary to keep the Log window open to record log messages: success/error messages are automatically recorded after each operation. The Log window is especially useful if an attempted procedure does not work: Log window will list possible reasons for failure of the procedure. For example, messages about invalid entries (such as an invalid username or password when connecting to a database or an illegal network name) will be displayed in the Log window. When Retina Workbench is closed, the information in the Log window is automatically saved to a text file called Log.txt.

Log

Database plugin initialized successfully

Connected to database 129.186.93.137 as root

Now using schema retina

Network retina loaded successfully

network junk2 deleted from database.
Network name cannot contain spaces, dots and underscores and cannot be longer than 10 chara

Network name cannot contain spaces, dots and underscores and cannot be longer than 10 chara

Saving to database...

network corrSpop saved to database.
Expression dataset demo loaded successfully

7. Help window

The Help window lists help topics on using Retina Workbench and provides instructions about various operations. Help window can be opened at any time to find necessary information. Once the Help window is opened, clicking on the topic of interest in the topics list on the left-hand side will display help information about the chosen topic.

8. Troubleshooting

Occasionally, you might run into following problems when using RetinaWorkbench:

1. Connection problems

    a. If you cannot connect to the database, there might be a problem with your database connection. Make sure you entered your username and password correctly. If you still cannot connect, close RetinaWorkbench and open it again from the "Plugins" menu in Cytoscape.

    b. Occasionally, a connection to the database may be lost while using RetinaWorkbench. If your results did not load properly, or if your requests do not get answered, close all windows in RetinaWorkbench and restart it again from the Plugins menu in Cytoscape.

2. Time problems

    Certain operations, such as saving a large network to database or querying a large network / expression dataset, take longer time than operations on small networks / datasets. Do expect that operations on large datasets may take a few minutes or longer. Speed is improved considerably if 1Gb or 2Gb of RAM is allocated for running java.

3. Results not loading

    If the results of your query do not load, yet RetinaWorkbench is finished processing your request (all buttons are enabled again), it is very likely that your request did not produce any results. If the result is a newly constructed network, it will only be loaded if it contains at least 2 nodes and 1 link. To see the status of your query, open the Log window, which monitors all operations.

4. Data format errors

    If you are loading an expression dataset from file, and the file is not correctly
    formatted (contains invalid characters or spaces, empty values), you will see
    an error message in the Log window and the dataset will not be loaded. Check
    the format of your dataset file and try again.

5. List of loaded expression sets / networks not updated properly

    If the list of loaded networks or datasets is not updated correctly, close the
    Expression data or Networks window you are working in and re-open it.
    Sometimes clicking on other tabs in the window will reload the network /
    dataset list as well.

6. Node names not showing in Cytoscape

    This is a Cytoscape feature not controlled by RetinaWorkbench. If node
    names are not displayed in the graphic window in Cytoscape for your newly
    loaded network, minimize or close the graphical display window in Cytoscape
    and re-open your network by clicking on the network name in the left panel.