DrugMatrix Tutorial

March 2011

NIEHS/NTP

111 TW Alexander Drive
Durham, NC 27709
http://ntp.niehs.nih.gov/
http://www.niehs.nih.gov/

Table of Contents

Getting Started

If this is your first time using DrugMatrix, it will be helpful to read the **DrugMatrix Overview** section of the **DrugMatrix Reference Guide** to familiarize yourself with parts of the DrugMatrix user interface and how information is organized. If you have questions regarding computations used for different types of data modeling in DrugMatrix, please refer to the **DrugMatrix Calculations** white paper. Both documents are available by clicking on **HELP** in the upper right hand corner of the User Interface.

Exercise 1: An Initial Query of DrugMatrix

As an introduction to DrugMatrix, the first exercise will show you how to use the very simple text search capability to extract information from the database and how to save information that you can use in other searches and with various other analysis tools.

You will learn to:
• Formulate simple queries and obtain background information about a compound.
• Populate Chemogenomic Domain Reports and use the information presented there to further investigate targeted drugs.
• Work with, manipulate and visualize lists of saved items.

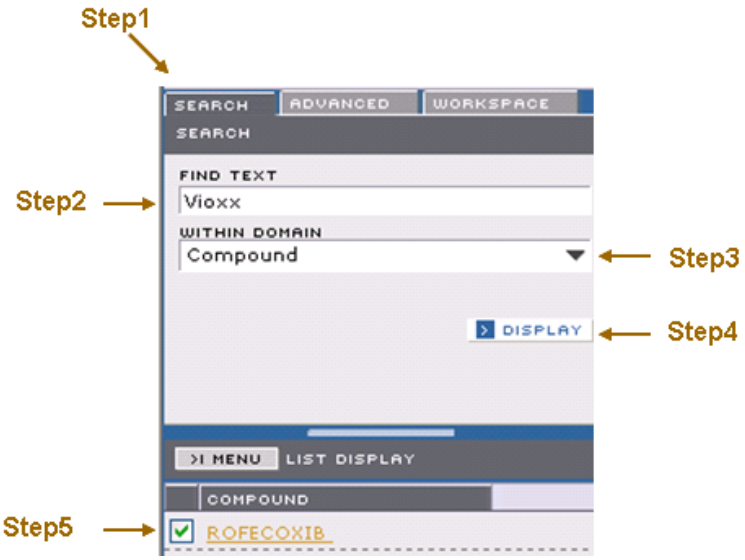
Exercise 1 ~ Part 1 ~ Simple Search

There have been various studies comparing the effectiveness of Vioxx® and Celebrex® over older NSAIDs for the treatment of arthritis. This section will demonstrate how to extract information from the database that will allow you to compare and contrast features of these compounds.

Begin with a simple search for compounds similar to Vioxx.
Step 1 Select the **SEARCH** tab (selected by default).
Step 2 In the **FIND TEXT** field, enter "**VIOXX**".
Step 3 Use the **WITHIN DOMAIN** drop down menu to select the **COMPOUND** domain for the query.
Step 4 Click the **DISPLAY** button.
Step 5 Click on **ROFECOXIB** to populate the appropriate Compound Domain Report with all the information available in DrugMatrix about this drug.

Notice in the List Display area that the compound **ROFECOXIB**, for which **VIOXX** is a brand name, has been retrieved by this query. When you search the **COMPOUND** domain using brand names or other synonyms, DrugMatrix will display the actual compound name at the top of the **COMPOUND** domain report.

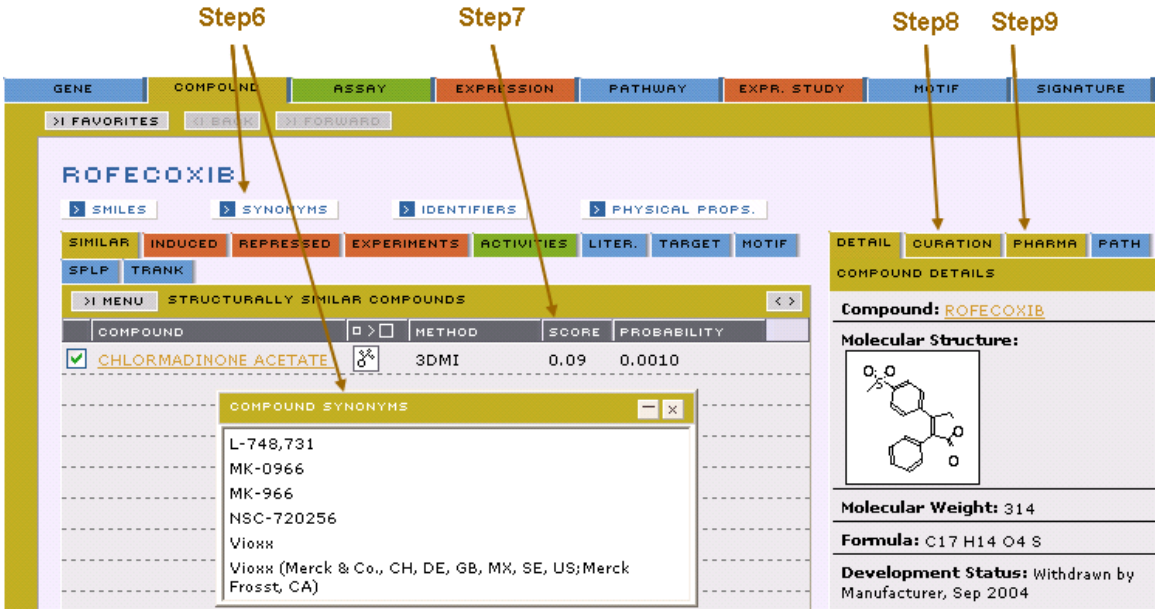
Exercise 1 ~ Part 1 ~ Steps 1 thru 5



This will display the Compound Domain report showing compounds similar in structure to **ROFECOXIB**.

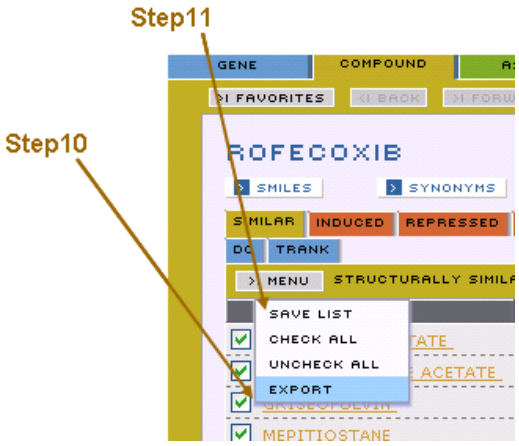
Step 6 Click on **SYNONYMS** to display a list of other names for **ROFECOXIB**. Notice the name **VIOXX** appears in this list. A search for any of these synonyms would have returned the same result.
Step 7 The **SCORE** tab indicates similarity between compounds.
Step 8 To find out what is known about **VIOXX** click on the **CURATION** tab. This panel will display extensive information such as the health conditions for which this drug is prescribed, toxicities of this drug, adverse effects and so forth.
Step 9 Click on **PHARMA** tab to view the pharmacological information. You can mouse over items on this panel for additional information. To read the original publication that describes the animal studies and pharmacology of this drug, click on one of the assays that has a "document" symbol on the left.

Exercise 1 ~ Part 1 ~ Steps 6 thru 9



Step 10 To export the **ROFECOXIB** similar compound list select **EXPORT** from the **MENU** at the **SIMILAR** panel.
This provides the option to open or save the file in excel format. Any list you generate is exportable to Excel. This enables you to work with lists outside of DrugMatrix, and to import them back into DrugMatrix.
Step 11 Select **SAVE LIST** from the menu on the **SIMILAR** panel to save the list of similar compounds in DrugMatrix workspace.

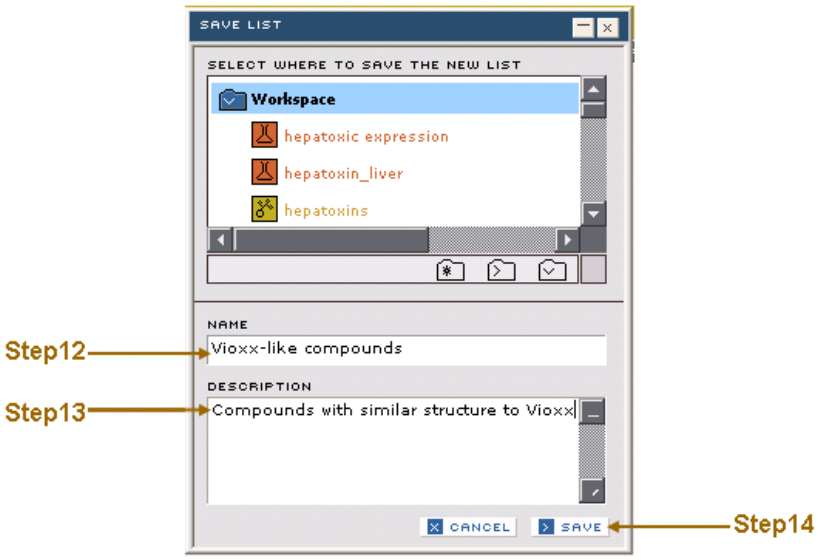
Exercise 1 ~ Part 1 ~ Steps 10 and 11



Step 12 Enter the name **Vioxx-like compounds** in the **NAME** field of the **SAVE LIST** dialog window.
Step 13 Give the name a brief description in the **DESCRIPTION** field, if you wish.

Step 14 Click the **SAVE** button to save the list.

Exercise 1 ~ Part 1 ~ Steps 12 thru 14

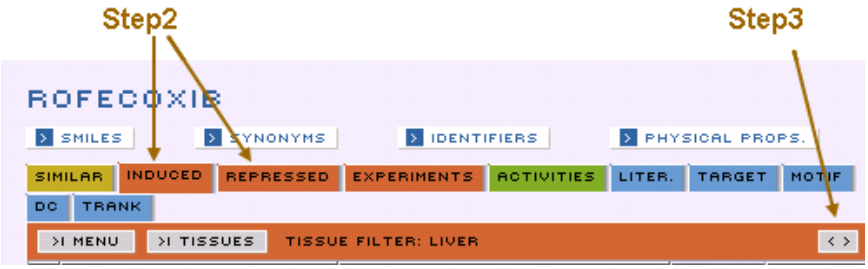


Exercise 1 ~ Part 2 ~ Working with List Management Tools

In order to gain familiarity with List Management tools, we will compare the gene expression profiles of the genes which are transcriptionally regulated by the administration of **VIOXX** and **CELECOXIB** in rat:

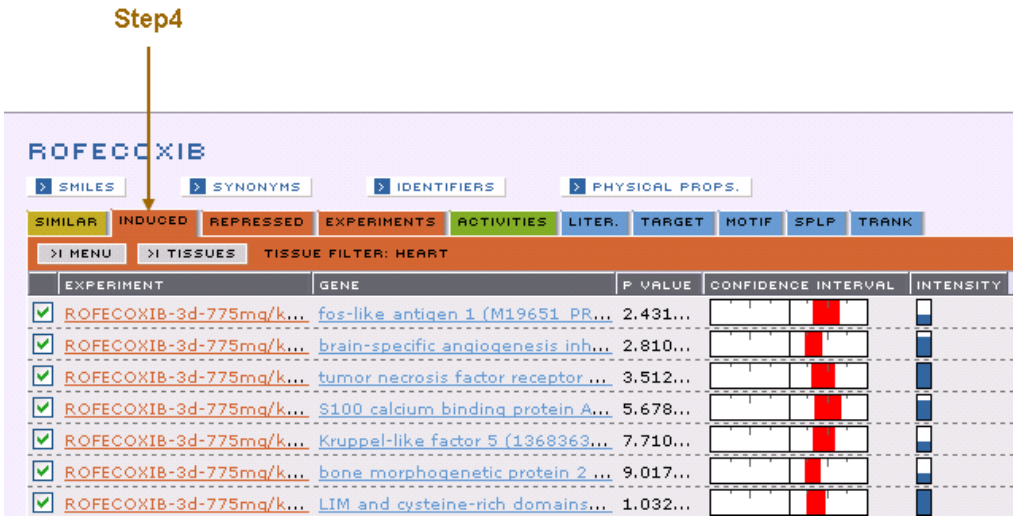
Step 1 If the compound domain panel is not currently populated with the list of **ROFECOXIB** compounds, follow **steps 1 thru 5 of Exercise 1 ~ Part 1** to generate a compound domain report on **ROFECOXIB.**
Step 2 Click on the **INDUCED** and **REPRESSED** tab separately to list the set of genes that are up-regulated and down-regulated respectively, by the administration of **VIOXX** and the experiments associated with the observation. Notice that **TISSUE FILTER: LIVER** is indicated next to the **TISSUES** button.
Step 3 Click on the **EXPAND** button to expand the window to show all columns.

Exercise 1 ~ Part 2 ~ Steps 2 and 3



Step 4 Now click on the **TISSUES** button and choose **HEART**. Notice that the **TISSUE FILTER** description changes to **HEART**.
With the **INDUCED** panel expanded you should see something like this.

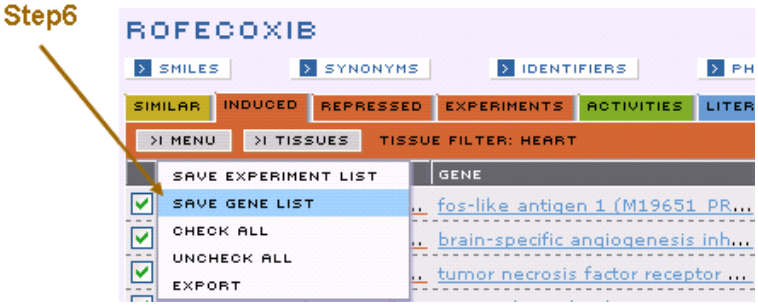
Exercise 1 ~ Part 2 ~ Step 4



Step 5 Move your mouse over individual experiments for a description of the *in vivo* studies.
Step 6 To save the list of genes induced by **ROFECOXIB**, select **SAVE GENE LIST** from the **MENU** on **DOMAIN REPORT** sub panels.

NOTE: When you save a list, only the checked items will be saved. This menu provides the option to check or uncheck all items in the list.

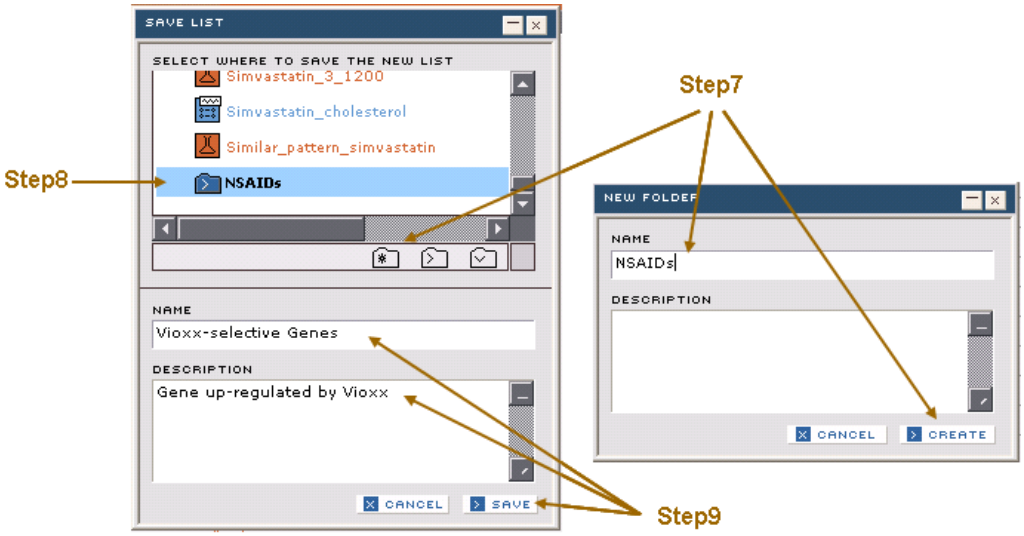
Exercise 1 ~ Part 2 ~ Step 6



Step 7 In the **SAVE LIST** dialog, click on the **NEW FOLDER** icon and **CREATE** a new folder in your **WORKSPACE** called **NSAIDs**.
Step 8 In the **SAVE LIST** dialog, select the **NSAIDs** folder you have just created.

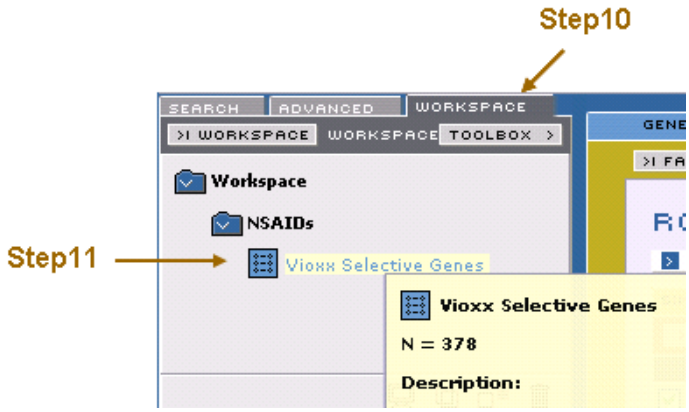
Step 9 Save your new list named as **Vioxx Selective Genes**. Provide an optional description if you like, e.g. **genes up-regulated by Vioxx**.

Exercise 1 ~ Part 2 ~ Steps 7 thru 9



Step 10 Click on the **WORKSPACE** tab in the **SEARCH-WORKSPACE** area on the left.
Step 11 Move your mouse over the saved list to see how many genes were induced by Vioxx.

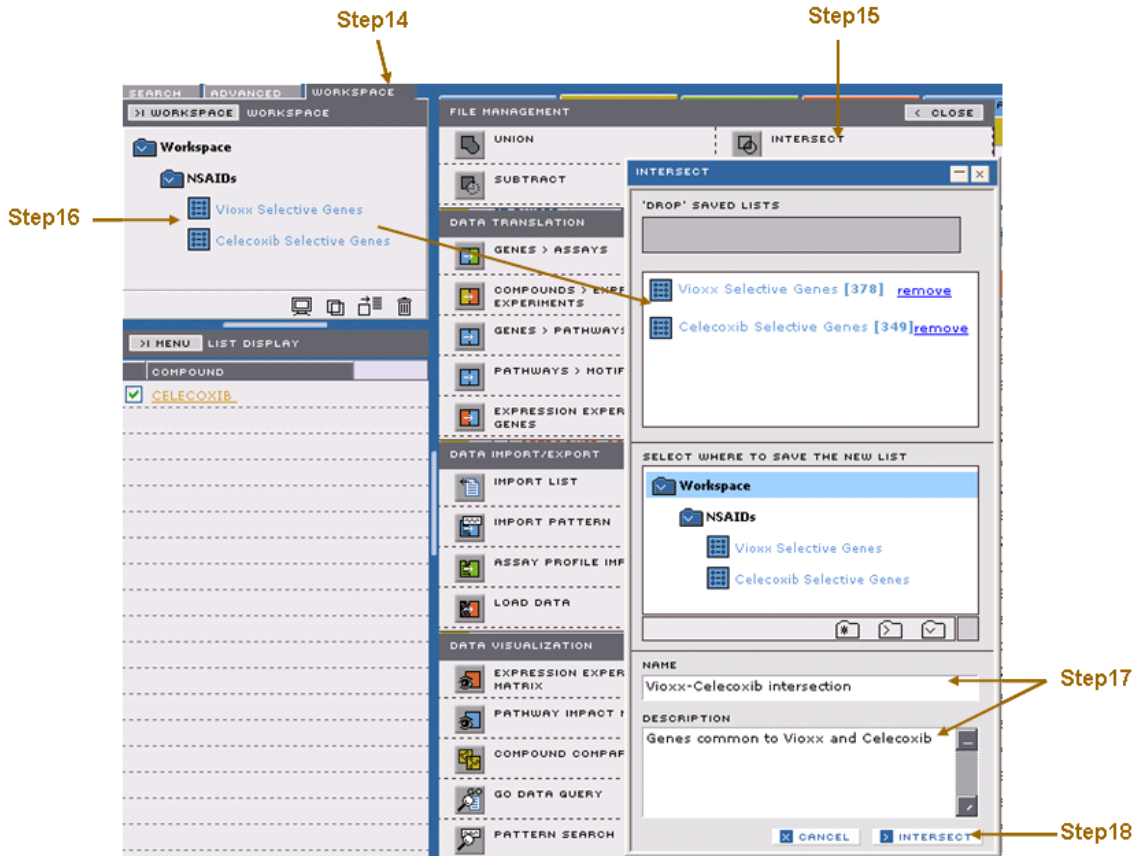
Exercise 1 ~ Part 2 ~ Steps 10 and 11



Step 12 Next do a **COMPOUND** search on **CELECOXIB** (as in **steps 1 thru 5 of Exercise 1 ~ Part 1**).
Step 13 Save the **CELECOXIB** induced genes in heart experiments with the name **Celebrex-selective genes** (as in **steps 6, 8 and 9 of Exercise 1 ~ Part 2)**.
Step 14 Click on the **TOOLBOX in the WORKSPACE**.
Step 15 In **FILE MANAGEMENT** area of the Toolbox, select the **INTERSECT** tool.

Step 16 Drag the previously saved **Vioxx-selective genes** and **Celecoxib-selective genes** lists into the **'DROP' SAVED LISTS** field of the **INTERSECT** tool.
Step 17 Name this new list for example, **Vioxx-Celecoxib Intersection** and give it a description, for example **Genes common to Vioxx and Celecoxib.**
Step 18 Click the **INTERSECT** button at the bottom of the dialog box to save the list in your general **WORKSPACE**. If you have created the folder called **NSAIDs** you can save the list within that folder.

Exercise 1 ~ Part 2 ~ Steps 14 thru 18



Although inspection of the genes in the intersect shows no obvious linkage, the number of overlapping genes is statistically significant, confirming that these two drugs are perturbing the expression patterns in the heart samples in a similar manner.

Other Possibilities ~

If you are interested in knowing the genes that are up-regulated only by Vioxx, but not Celecoxib, click on the **SUBTRACT** tool from the **TOOLBOX** and drag the **VIOXX** file onto the **SUBTRACT FROM** tool and drag the **CELECOXIB** file into the **LIST/FAVORITE** field.

Conclusions to Exercise 1

This exercise has introduced you to some of the basic features and functions of DrugMatrix query tools. You have learned some techniques that will help you begin to extract information from DrugMatrix that is specific to your particular research.

Exercise 2: Advanced Search Queries

In the previous exercise, you learned to extract information from the database using the simple **SEARCH** tool and you learned how to save lists of information you extracted.
DrugMatrix also offers more advanced search capabilities via the **ADVANCED** tab of the **SEARCH WORKSPACE** area. With this tool, a number of filters can be combined to narrow searches for relevant data. The DrugMatrix database has been structured so that these queries can combine information across domains, allowing very powerful searches. Combined with the use of result lists and list manipulation tools, highly complex searches can be carried out. As you become more familiar with these tools you can organize information extracted from the data base in a way that best suits your specific research goals.

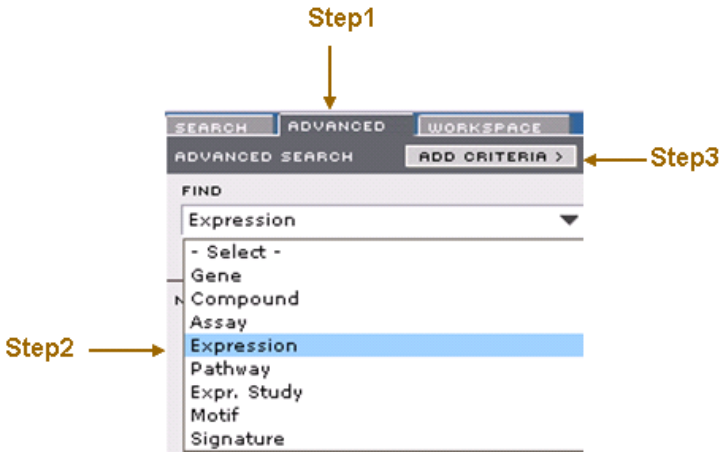
In this exercise you will learn to:
• Use the **ADVANCED** search tools to narrow search results with a combination of predefined terms and using Boolean operators.
• Explore information presented in various Domain Reports.
• Generate a number of specific database queries based on examples provided.

Exercise 2 ~ Part 1 ~ Build an Expression Experiment Advanced Query

In this exercise you will use the **ADVANCED** search tools to identify Vioxx-like expression profiles.

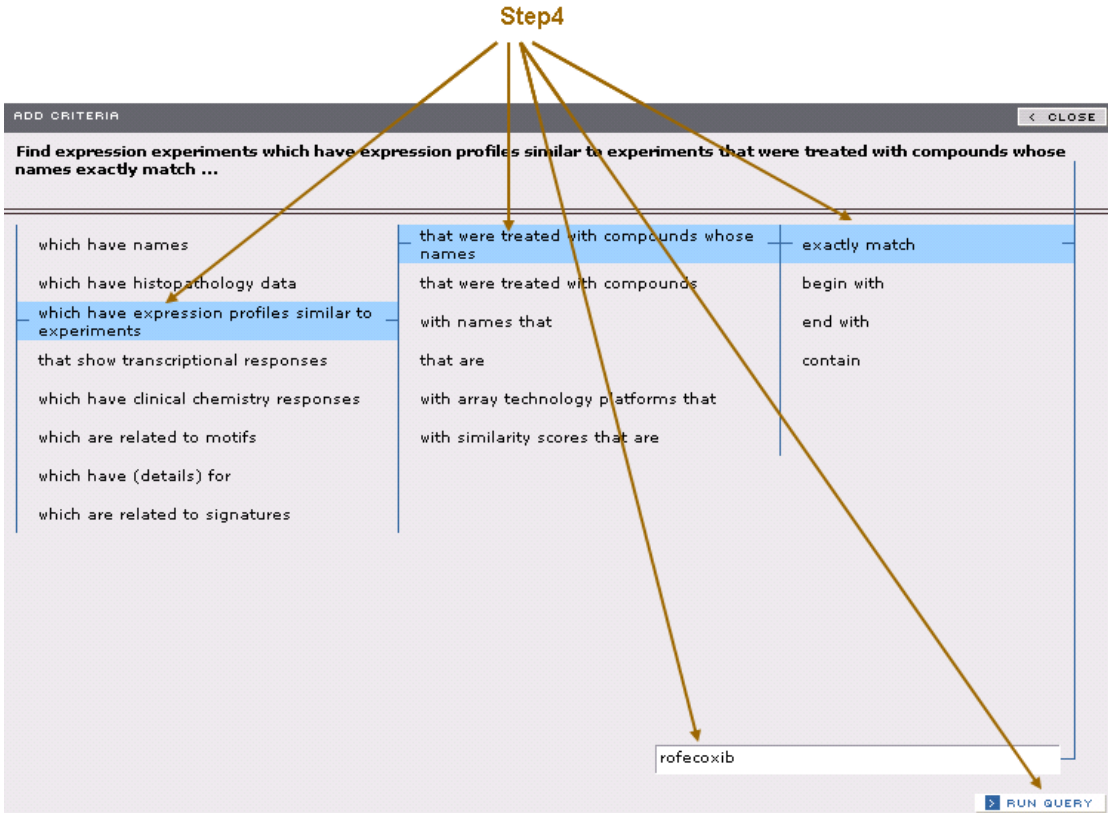
In order to identify Vioxx-like expression profiles:
Step 1 Click on the **ADVANCED** search tab.
Step 2 Select **EXPRESSION** from the pull-down menu.
Step 3 Click on **ADD CRITERIA**.

Exercise 2 ~ Part 1 ~ Steps 1 thru 3



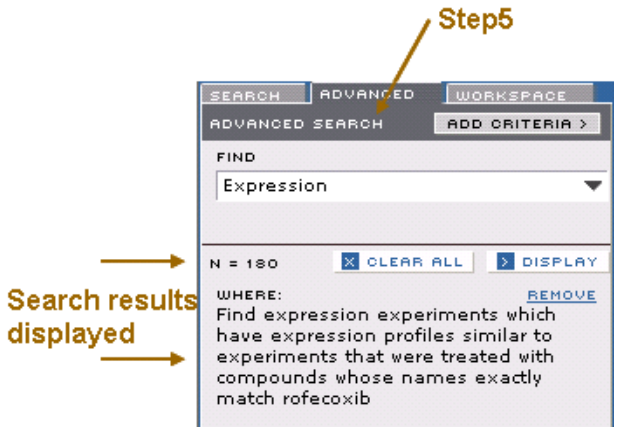
Step 4 In the **ADD CRITERIA** panel click on the criteria indicated in the color cyan below. Then type “rofecoxib” into the text box and press the **RUN QUERY** button.

Exercise 2 ~ Part 1 ~ Step 4



Step 5 When the search is complete, the **ADVANCED SEARCH** panel shows the criteria by which you have searched. It also indicates the resulting number of records.

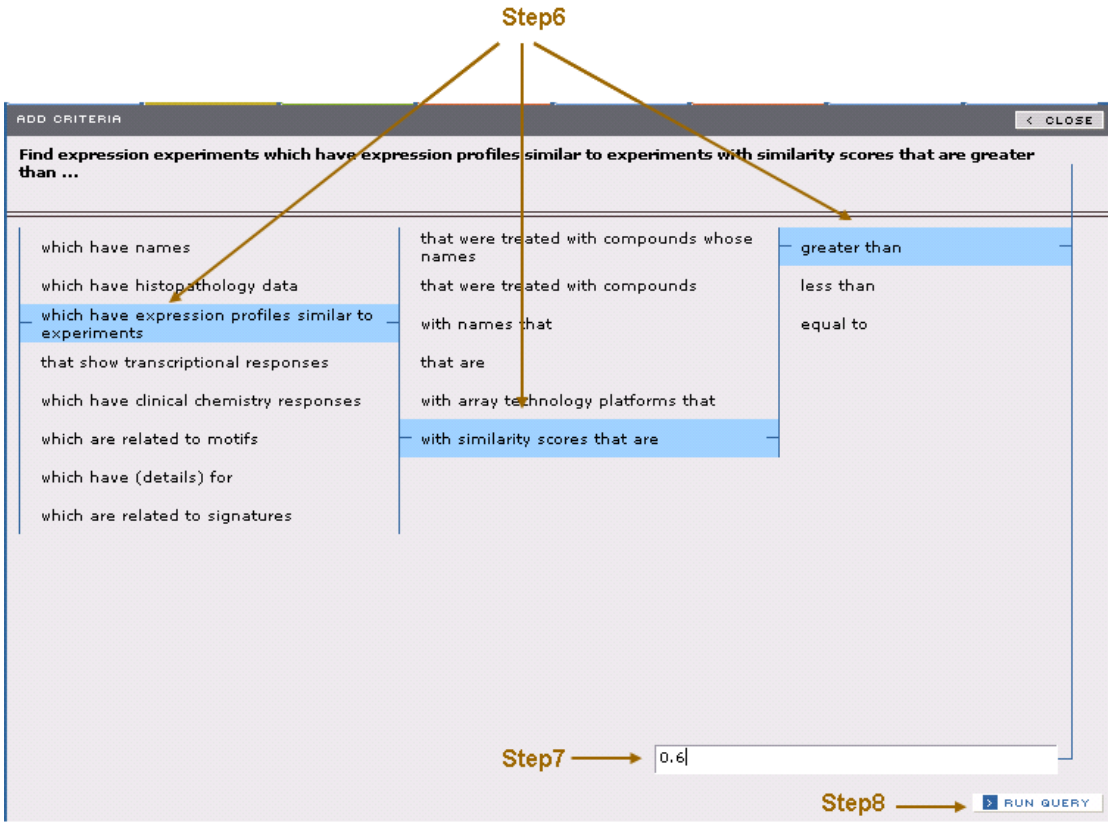
Exercise 2 ~ Part 1 ~ Step 5



To limit the search to experiments that are very similar you can require that the Pearson's similarity score be above a specific threshold.

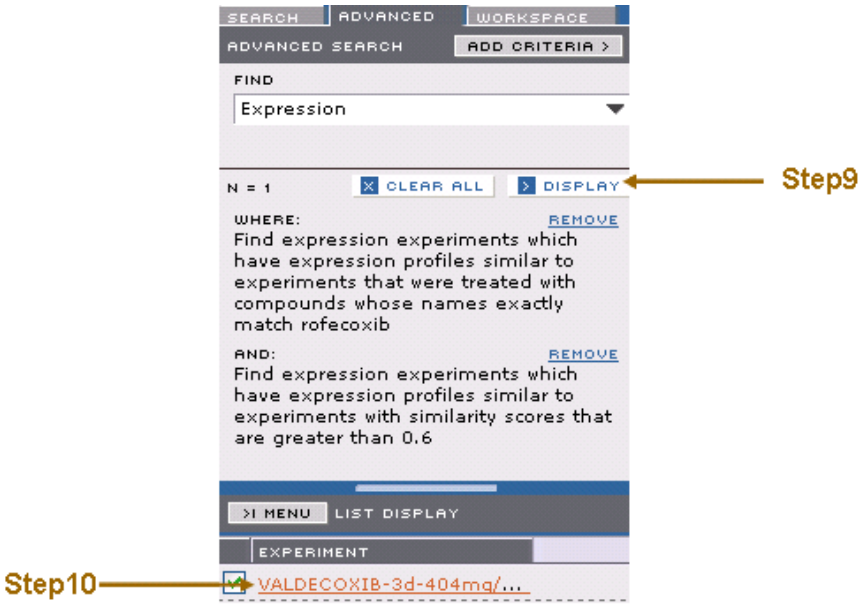
Step 6 Click **ADD CRITERIA**, add the criteria indicated in cyan below.
Note: As you select criteria in each column of an **ADD CRITERIA** panel, the next column of related choices will be displayed. The choice you make in each column, determines the selections that will be presented in the following column.
Step 7 Type **"0.6"** into the entry field.
Step 8 Click **RUN QUERY**.

Exercise 2 ~ Part 1 ~ Steps 6 thru 8



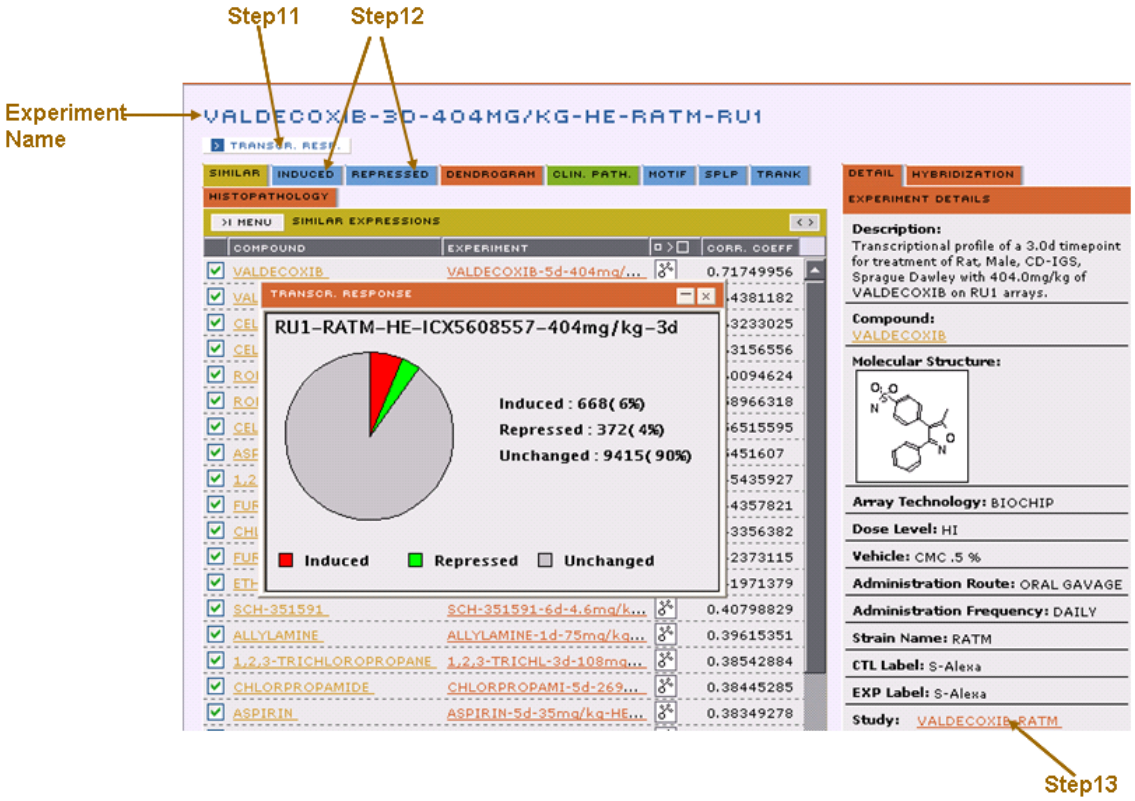
When the search is complete, the **ADVANCED SEARCH** panel shows all of the criteria by which you have searched. It also indicates the number of expression experiments you have identified using these criteria.
Step 9 Click the **DISPLAY** button to display this list of experiments below the panel.
Step 10 Click on the experiment **VALDECOXIB-3d-404mg/kg-HE-RATM-RU1**.

Exercise 2 ~ Part 1 ~ Steps 9 and 10



This will populate an **EXPRESSION** domain report that you can use to observe how strongly **VALDECOXIB** perturbed gene expression in this experiment. You can look at the percentages of genes up regulated and down regulated, and you can view which genes were perturbed.
Notice that the name of the experiment you selected appears top of this report. On the **SIMILAR** panel, experiments are listed by similarity of their expression profiles (as shown in the **CORR. COEFF** column) to the **VALDECOXIB** experiment you selected.
Step 11 Click on the **TRANSCR. RESP.** button to display a graph showing the number and percent of genes that are up-regulated, down-regulated and unchanged in this experiment.
Step 12 Click on the **INDUCED** and **REPRESSED** tabs to see detailed information about genes that were up and down regulated in this experiment.
To look at the whole study, of which the above individual experiment is a component:
Step 13 Click on the link **VALDECOXIB-RATM** in the **DETAIL** panel.

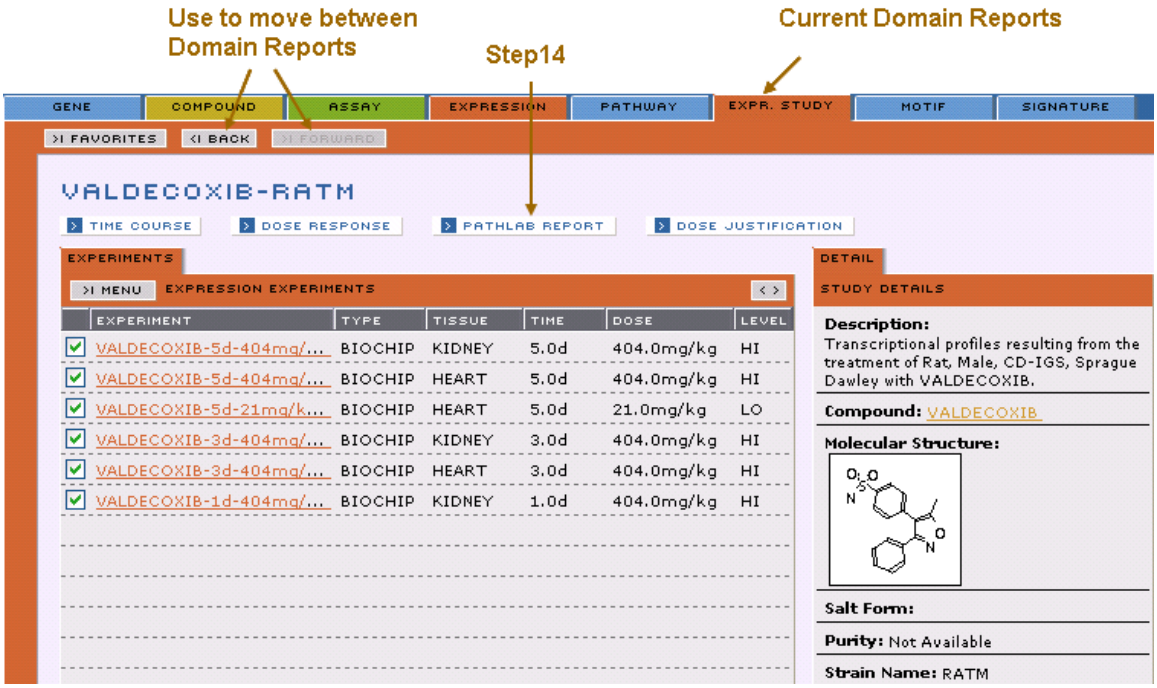
Exercise 2 ~ Part 1 ~ Steps 11 thru 13



This will now populate an **EXPR. STUDY** Domain Report. An Expression Study is a set of experiments done during a narrow window of time using various doses of a particular drug at various time intervals. This report gives us access to information about how the study was derived, the other experiments in the study, histopathology and clinical pathology information, and a variety of other aggregated information to help provide context for this experiment.
Step 14 Click on the **PATHLAB REPORT** button to view histopathology and clinical pathology reports. This will display a PDF report in a separate browser window.

Note: You can use the **BACK** and **FORWARD** buttons to move between Domain Reports you have just viewed. You can also navigate using the Domain Report tabs.

Exercise 2 ~ Part 1 ~ Step 14



Exercise 2 ~ Part 2 ~ Examples of Various Queries

This section provides several examples of advanced queries. You can generate advanced queries for any of the six different chemogenomic domains: Genes, Compounds, Assays, Expression, Pathways and Expression studies. This section provides several examples of advanced queries.

Exercise 2 ~ Part 2 ~ Search for Expression Experiments ~ Example 1

Suppose we want to find experiments that impact cell division, such as treatment with certain classes of anti-cancer agents. Since white blood cells rapidly divide, a count of these cells in blood is used as an indicator of compound activity. DrugMatrix has data for white cell counts, so we can query for experiments where the animal's blood had a reduced white blood cell count.

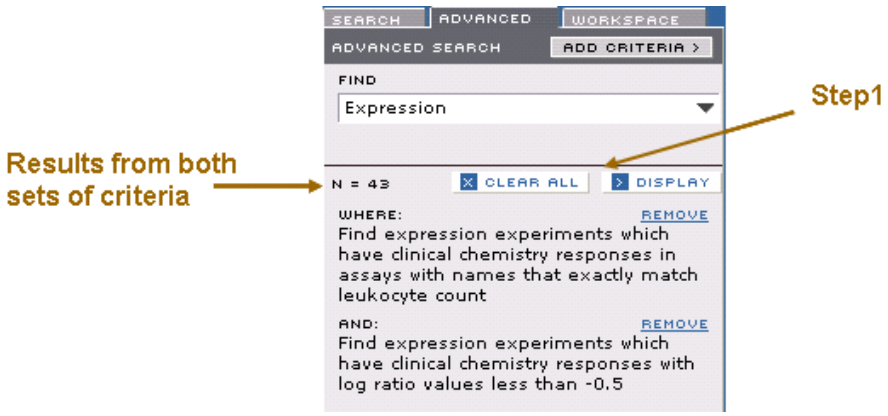
The following query results were obtained by running an initial query which retrieved a large number of experiments. The query was then refined by using a second set of criteria.

Step 1 Click the **CLEAR ALL** button at the bottom of the **ADVANCED** panel to remove all previous search criteria.

Step 2 Click **ADVANCED** tab and select **EXPRESSION**.

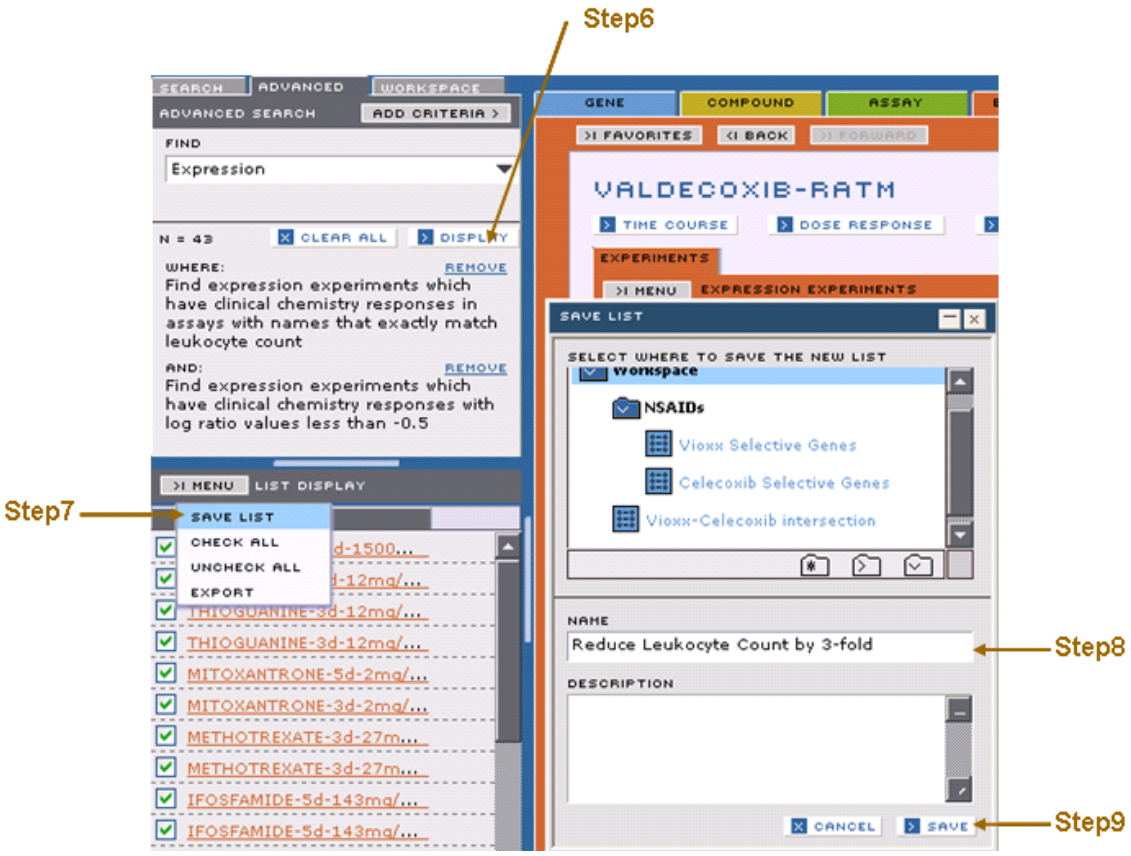
Step 3 Click **ADD CRITERIA** and select
-which have clinical chemistry responses
-in assays with names that
-exactly match
-[leukocyte count]
Step 4 Click the **RUN QUERY** button at the bottom of the **ADD CRITERIA** panel
Step 5 Click on **ADD CRITERIA** (this is equivalent to the Boolean operator "**AND**") and select
-which have clinical chemistry responses
-with log ratio values
-less than
[-0.5]
We are looking for an approximate 3x reduction or $\log_{10}(1/3) \sim -0.5$

Search results from Exercise 2 ~ Part 2 ~ Example 1 ~ Steps 1 thru 5



Step 6 Click on the **DISPLAY** button.
Notice that most of the experiments populated are for anti-cancer drugs.
Step 7 Select **SAVE LIST** from the **MENU** button on the **LIST DISPLAY** panel.
Step 8 Name the list in the **SAVE LIST** dialog window, for example as **Reduce Leukocyte Count by 3-fold**.
Step 9 Click the **SAVE** button at the bottom of the **SAVE LIST** dialog window.

Exercise 2 ~ Part 2 ~ Example 1 ~ Steps 6 thru 9



We can use the DrugMatrix **EXPRESSION EXPERIMENTS > COMPOUNDS** tool to get a discrete list of compounds from this set of experiments.

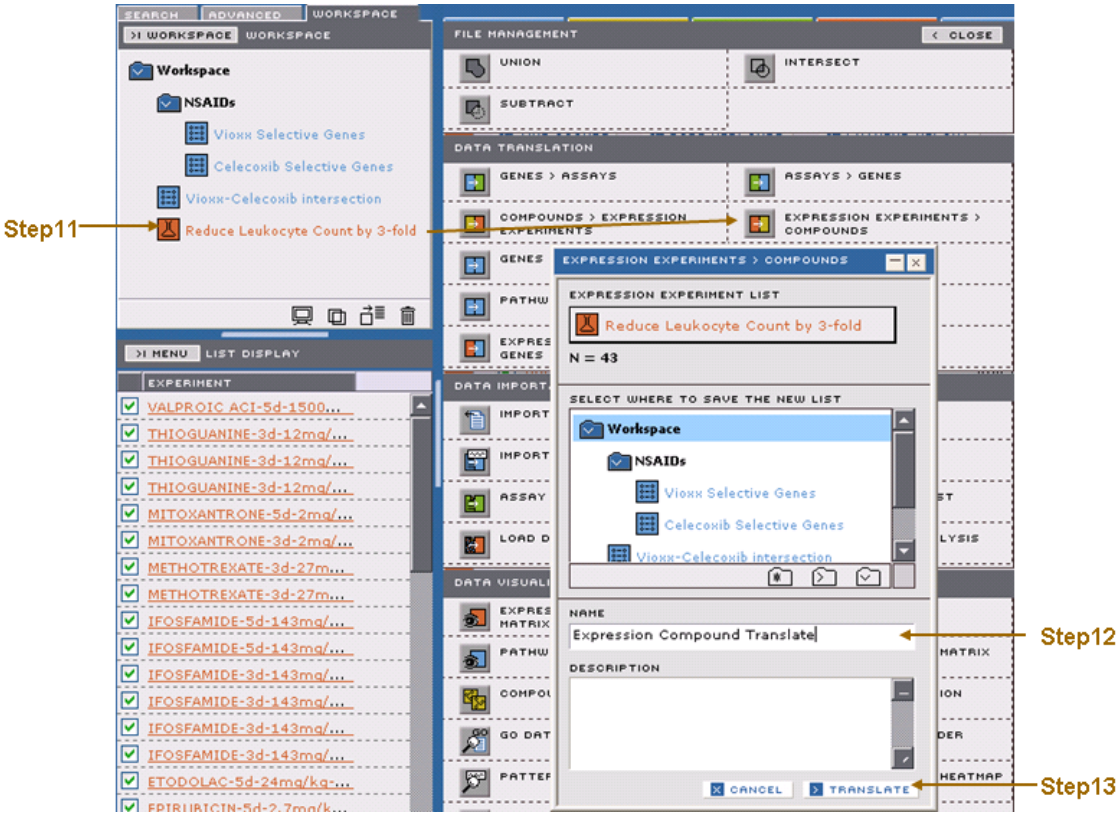
Step 10 Select the **WORKSPACE** tab and open the **TOOLBOX**.

Step 11 Drag this saved list into the **EXPRESSION EXPERIMENTS > COMPOUNDS** tool.

Step 12 Give the new list a name, for example **Expression Compound Translation**.

Step 13 Click the **TRANSLATE** button at the bottom of the panel.

Exercise 2 ~ Part 2 ~ Example 1 ~ Steps 11 thru 13



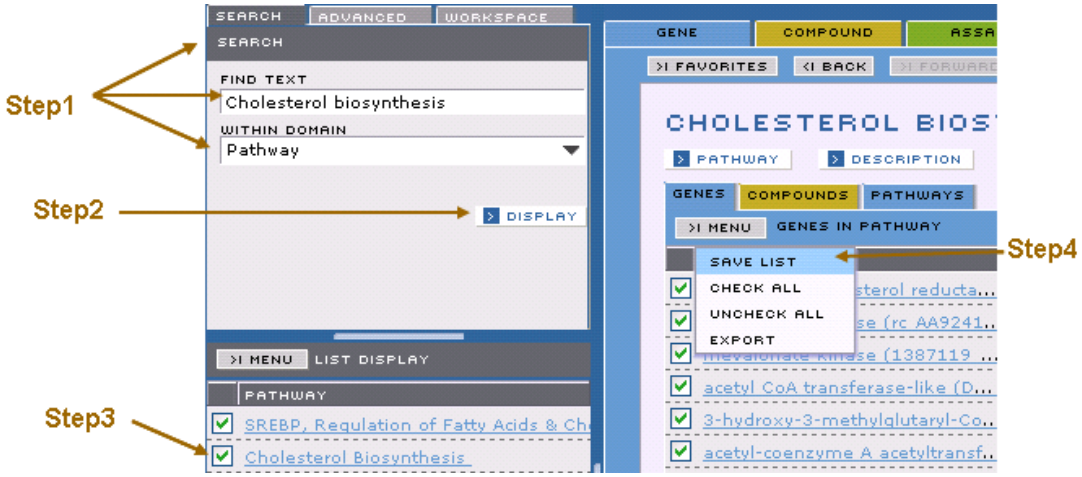
Step 14 Now in the **WORKSPACE** panel, mouse over this list to see how many compounds are represented by the selected experiments.

Exercise 2 ~ Part 2 ~ Search for Expression Experiments ~ Example 2

A popular class of drugs on the market is used to reduce circulating cholesterol, one form of which is associated with increased risk of heart toxicity. These drugs act by directly inhibiting enzymes in the cholesterol biosynthesis pathway.
As an example, you might want to investigate whether other compounds have similar effects on this pathway. To do this, we will first select the genes involved in this pathway and then identify other compounds that perturb these genes similarly to compounds of this class.
Step 1 Select the **SEARCH** tab and do a basic text query on **Cholesterol Biosynthesis** within the **PATHWAY** domain.
Step 2 Click on the **Cholesterol Biosynthesis** link to populate the **PATHWAY** Domain Report.

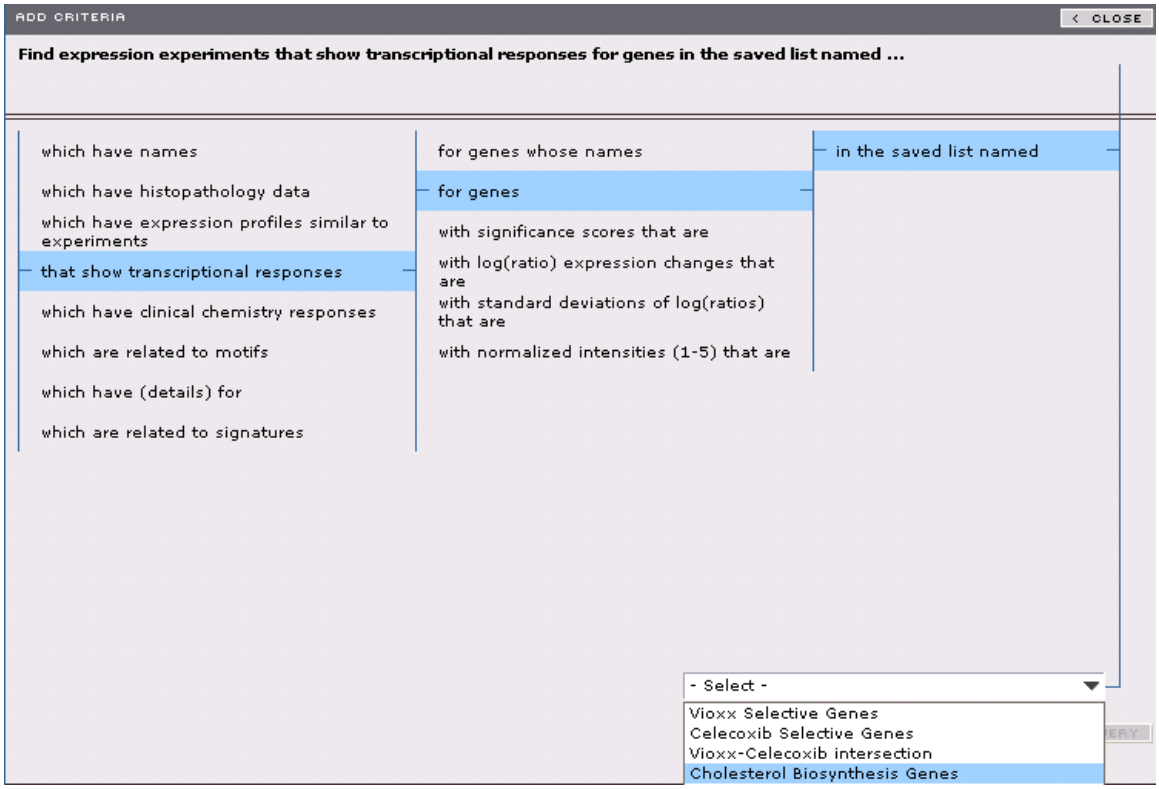
Step 3 Click on the **MENU** button on the **PATHWAY** Domain Report panel.
Step 4 Save the list of genes involved in this pathway as **Cholesterol biosynthesis genes**.

Exercise 2 ~ Part 2 ~ Example 2 ~ Steps 1 thru 4

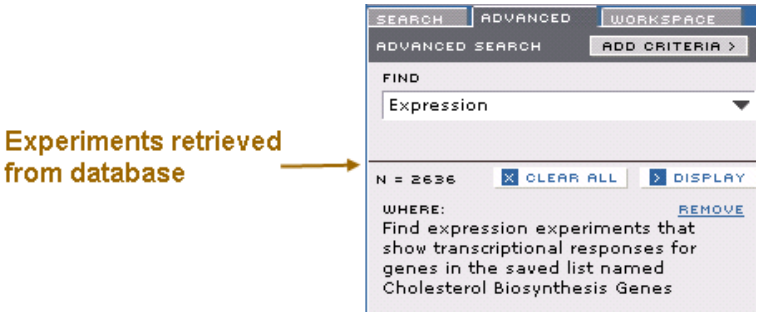


Now search for expression experiments in which cholesterol biosynthesis genes were perturbed.
Step 5 Click on the **ADVANCED TAB** and select **EXPRESSION** from the drop-down menu.
Step 6 Generate the advanced query shown below.

Exercise 2 ~ Part 2 ~ Example 2 ~ Step 6



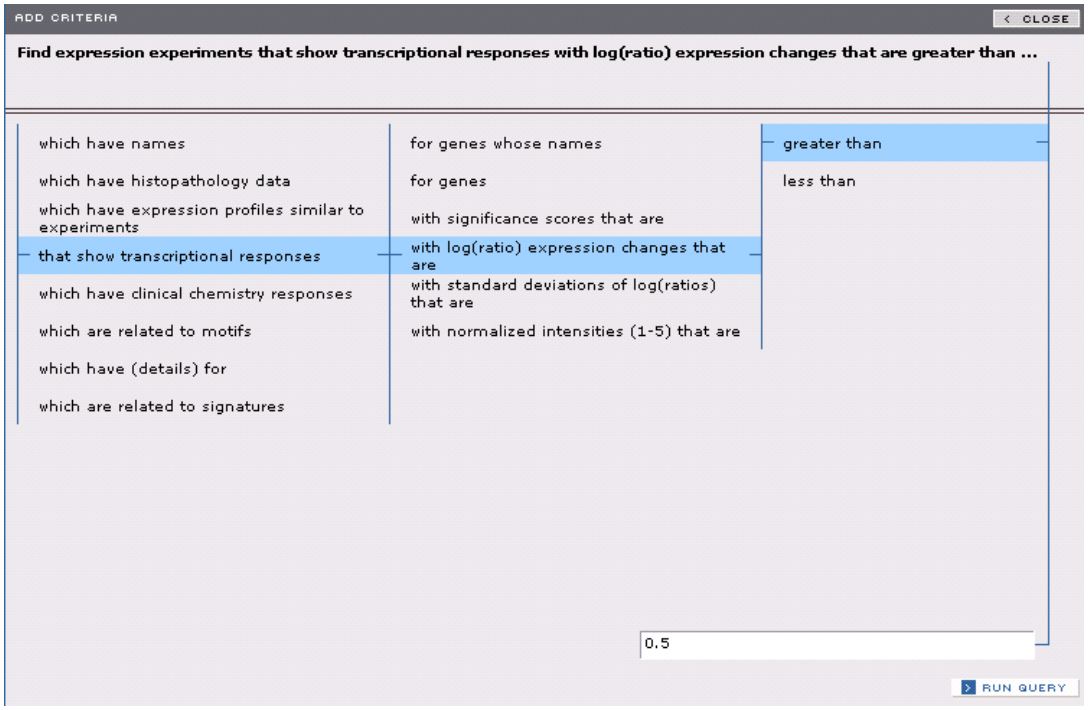
Notice the large number of expression experiments retrieved from the data base



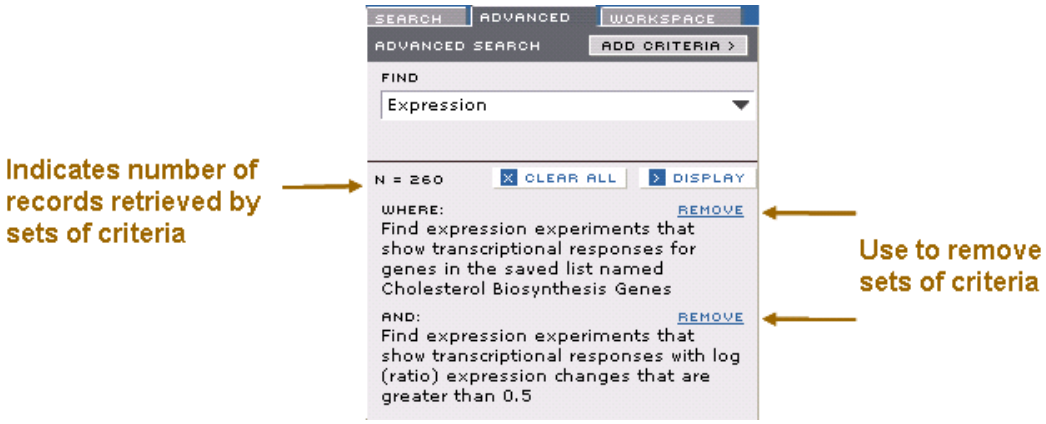
One may assume that if cholesterol was reduced, these genes would be repressed. But, in fact, the organism detects the drop in cholesterol and increases transcription of these genes to compensate. Therefore, we want to search for expression experiments showing strong induction of these pathway genes. Alternatively, we could select experiments for cholesterol-lowering drugs to determine how these genes are generally perturbed and then use this information to filter for experiments.

Step 7 In order to search for experiments showing strong induction of cholesterol biosynthesis genes, continue to refine the search by adding the criteria indicated below.

Exercise 2 ~ Part 2 ~ Example 2 ~ Step 7



Notice that the addition of this set of criteria has narrowed the results considerably. You can continue refining your searches by adding as many sets of criteria as you like. You can also remove individual sets of criteria with the **REMOVE** button on the **ADVANCED** panel.

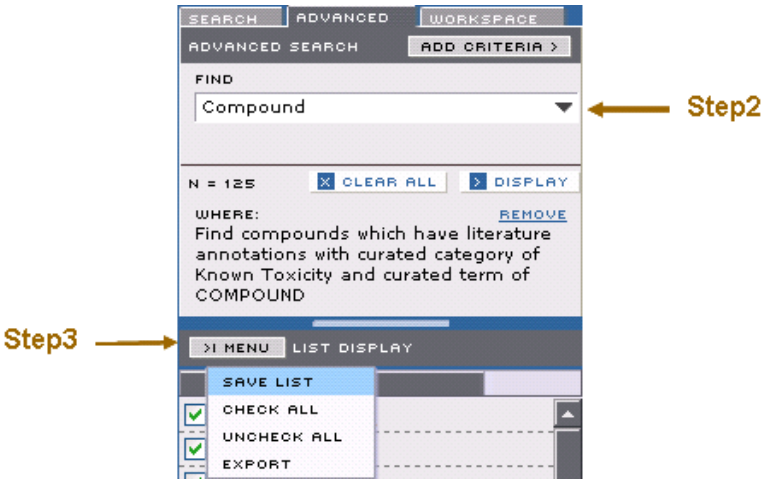


Exercise 2 ~ Part 3 ~ Search for Compounds ~ Example 1

In this part of the exercise, we will search for compounds based on either experimental or literature annotation.
First, suppose we want to find compounds that are annotated as hepatotoxicants
Step 1 Select the **ADVANCED** tab, click **ADD CRITERIA** and select

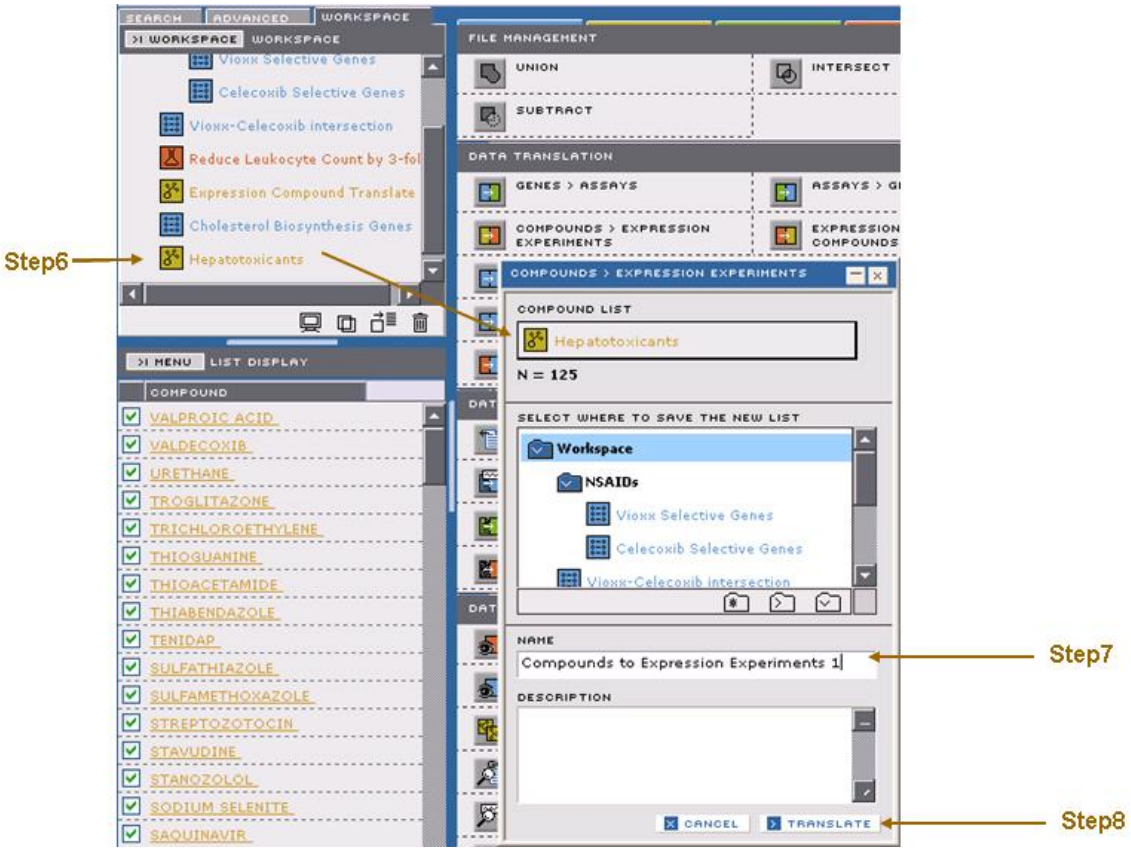
-which have literature annotations with curated category of
-Known Toxicity
-and curated term of
-[Hepatotoxicity]
Step 2 **DISPLAY** the list.
Step 3 Select **SAVE LIST** from the menu on the **LIST DISPLAY** panel.

Exercise 2 ~ Part 3 ~ Example 1 ~ Steps 2 and 3



Step 4 In the **SAVE LIST** dialog box, name the list as **Hepatotoxicants**.
Step 5 Click the **SAVE** button at the bottom of the dialog box to save the list.
If we now want to ask what genes might be responsible, a first step would be to identify expression experiments with these compounds that are in the database. In order to do this, we will need to use the tool that will convert these compounds to expression experiments.
Step 6 Open the **TOOLBOX** and Drag the **Hepatotoxicants** list onto the **COMPOUNDS > EXPRESSION EXPERIMENTS** tool.
This will open the dialog window for that tool with **Hepatotoxicants** in the **COMPOUND LIST** field.
Step 7 In dialog box that appears, name the list to be translated, for example **Compounds to Expression Experiments 1**.
Step 8 Click the **TRANSLATE** button at the bottom of the panel.

Exercise 2 ~ Part 3 ~ Example 1 ~ Steps 6 thru 8



Step 9 Mouse over the list name in the **WORKSPACE** panel to see how many experiments these compounds were used in.
If you inspect the experiment list, you will notice that all tissues are represented in the experiment list. How might you refine this list to filter the experiments only for liver tissues?

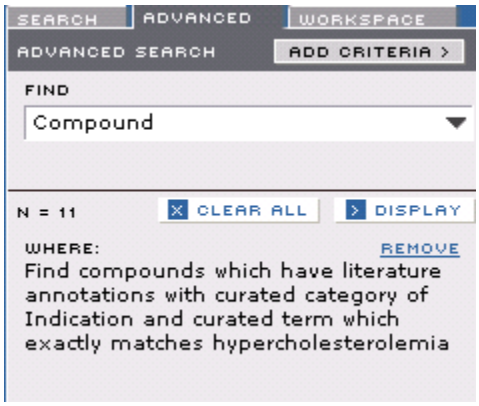
HINT: Use the Advanced Query Tool:
Find experiments which have names found in the saved list named “Compounds to Expression Experiments 1”
And
Find experiments which have (details) for experiments run on tissues that exactly match “liver”

Exercise 2 ~ Part 3 ~ Search for Compounds ~ Example 2

Here we want to query the data base for compounds that are used to treat hypercholesterolemia.

Step 1 Conduct the search indicated in the panel below.

Search results ~ Exercise 2 ~ Part 3 ~ Example 2

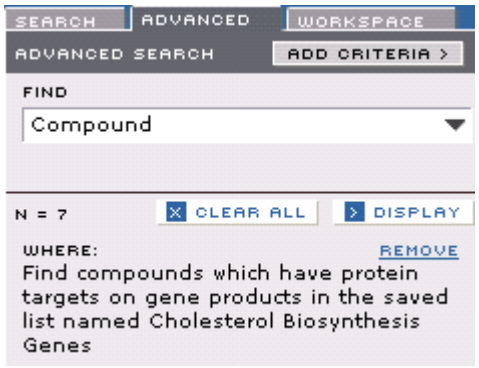


Exercise 2 ~ Part 3 ~ Search for Compounds ~ Example 3

We can even query the data base for compounds that are known to hit targets in certain pathways, a useful reference when looking for expression profiles to compare compounds which are suspected to act on the same pathway.
First we will extract list of genes in the Cholesterol Biosynthesis pathway as we have done in **Exercise 2 Part 2 Example 2**. Alternatively you can take the list saved as **Cholesterol biosynthesis genes** in **WORKSPACE** tab.

Now we will query the data base for compounds that have known targets that are found in the Cholesterol biosynthesis pathway.
Step 1 Use the **ADVANCED** search panel to conduct the search indicated by the results in the panel below.

Search results Exercise 2 ~ Part 3 ~ Example 3



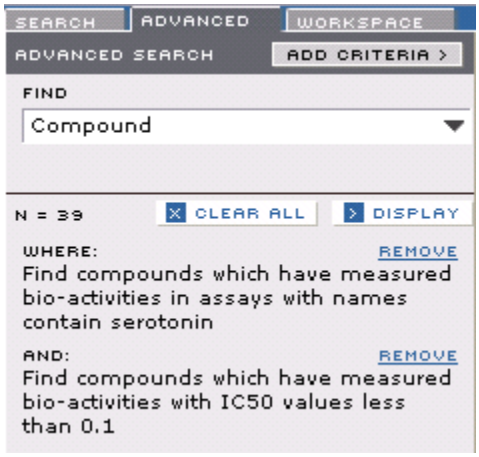
Consider how you might use a different query to find compounds that perturb genes in this pathway.

Exercise 2 ~ Part 3 ~ Search for Compounds ~ Example 4

Besides expression data there is a wealth of in vitro assay data available in the DrugMatrix database. For example, we can search for compounds that bind to serotonin receptor assays with low IC50 concentrations, such as less than 0.1M.

Step 1 Conduct the search indicated by the results in the panel below.

Search results ~ Exercise 2 ~ Part 3 ~ Example 4

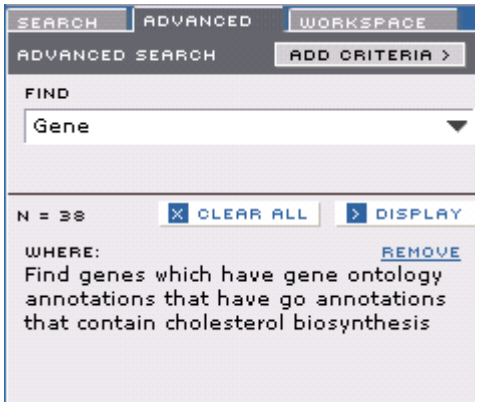


Exercise 2 ~ Part 4 ~ Search for Genes ~ Example 1

The Gene Ontology Consortium has annotated many genes according to highly curated categories. Another way to get a list of genes involved in the cholesterol biosynthesis pathway is to search for genes that match a GO annotation such as cholesterol biosynthesis.

Step 1 Conduct the search indicated by the results in the panel below.

Search results ~ Exercise 2 ~ Part 4 ~ Example 1

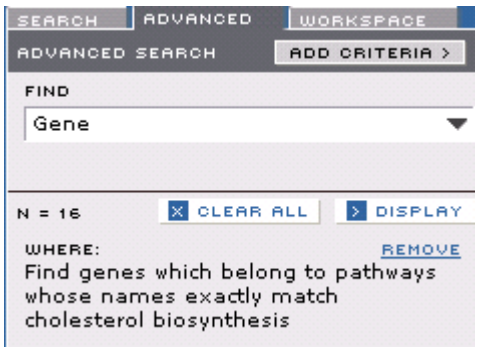


Exercise 2 ~ Part 4 ~ Search for Genes ~ Example 2

We can also extract gene lists for Iconix curated pathways directly rather than loading a selected pathway as we did in **Exercise 2 ~ Part 2 ~ Example 2**.

Step 1 Conduct the search indicated by the results in the panel below.

Search results ~ Exercise 2 ~ Part 4 ~ Example 2

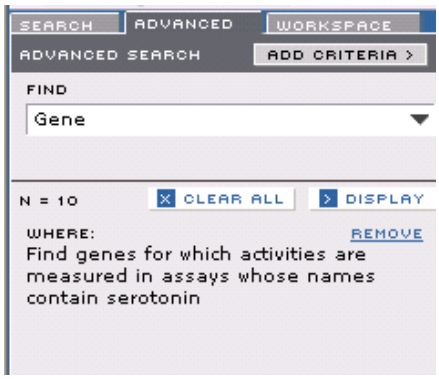


Exercise 2 ~ Part 4 ~ Search for Genes ~ Example 3

Another cross domain link allows us to search for genes whose protein product is the basis of a biochemical assay in DrugMatrix.

Step 1 Conduct the search indicated by the results in the panel below.

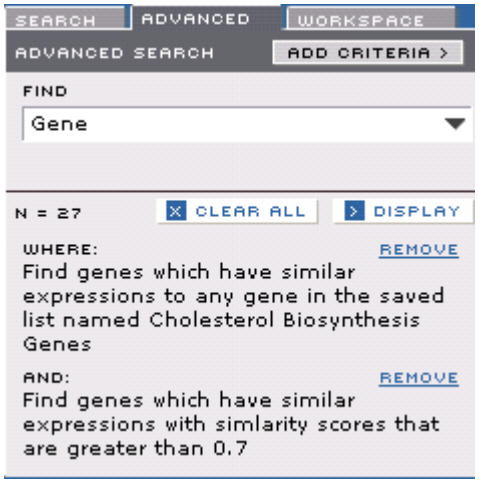
Search results ~ Exercise 2 ~ Part 4 ~ Example 3



Exercise 2 ~ Part 4 ~ Search for Genes ~ Example 4

This example shows you how to find genes whose expression profiles are similar to genes involved in the cholesterol biosynthesis pathway. This is useful if you want to find new targets that may be associated with the pathway, or genes in other pathways that are co-regulated by the selected pathway.

Search results ~ Exercise 2 ~ Part 4 ~ Example 4



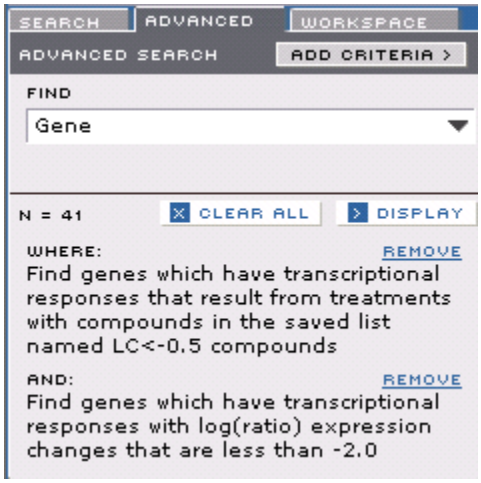
Exercise 2 ~ Part 4 ~ Search for Genes ~ Example 5

This example shows you one way to identify genes that are strongly repressed by any of the compounds in a specific list.

Step 1 Create a list of compounds that reduce leukocyte count (as in **Exercise 2 ~ Part 2 ~ Example 1**).
Step 2 Save the list (for example **Reduce Leukocyte Count by 3-fold**).
Step 3 Open the **TOOLBOX** and Drag the **Reduce Leukocyte Count by 3-fold** list onto the **EXPRESSION EXPERIMENTS > COMPOUNDS** tool.

Step 4 In dialog box that appears, name the list to be translated in the **ENTER NEW LIST NAME** field, for example **LC<-0.5 compound**.
Step 5 Click the **TRANSLATE** button at the bottom of the panel.
Step 6 Conduct the search indicated by the results in the panel below.
This will be a relatively slow query.

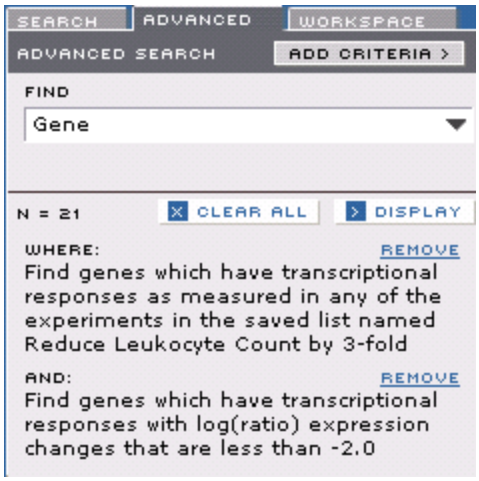
Search results ~ Exercise 2 ~ Part 4 ~ Example 5



Exercise 2 ~ Part 4 ~ Search for Genes ~ Example 6

Below is an alternative approach to identifying genes that are strongly repressed by compounds that reduce leukocyte counts. In this example, we are looking for strongly repressed genes in experiments that are already defined as reducing white cell counts.
Step 1 Create a list of expression experiments that reduce leukocyte count (as in **Exercise 2 ~ Part 2 ~ Example 1**).
Step 2 Save the list (for example **Reduce Leukocyte Count by 3-fold**).
Step 3 Conduct the search indicated by the results in the panel below.
This should be a much faster search than the previous example.

Search results ~ Exercise 2 ~ Part 4 ~ Example 6

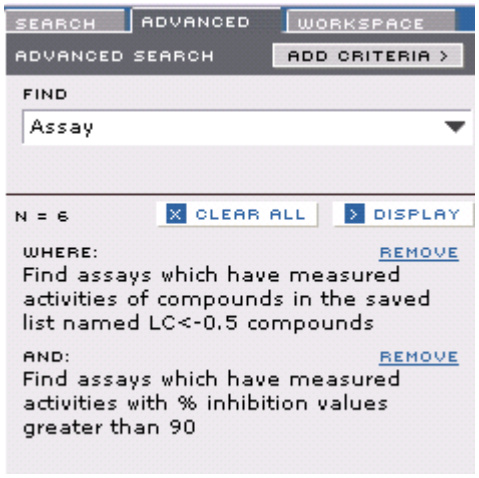


Exercise 2 ~ Part 5 ~ Search for Assays ~ Example 1

This example illustrates how you can identify relevant assays that are strongly affected by compounds in a list. In this case, we will search for biochemical assays which are strongly inhibited by compounds that decrease leukocyte count.

Step 1 Conduct the search indicated by the results in the panel below.

Search results ~ Exercise 2 ~ Part 5 ~ Example 1



Conclusions to Exercise 2

In this exercise, you have learned to use the advanced query tool in DrugMatrix and see examples of several different types of queries you can generate. As you become familiar with the DrugMatrix suite of tools, you will be able to generate sophisticated queries specific to your particular research.

Exercise 3: Upload Customer Expression Data and Interpret Affymetrix Gene Expression Data Using Gene Ontology Tool

DrugMatrix can be used to interpret the effects of candidate compounds on gene expression. As long as the experiments are structured appropriately, this analysis can be done whether you use Codelink, Affymetrix or Agilent arrays.

To prepare for this exercise, you will need to download the file **AffyData.zip** from the **"Help"** page in DrugMatrix (link to "Help" page is in the upper right of page after login), and extract the contents to a folder on your computer. These files contain Plier normalized, expotentiated data. Each file represents an individual sample. In these files, column 1 entitled "Probe Set Name "contains the probe names and column 2 entitled "Signal" contains the Plier normalized expotentiated intensity values.

DrugMatrix can help you interpret gene expression data using tools that combine expression information with curated functional information. For example, the Gene Ontology Consortium has curated information for a large fraction of the genes that are used on the expression arrays. This information can be very useful when trying to understand the pattern of expression changes observed in an experiment.

Gene Ontology annotations describe a gene product in terms of three hierarchies:
Molecular Function (MF, the biochemical activity of the protein, e.g. Kinase)
Cellular Component (CC, the place in a cell where the protein is active, e.g. Extracellular)
Biological Process (BP, biochemical 'objective' to which the protein contributes, e.g.: Induction of apoptosis)

In this exercise, you will learn to:
• Upload Affymetrix gene expression data using the load data tool
• Use the Gene Ontology tool to
• Explore a gene expression profile
• Determine which genes participate in a particular biological process
• Create lists of genes that meet certain criteria

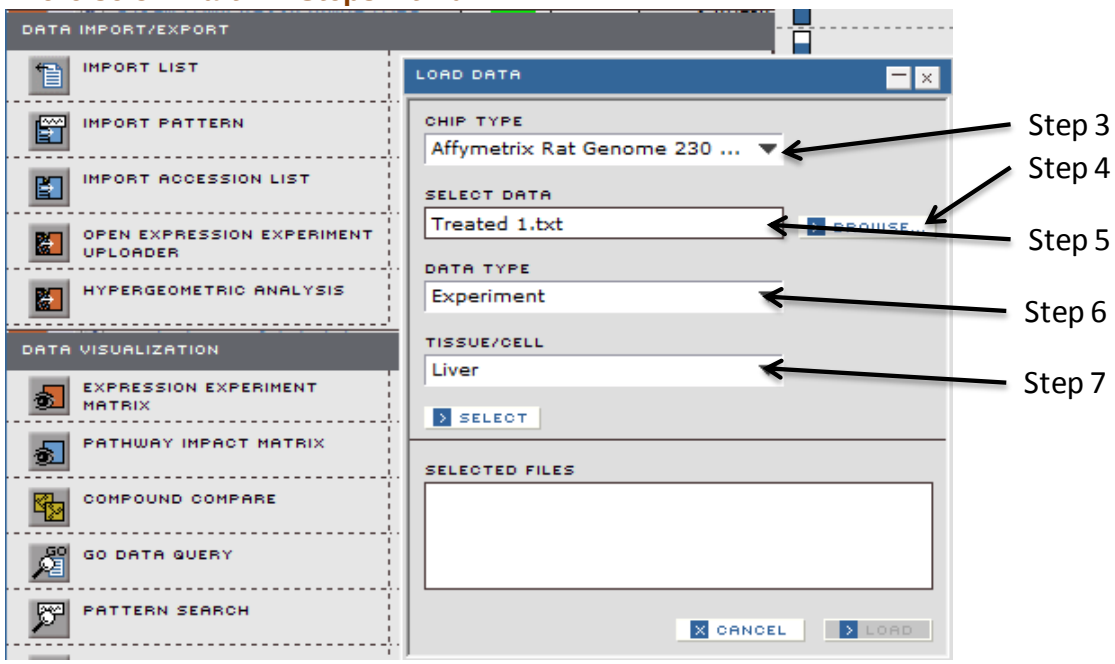
Exercise 3 ~ Part 1 ~ Upload Data Derived from Affymetrix RG230-2 chips

In this exercise, you will upload microarray data and explore information presented in the **EXPRESSION** Domain Report.

Step 1 Click **TOOLBOX** button in **WORKSPACE**.
Step 2 Click **OPEN EXPRESSION EXPERIMENT UPLOADER**.
Step 3 In the **LOAD DATA** dialog, choose **Affymetrix Rat Genome 230 2.0** from the **CHIP TYPE** select menu.

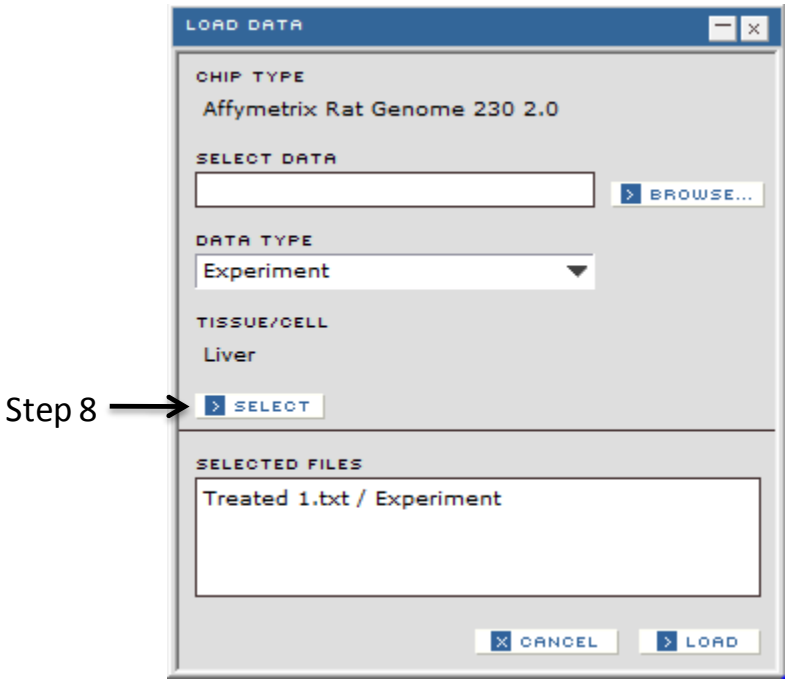
Step 4 Click on the **BROWSE** button and navigate to the file **Treated 1.txt**.
Step 5 Click **OPEN** on the Windows **"CHOOSE FILE"** dialog window. The experiment file name should appear in the **SELECT DATA** field.
Step 6 Choose **EXPERIMENT** from the **DATA TYPE** select menu.
Step 7 Choose Liver from the **TISSUE** select menu.
At this point the **LOAD DATA** dialog should look like this.

Exercise 3 ~ Part 1 ~ Steps 2 thru 7



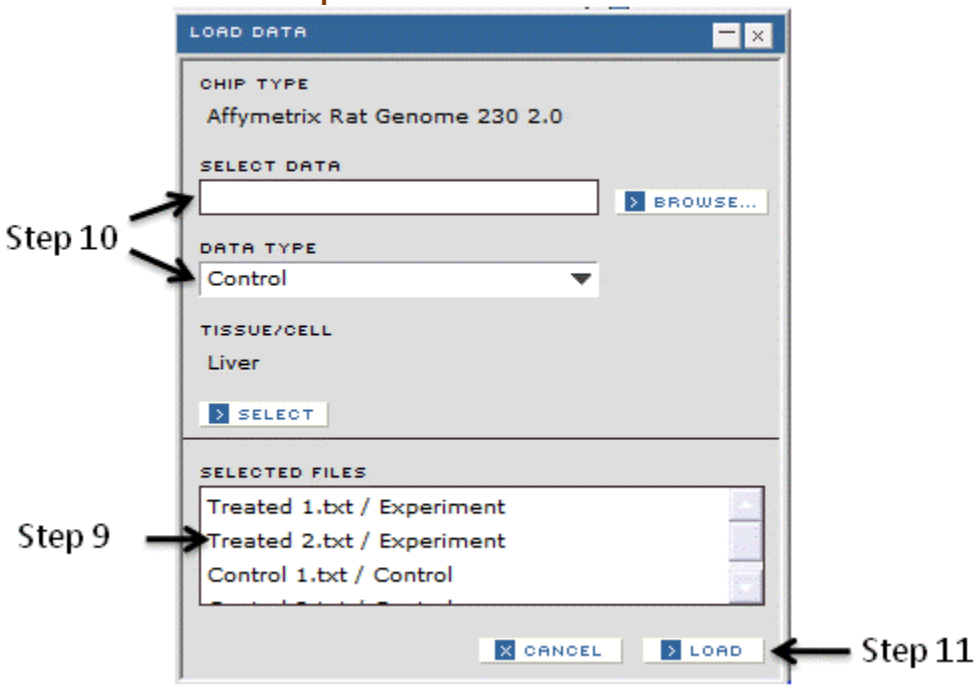
Step 8 Click the **SELECT** button.
The name of the uploaded file should appear in the **SELECTED FILES** field.

Exercise 3 ~ Part 1 ~ Step 8



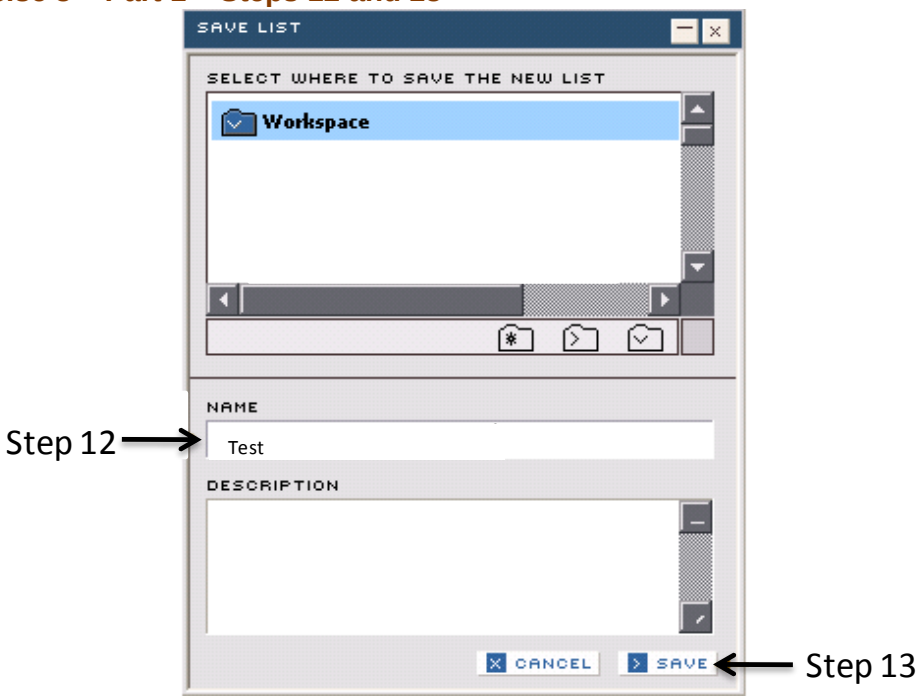
Step 9 Click the **BROWSE** button and upload the other experiment sample, **Treated 2.txt**, as in the previous steps.
Step 10 Click the **BROWSE** button and select file **Control 1.txt,** choose **CONTROL** from the **DATA TYPE** select menu and then follow the same steps to upload the control experiment **Control 2.txt.**
Step 11 Once all files are uploaded, click on the **LOAD** button.

Exercise 3 ~ Part 1 ~ Steps 9 thru 11



Step 12 A **SAVE LIST** pop-up dialog window will appear. Save the experiments as **Test**
Step 13 Click on the **SAVE** button.

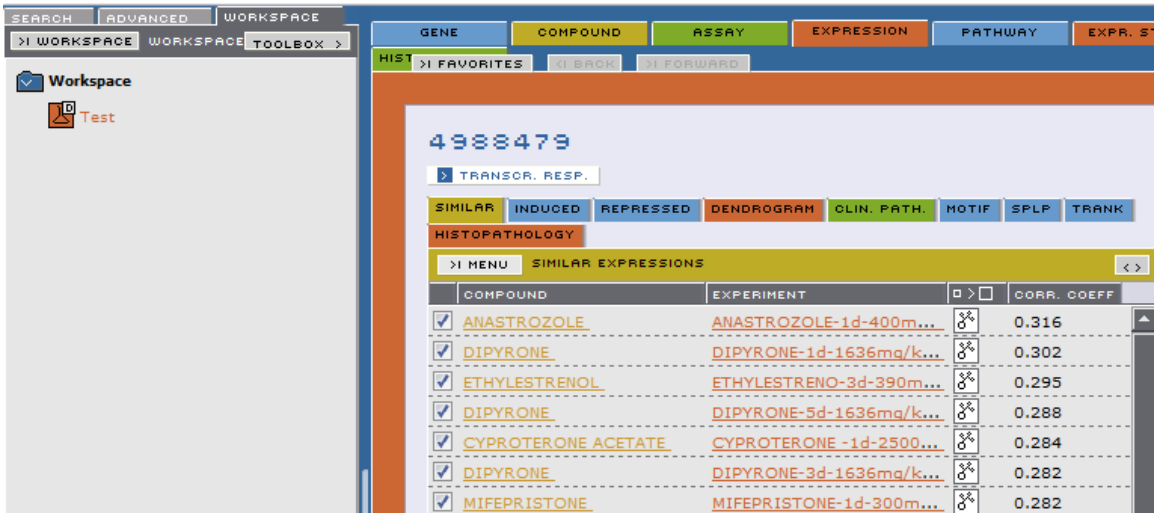
Exercise 3 ~ Part 1 ~ Steps 12 and 13



The information from the uploaded files will populate an **EXPRESSION** Domain Report panel. This information is further explored in Part 2 of this exercise.

Exercise 3 ~ Part 2 ~ Explore Affymetrix Array Results

Once you have uploaded the Affymetrix array files, DrugMatrix will automatically calculate the ratio, and the results will populate an **EXPRESSION** Domain Report. The first panel displayed in this report is the **SIMILAR** panel.

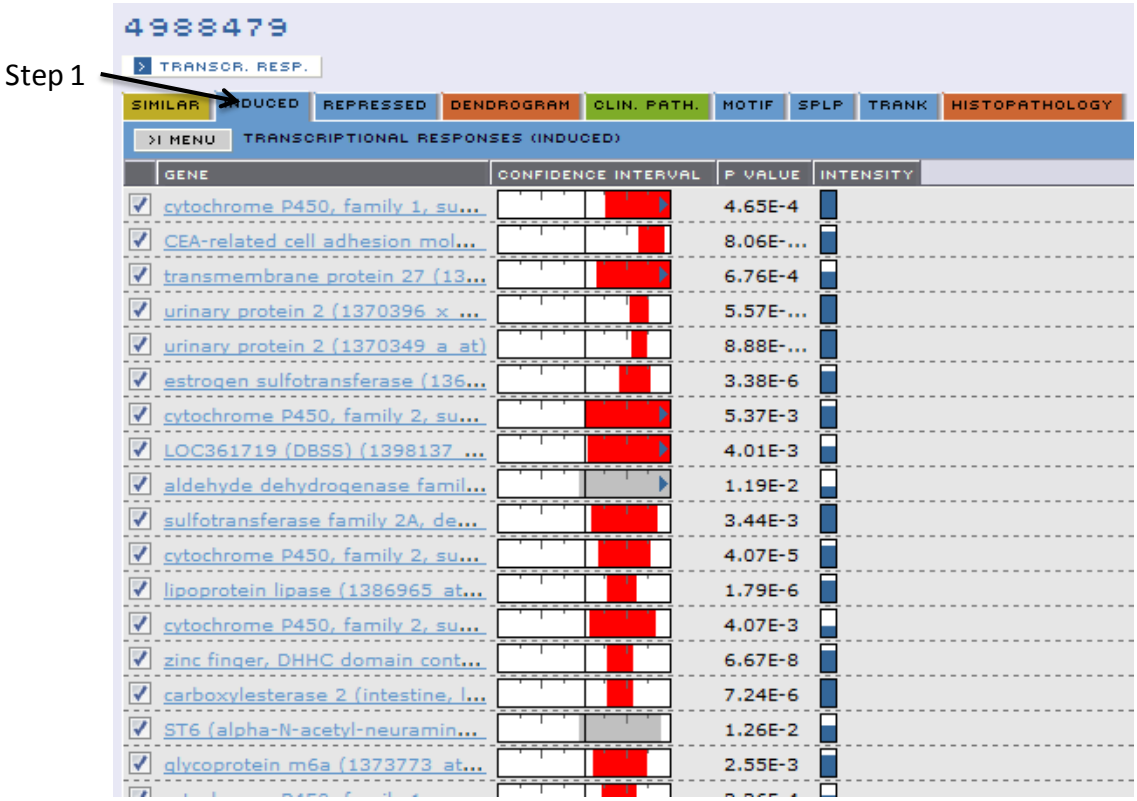


This panel presents a list of expression experiments similar to the uploaded Affymetrix experiments. The experiments are ordered by their similarity to the uploaded experiments in expression profile measured by the Pearson's correlation coefficient as indicated in the **CORR. COEFF.** column. The **COMPOUND** column shows the compound used in each of these experiments. As expected, the experiments most similar to the uploaded Affymetrix experiments are bacitracin treatments.

Next you will use the Gene Ontology tool in DrugMatrix to investigate gene expression changes induced by the uploaded Affymetrix data.

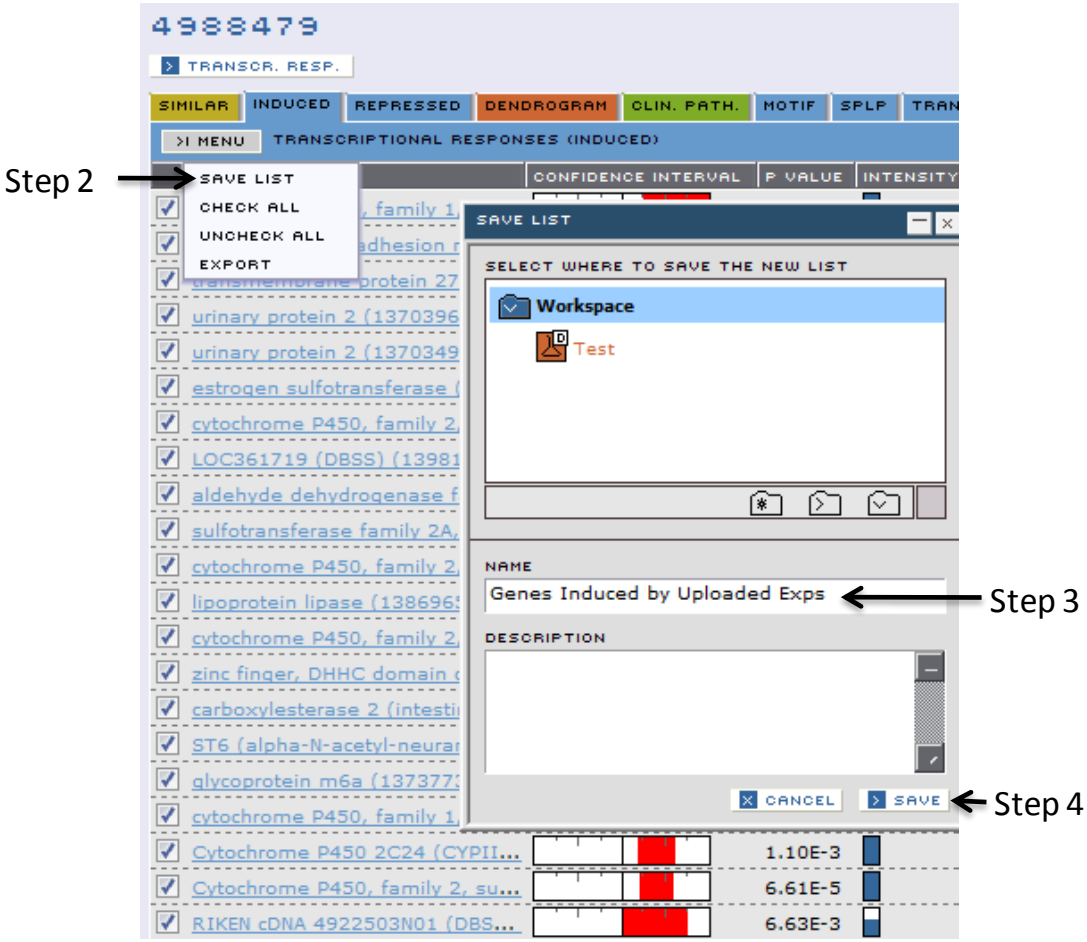
Step 1 Click on the **INDUCED** sub tab within the **EXPRESSION** Domain Report to get a rank ordered, list of genes that are induced by your compound. These are all genes on the Affymetrix gene chip platform.

Exercise 3 ~ Part 2 ~ Step 1



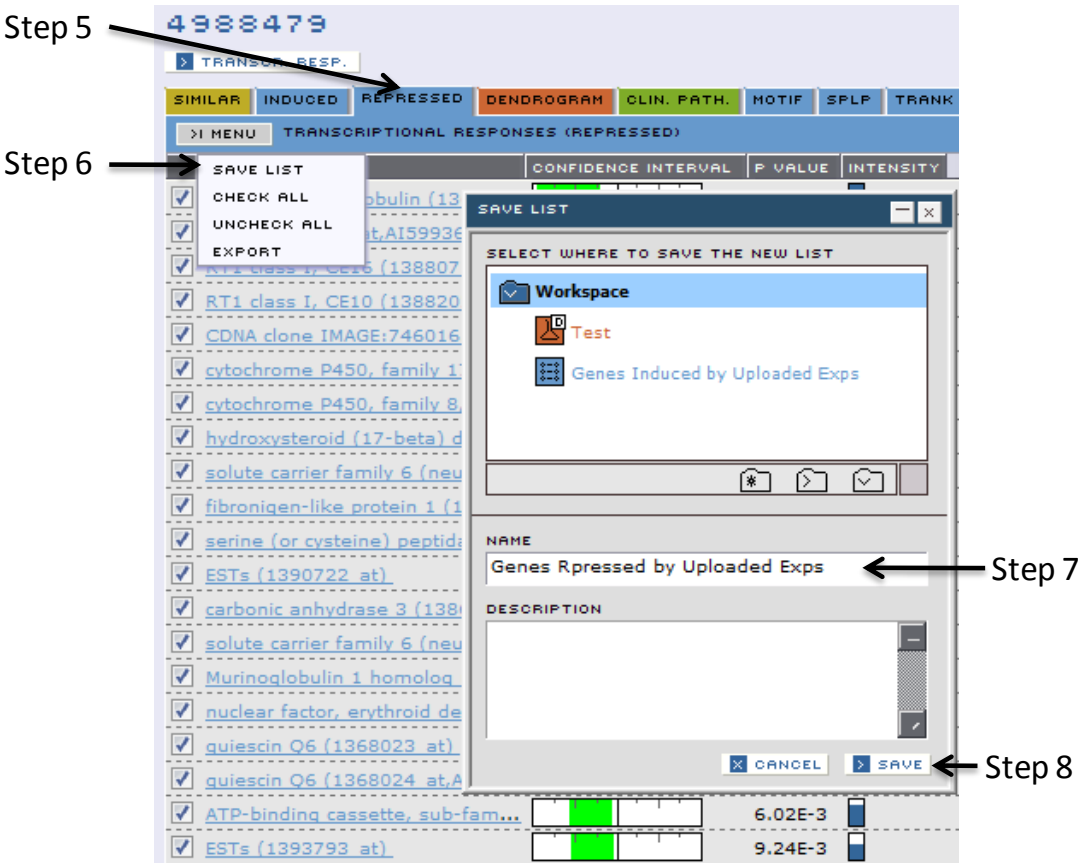
Step 2 Click on the sub-panel **MENU** button and select **SAVE LIST**.
Step 3 In the **SAVE LIST** dialog window, enter **Genes Induced by Uploaded Exps** in the **NAME** field.
Step 4 Click on the **SAVE** button to save the list.

Exercise 3 ~ Part 2 ~ Steps 2 thru 4



Step 5 Click on the **REPRESSED** sub tab within the **EXPRESSION** Domain Report to get a rank ordered, list of genes that are repressed by your compound.
Step 6 Click on the sub-panel **MENU** button and select **SAVE LIST**.
Step 7 In the **SAVE LIST** dialog window, enter **Genes Repressed by Uploaded Exps** in the **NAME** field.
Step 8 Click on the **SAVE** button to save the list.

Exercise 3 ~ Part 2 ~ Steps 5 thru 8



Since all of the bioinformatics work to map these lists of genes to other chips and to public annotations has been done, you can begin to explore the meaning of this list of genes in a statistical analysis tool called the **GO DATA QUERY** tool.

Step 9 Select the **WORKSPACE** tab, you can mouse over the names of the lists you just saved and see a hover indicating the number of genes saved in the list.

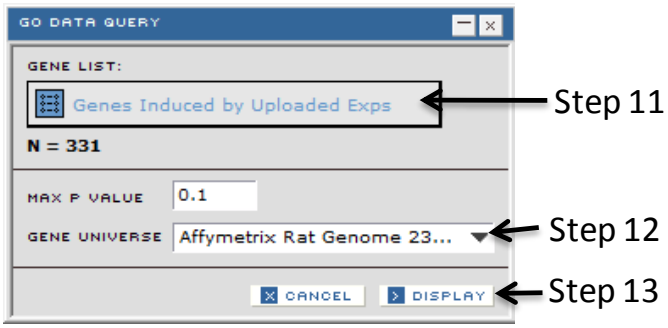
Step 10 Click on the **TOOLBOX** button underneath the **WORKSPACE** tab.

Step 11 Drag the **Genes Repressed by the Uploaded Exps** list onto the **GO DATA QUERY** tool in the **DATA VISUALIZATION** section of the **TOOLBOX**.
This will open a **GO DATA QUERY** dialog window.

Step 12 Leave the default values in the **MAX P VALUE** field and select **Affymetrix Rat Genome 230 2.0** in the **GENE UNIVERSE** field.

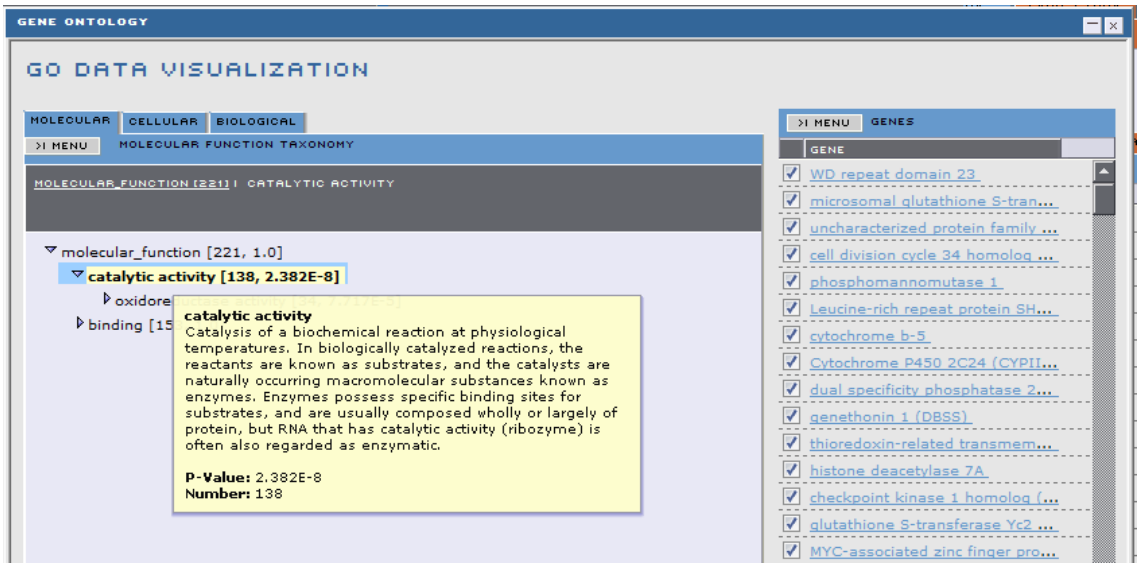
Step 13 Click on the **DISPLAY** button.

Exercise 3 ~ Part 2 ~ Steps 11 thru 13

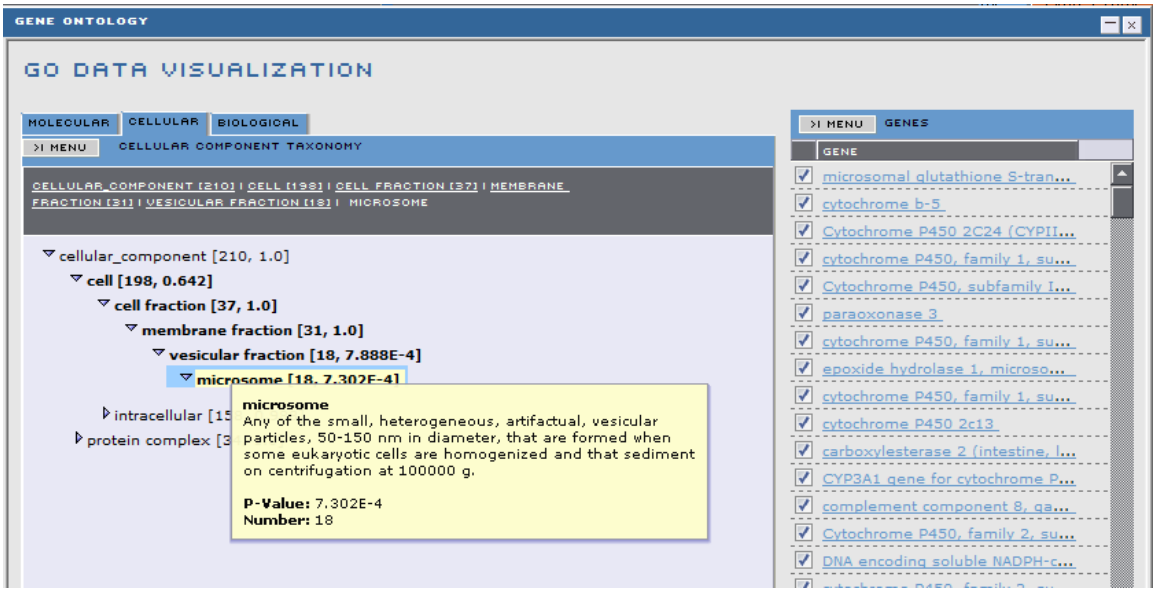


This will display the **GO DATA VISUALIZATION** with three tabs for three Gene Ontology hierarchies, and a list of genes on the right. The brackets next to each **GO** term indicate the number of times the term is found to be associated with a gene in the provided gene list and the **P-Value** (the probability that these gene counts would occur in a random sample of genes). The most statistically significant term within each node is **bolded.**

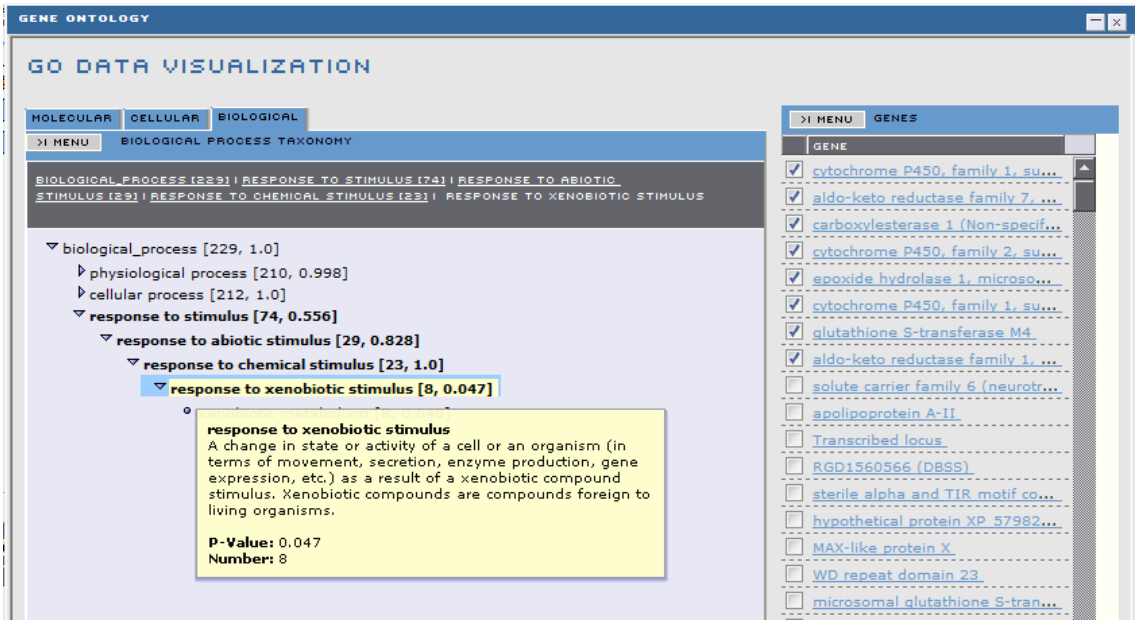
MOLECULAR panel showing the hierarchy for the Molecular Function



CELLULAR panel showing the hierarchy for the Cellular Component



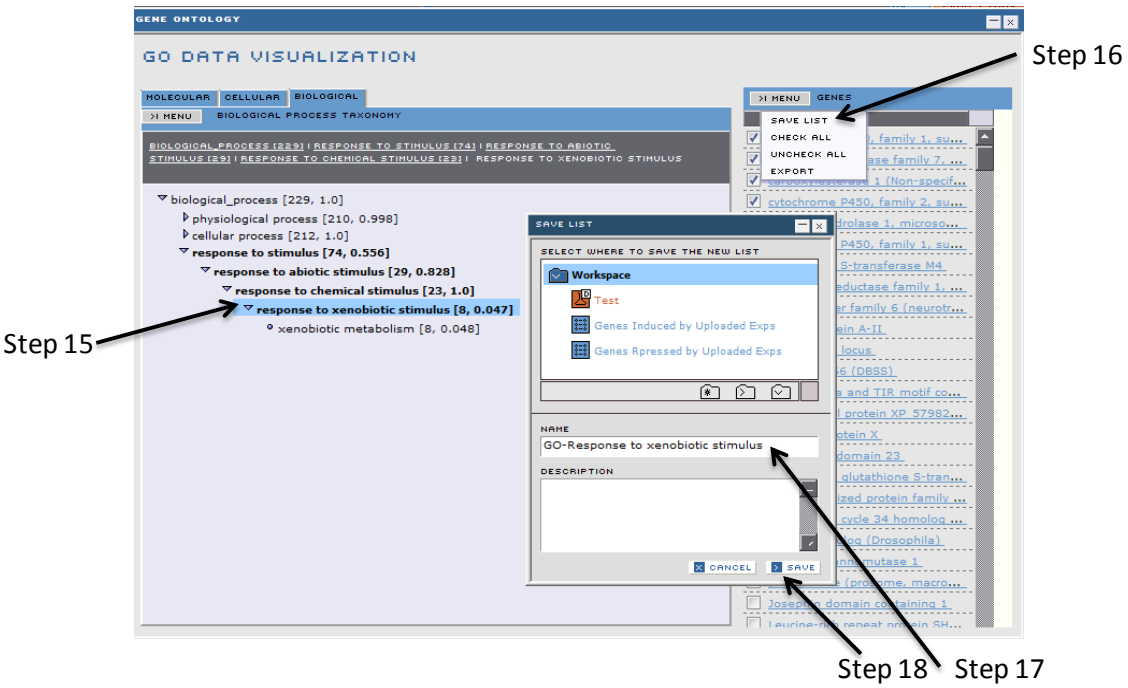
BIOLOGICAL panel showing hierarchy for Biological Process



Step 14 Mouse over the Gene Ontology term of interest in the list and a hover will display description of the Gene Ontology annotation.
Step 15 Click on a term of interest and the corresponding genes associated with that term will be highlighted in the right hand gene panel.
Step 16 If you want to save this list of genes, click on the sub-panel **MENU** button and select **SAVE LIST**.
Step 17 In the **SAVE LIST** dialog window, enter a gene list name in the **NAME** field.

Step 18 Click on the **SAVE** button.

Exercise 3 ~ Part 2 ~ Step 15 thru 18



Conclusions to Exercise 3

In this exercise you learned how to upload microarray data from the Affymetrix platform into the DrugMatrix database, and analyze the data you uploaded using the **GO QUERY** tool to understand expression profiles and functions of various genes with Gene Ontology annotations. You also learned how to create lists of genes associated with a particular Gene Ontology annotation.

Exercise 4 Visualizing Expression Data in the Context of Pathways

Understanding the pattern of gene expression in the context of biological pathways can be a useful aid in interpreting results. DrugMatrix has 137 hand curated pathways involved in key biological and toxicological processes. The following exercise will demonstrate how to look at treatment, compounds or tissue-related gene transcription changes in a pathway.
In this exercise you will visualize the gene transcription changes in a pathway to allow comparisons between different compounds and tissues.

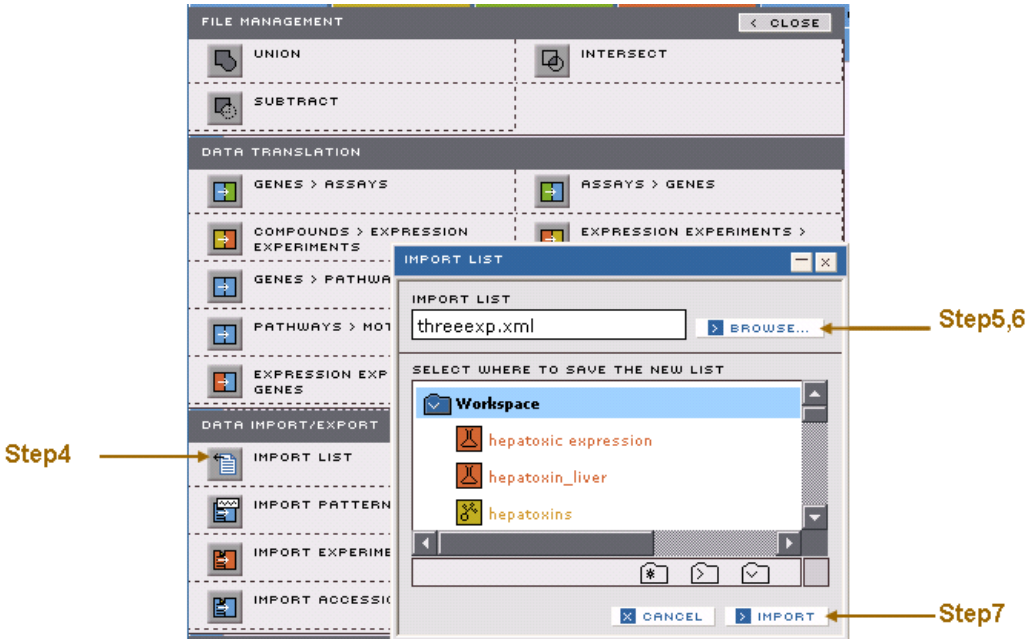
Exercise 4 ~ Part 1 ~ Examining Gene Expression Via Pathways

To prepare for this exercise, you will need to obtain sample files from the Iconix demo website.
Step 1 Go to **"Help"** page in DrugMatrix (link to "Help" page is in the upper right of page after login)
Step 2 Right click on **threeexp.xml** and select **Save Target As...** from the menu and save these files with the same name to a folder on your computer. This file contains the following three experiments:

NORETHINDRON-3d-375mg/kg-LI-RATM-RU1 (norethindrone liver gene expression)
BEZAFIBRATE-3d-100mg/kg-LI-RATM-RU1 (bezafibrate liver gene expression)
BEZAFIBRATE-3d-100mg/kg-KI-RATM-RU1 (bezafibrate kidney gene expression)

Step 3 Select the **WORKSPACE** tab and open **TOOLBOX**.
Step 4 Click on the **IMPORT LIST** tool to open the **IMPORT LIST** dialog.
Step 5 Click the **BROWSE** button on the **IMPORT LIST** dialog window
Step 6 Locate the file **threeexp.xml** and click the **OPEN** button on the **CHOOSE FILE** window. The file name will appear in the **IMPORT LIST** field of the **IMPORT LIST** dialog window.
Step 7 Click on the **IMPORT** button to import the file into Workspace.

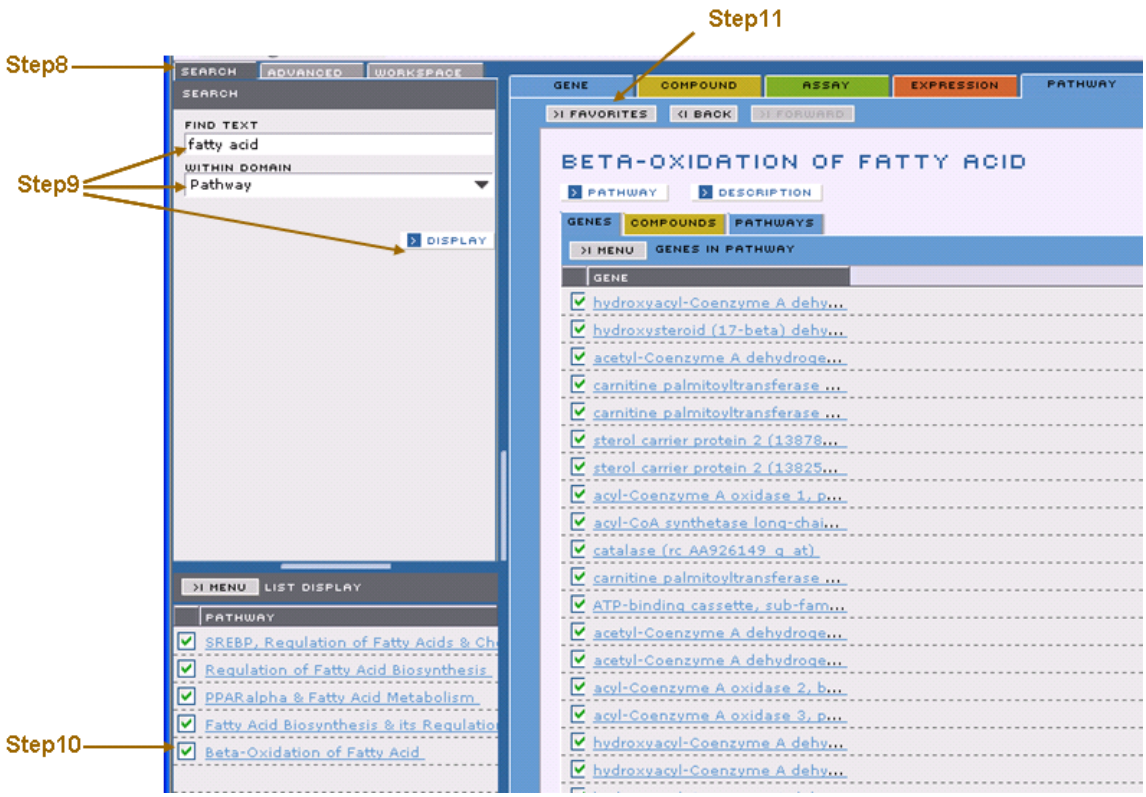
Exercise 4 ~ Part 1 ~ Steps 4 thru 7



Next we will identify a pathway that we want to visualize these data on, and put it into "**FAVORITES**" so that we can easily retrieve it.

Step 8 Click on the **SEARCH** tab
Step 9 Select **PATHWAYS** from the pull-down menu, type **FATTY ACID** in the **FIND TEXT** field and click on the **DISPLAY** button.
Step 10 Click on **Beta-oxidation of Fatty Acid** pathway to population the **PATHWAY** domain report with all the information available in DrugMatrix about the fatty acid beta-oxidation pathway.
This pathway is a catabolic pathway that oxidizes long chain fatty acids into acetyl Coenzyme A, a common substrate in many other pathways. It is the primary way your body retrieves energy from stored fats.
Step 11 Click on **FAVORITES** button and select **ADD TO FAVORITES.**

Exercise 4 ~ Part 1 ~ Steps 8 thru 11



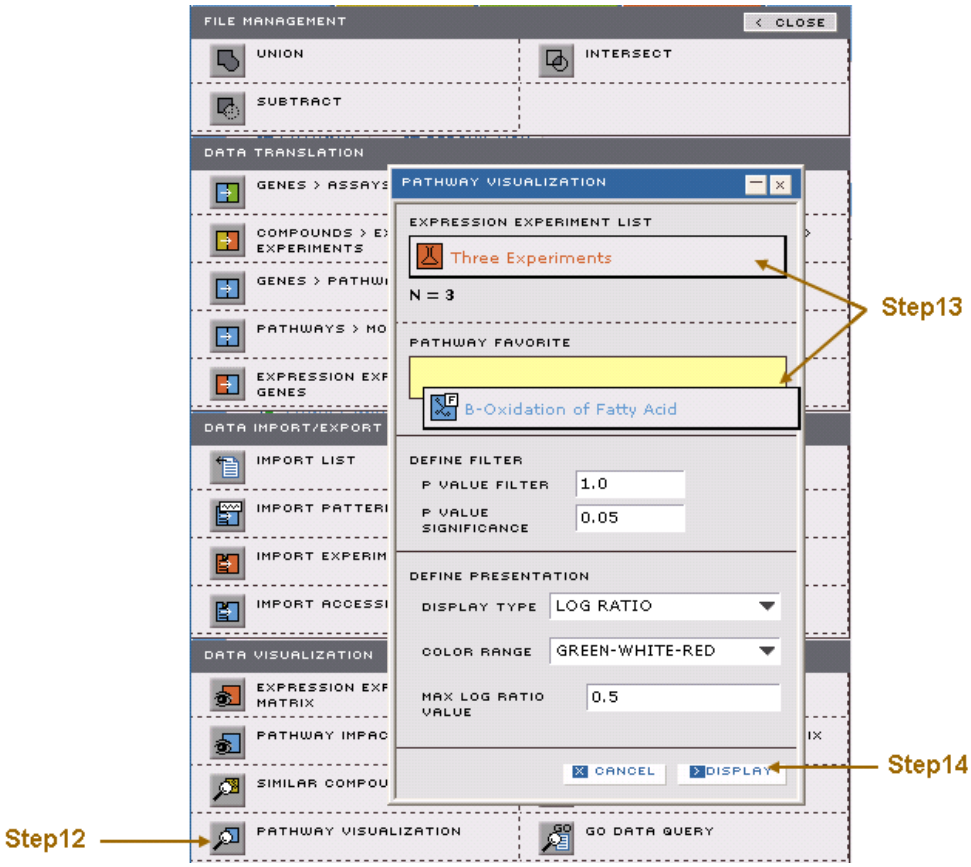
Now we will overlay the Three Experiments that we uploaded onto the **B-oxidation of Fatty Acid** pathway that we saved as a Favorite.

Step 12 Open the **TOOLBOX** and select the **PATHWAY VISUALIZATION** tool from the **DATA VISUALIZATION** section at the bottom of the panel.

Step 13 Drag the **Three experiments** list into the **EXPRESSION EXPERIMENT LIST** entry box and the **B-oxidation of Fatty Acid** pathway into the **PATHWAY FAVORITE** entry box.

Step 14 Leave all the options as default, click on **DISPLAY** button at the bottom of the dialog box.

Exercise 4 ~ Part 1 ~ Steps 12 thru 14

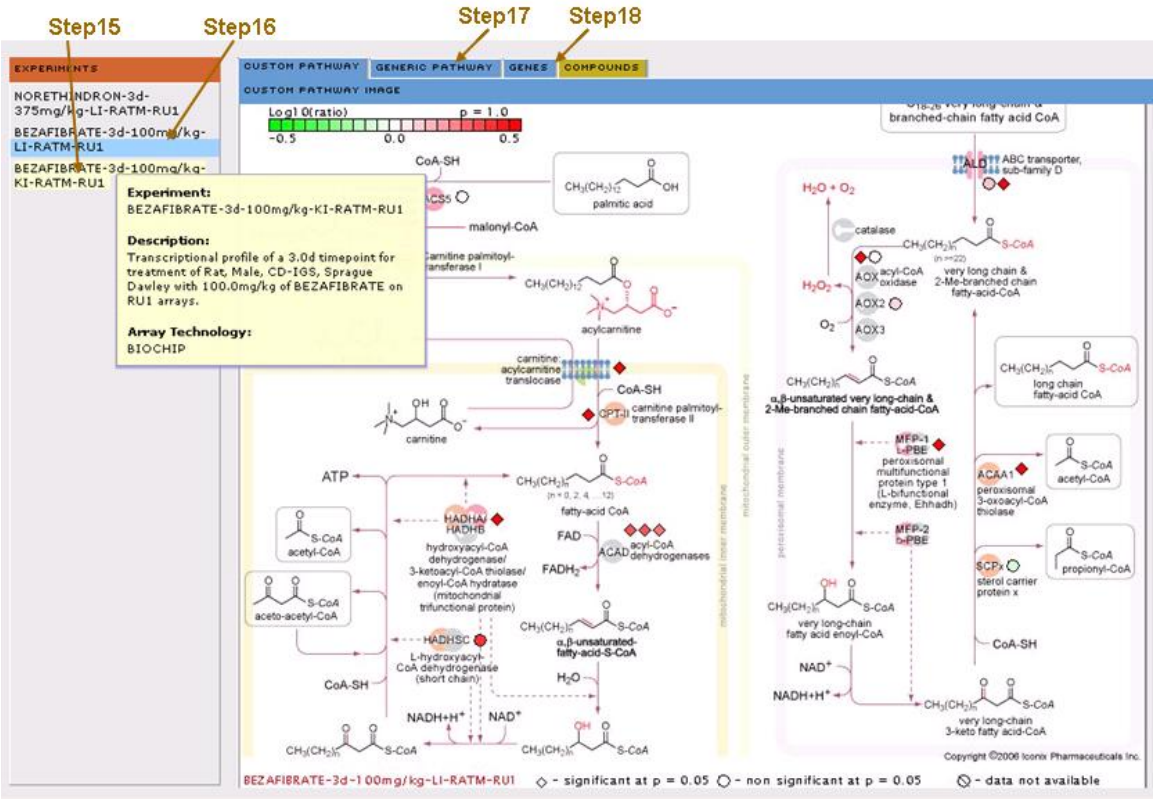


The DrugMatrix application will retrieve the gene expression data corresponding to the genes known to be involved in the fatty acid beta-oxidation pathway. The colored dots next to each gene are the expression values for each probe or probe-set that can be associated with the described gene. Red indicates up-regulation and green indicates down-regulation. Significantly changed genes ($p<0.05$) are represented by a diamond.

Within this visualization, you can:
Step 15 Mouse over experiments in the **EXPERIMENTS** panel to display a hover containing descriptive information about each experiment. The tissue used appears as part of the experiment name (e.g. **LI-RATM-RU1** or **KI-RATM-RU1**).
Step 16 When you click on a different experiment name it will display the custom pathway for that experiment in the **CUSTOM PATHWAY** panel.
Step 17 Click on the **GENERIC PATHWAY** tab to view the generic (no expression data overlayed) beta-oxidation pathway.
Step 18 Click on the **GENES** tab to view the list of genes in this curated beta-oxidation pathway.

Note that **Bezafibrate** is a drug that causes peroxisome proliferation by increasing the number of cell organelles that oxidize fat. Most of the genes in the beta-oxidation pathway are significantly induced in the liver by the administration of 100mg/kg of bezafibrate. This pathway appears to be composed of co-regulated genes.

Exercise 4 ~ Part 1 ~ Steps 15 thru 18



Step 19 Select the third experiment on the **EXPERIMENT** panel, which is also the 100mg/kg treatment with bezafibrate, but in this case we are visualizing kidney gene expression data.
Note that in this experiment, far fewer genes seem to be affected. This is likely because peroxisomes are induced in the liver, not the kidney.

Conclusions to Exercise 4

In this exercise, you learned to use the pathway visualization tool to study the regulation of genes in the beta-oxidation pathway in different treatments and tissues.

Exercise 5: Finding Similarities and Differences Among Compounds Using Pathway Analysis

This exercise demonstrates the use of several DrugMatrix tools for expression data comparison in the context of biological pathways as well as carrying out a multi-step drill-down to relevant data.

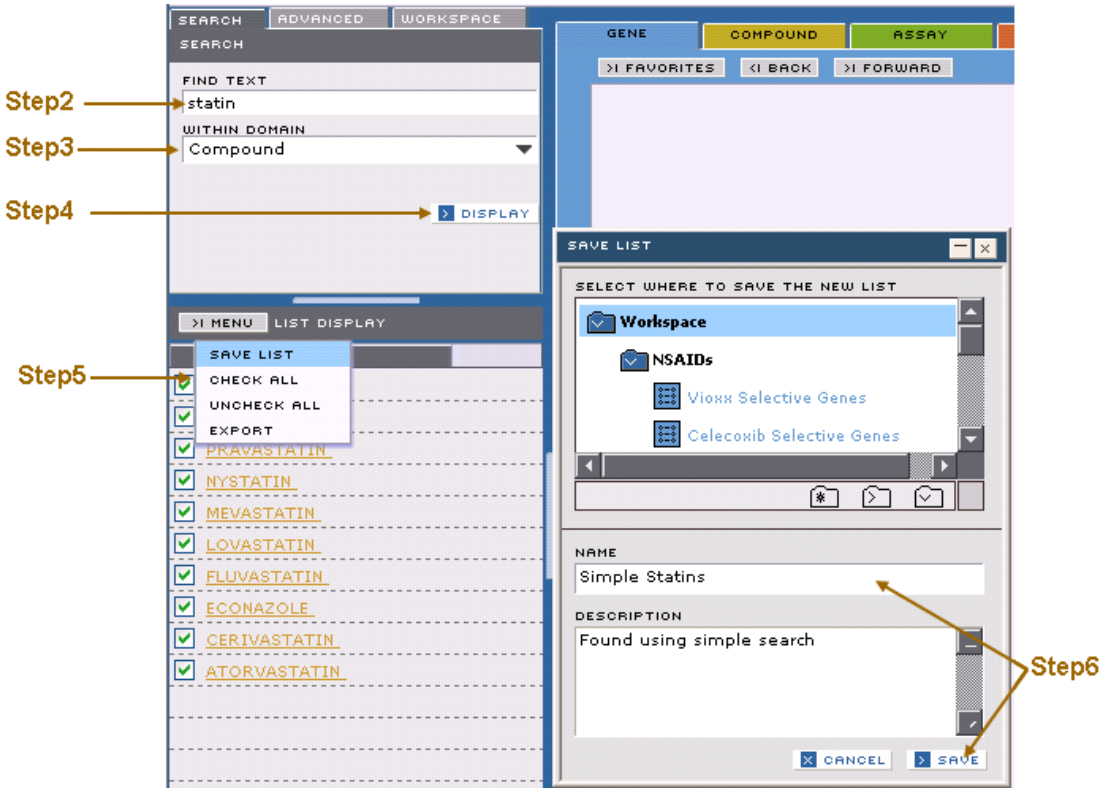
In this exercise, you will learn to:
• Use the simple and advanced search features to construct compound lists and convert a compound list into an experiment list.
• Use the **PATHWAY IMPACT** tool to identify most impacted pathways.
• Use the **PATHWAY > GENE** translation tool to convert a pathway to a gene list.
• Use the **EXPRESSION EXPERIMENT MATRIX** tool to explore similarities and differences between compounds at the gene expression level.

Exercise 5 ~ Part 1 ~ Creating an Expression Experiment List

The goal in this section is to find statins (cholesterol lowering compounds) in DrugMatrix, identify experiments in DrugMatrix where animals were treated with these compounds and generate a list of these experiments in order to carry out comparisons.

Step 1 Select the **SEARCH** tab.
Step 2 Type **STATIN** in the **FIND TEXT** field.
Step 3 Select **COMPOUND** from the drop down menu.
Step 4 Click on the **DISPLAY** button.
Step 5 Select **SAVE** from the **MENU** on the **LIST DISPLAY** panel.
Step 6 In the **SAVE LIST** dialog window, save the list as **Simple statins**.

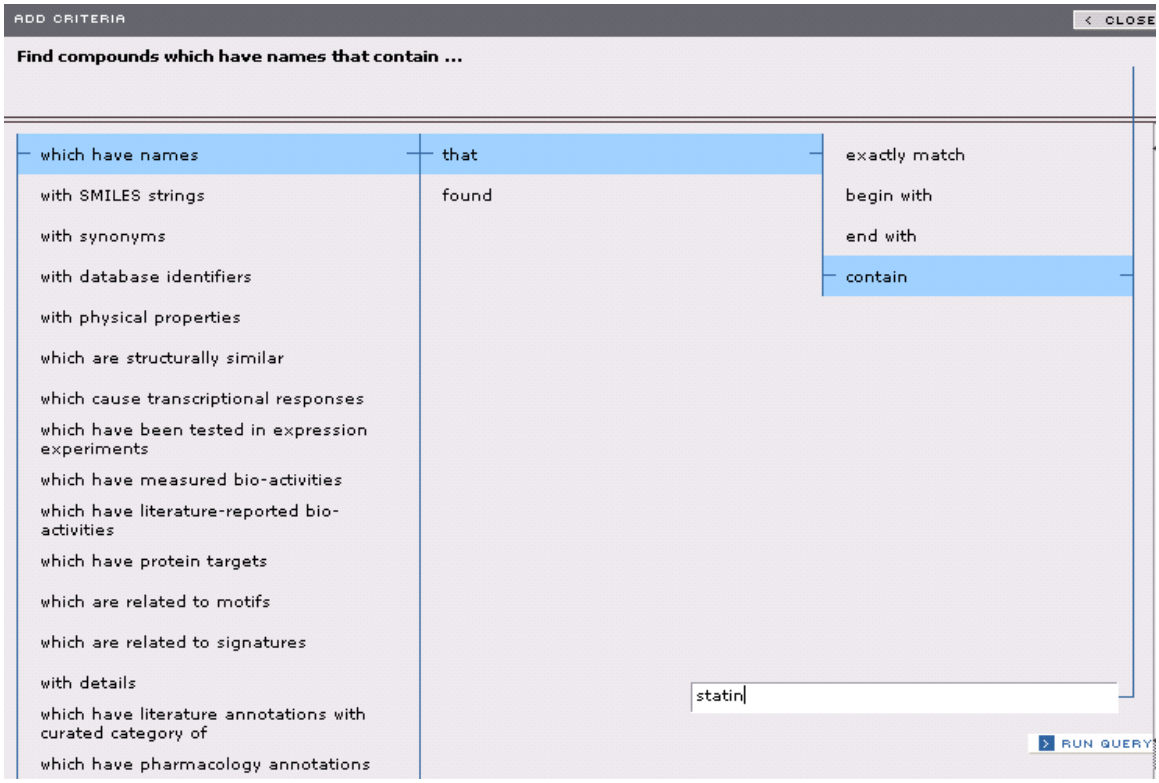
Exercise 5 ~ Part 1 ~ Steps 2 thru 6



This same search can be carried out using the **ADVANCED** search tool.

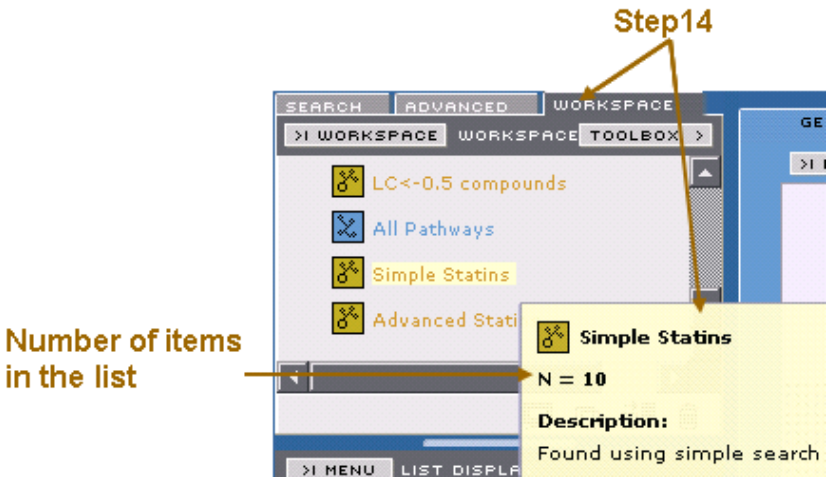
Step 7 Click on the **ADVANCED SEARCH** tab and click the **CLEAR ALL** button to remove all previous criteria
Step 8 Choose **COMPOUND** from the **FIND** select menu.
Step 9 Click **ADD CRITERIA**.
Step 10 Make the selections indicated on the **ADD CRITERIA** panel below.
Step 11 Click the **RUN QUERY** button.

Exercise 5 ~ Part 1 ~ Steps 10 and 11



The results of this search will appear under the **ADVANCED** tab.
Step 12 Click the **DISPLAY** button to display the list
Step 13 Save the resulting list as in steps 5 and 6 but give this list the name **Advanced statins**
Step 14 Select the **WORKSPACE** tab and mouse over the two lists **Simple Statins** and **Advanced Statins**. Note that the two lists contain different numbers of compounds.

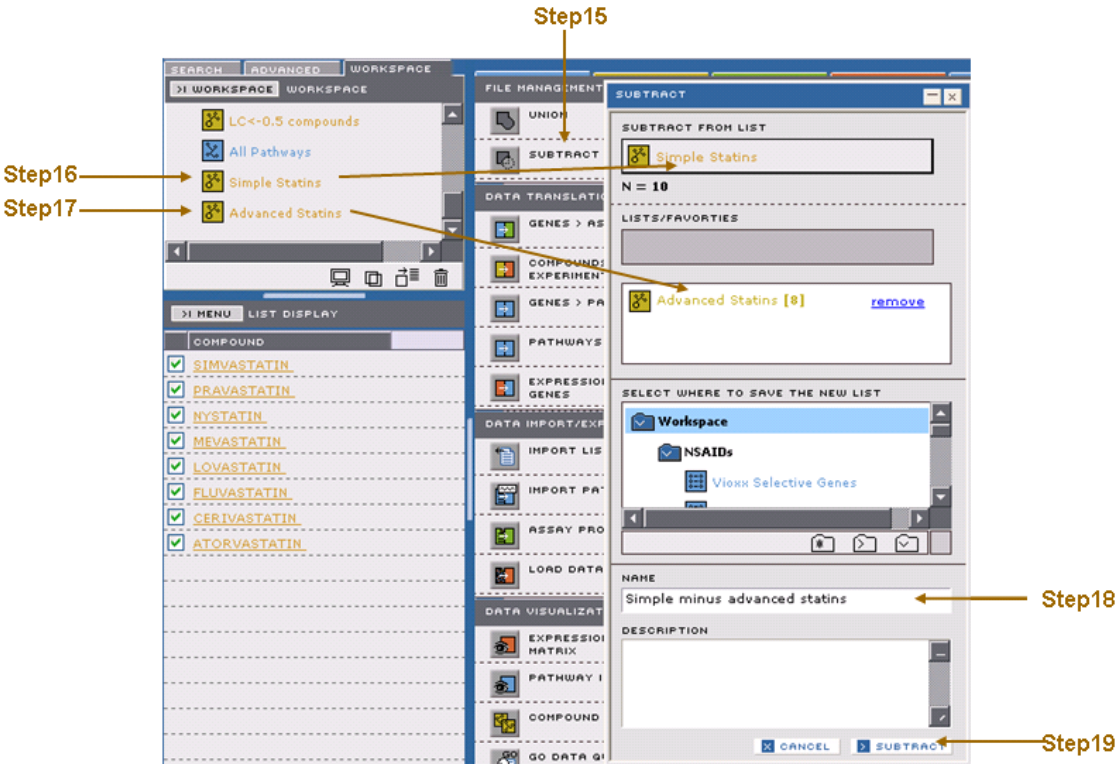
Exercise 5 ~ Part 1 ~ Step 14



Step 15 Click the **TOOLBOX** button on the **WORKSPACE** panel and open the **SUBTRACT** tool from the **LIST EDIT** section at the top of the toolbox.
Step 16 Drag the **Simple Statins** list to the **SUBTRACT FROM LIST** field of the **SUBTRACT** dialog window.
Step 17 Drag the **Advanced Statins** list to the **LIST/FAVORITES** field.

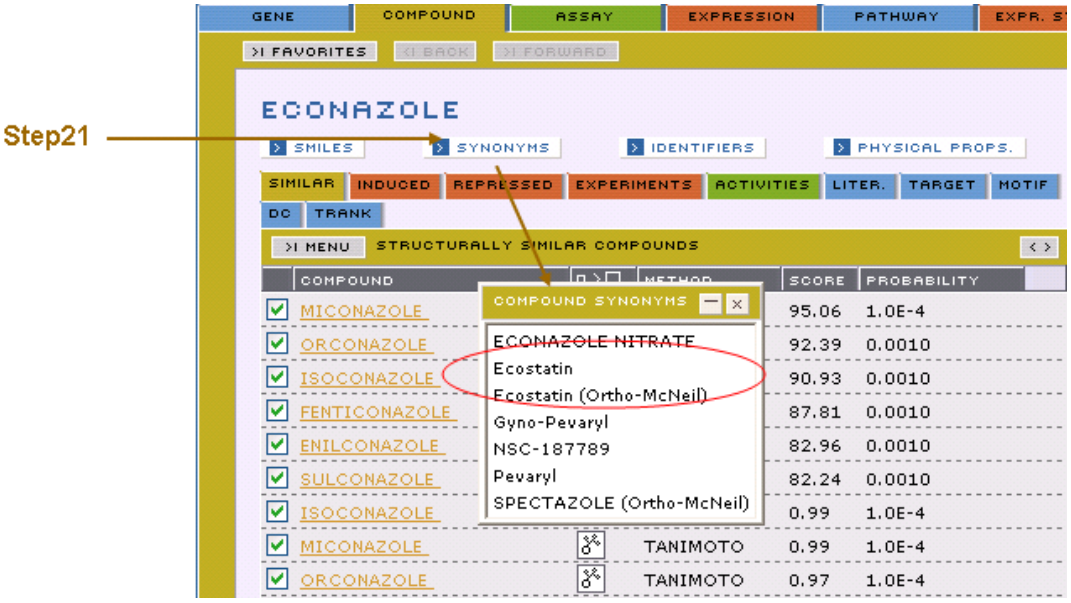
Note: You can put only one list into the **SUBTRACT FROM LIST** field. All lists you drag into the **LIST/FAVORITES** field will be subtracted from that list.
Step 18 Name the list **Simple minus advanced statins**
Step 19 Click the **SUBTRACT** button.

Exercise 5 ~ Part 1 ~ Steps 15 thru 19



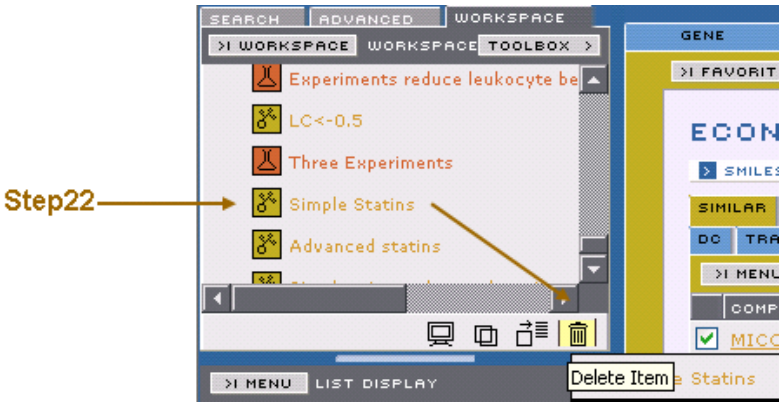
The list **Simple minus advanced statins** will be placed in your **WORKSPACE** and displayed in the **LIST DISPLAY** area. This list contains the items that were in the **Simple Statins** list but not in the **Advance Statins** list.
To understand why these compounds were extracted from the data base using the simple search technique...
Step 20 Click on **ECONAZOLE**.
This will populate a **COMPOUND** domain report with information about the compound **ECONAZOLE**.
Step 21 Click on the **SYNONYMS** button

Exercise 5 ~ Part 1 ~ Step 21



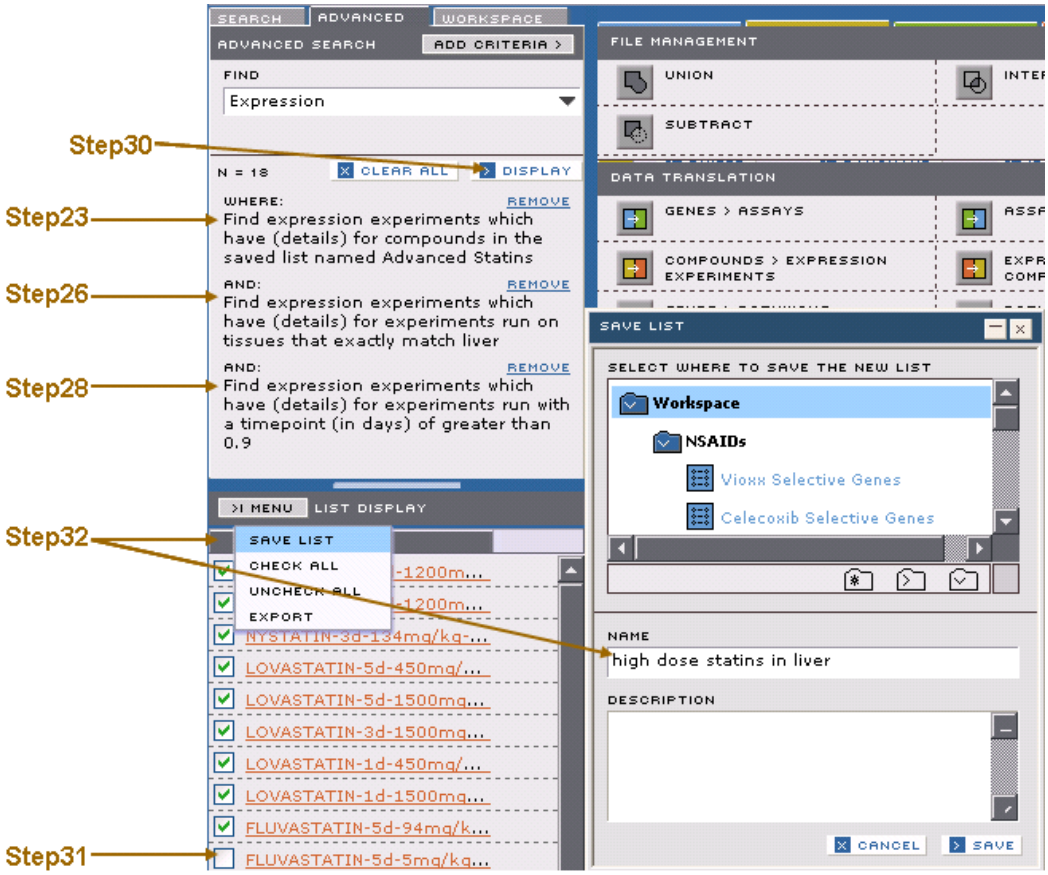
The Simple Search found all instances of the string **statin**, but the Advanced Search found only compounds that contained **statin** in the compound name, since that is what you specified.
The **Advanced Statins** list is the one we'll use, so delete the **Simple Statins** list to avoid confusion.
Step 22 Drag and drop the **Simple Statins** list onto the **DELETE ITEM** icon (trash can).

Exercise 5 ~ Part 1 ~ Step 22



Now add several sets of criteria to query for a specific set of expression experiments.

Click **ADVANCED** tab, and select **EXPRESSION**.
Step 23 Click the **ADD CRITERIA** button to find expression experiments...
-which have (details) for
-compounds
-in the saved list named
-[Advanced statins]
Step 24 Click **RUN QUERY**.
Step 25 Click on the **DISPLAY** list button to see the list of results.
This list returns all experiments that were run on any of the statins in our list. Take note of the number of experiments returned (shown on the left of **ADVANCED SEARCH** panel). We want to refine our query so we will add more criteria to the search. First, let's limit the list to experiments run on liver only.
Step 26 Click **ADD CRITERIA** and select
-which have (details) for
-experiments run on tissues
-that exactly match
-[liver]
Step 27 Click the **RUN QUERY** button.
This should narrow the list of experiments to approximately one third of the number. Next, let's remove time-points of less than 1 day.
Step 28 Click **ADD CRITERIA** and select
-which have (details) for
-experiments run with a timepoint (in days) of
-greater than
-[0.9]
Step 29 Click the **RUN QUERY** button.
This should further narrow the list of experiments to a manageable size.
Step 30 Click **DISPLAY** to display the list in the **LIST DISPLAY** area.
Next, let's remove all of the low dose (below 15 mg/kg) experiments.
Step 31 Go through the list and **UNCHECK** all of the low dose experiments (i.e. below 15mg).
Step 32 Now save the list as **high dose statins in liver**.



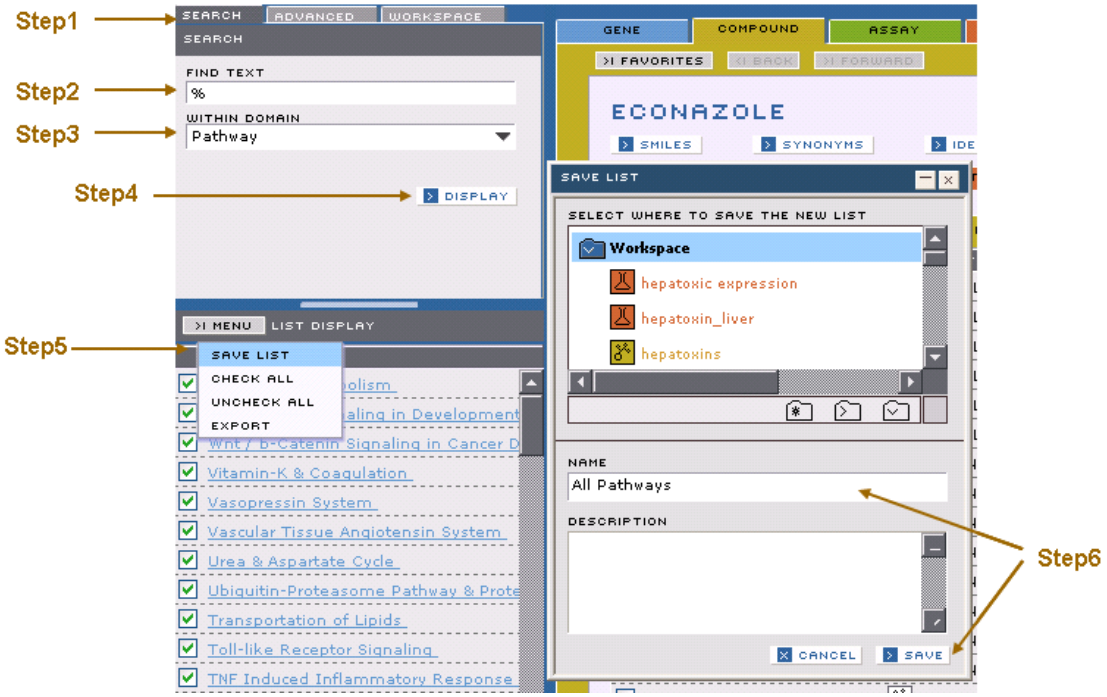
Exercise 5 ~ Part 2 ~ Pathway Impact Analysis

In this exercise, we will visualize the impact of the saved experiments on all 137 of the curated pathways using the Pathway Impact Matrix Tool. The pathway impact analysis tool is an approach to rapidly quantify the impact of an experiment on the 137 pathways in DrugMatrix. Statistical enrichment of significantly impacted genes in a pathway is determined using a hypergeometric distance metric, which is based on the fraction of the pathway genes that are significantly induced ($p<0.05$) by the experiment relative to the total number of all genes significantly induced by the experiment. The impact score is expressed as a –log (p-value), which reflects the likelihood of randomly perturbing that pathway.

First, we will conduct a Simple Search to create a list of all pathways.
Step 1 Click on the **SEARCH** tab.
Step 2 Type **%** in the **FIND TEXT** field.
Note: The **%** sign is a universal wildcard symbol within DrugMatrix, so this search will find all pathways within the **PATHWAY** domain.
Step 3 Select **PATHWAY** from the **WITHIN DOMAIN** select menu.
Step 4 Click the **DISPLAY** button to display the list.
Step 5 Select **SAVE LIST** from the **MENU** button on the **LIST DISPLAY** panel.

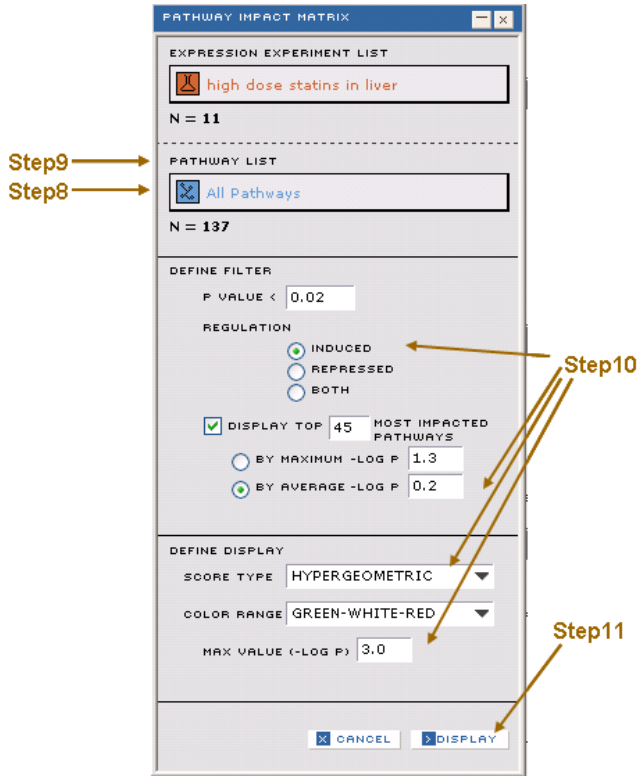
Step 6 Save the resulting list as **all pathways**.

Exercise 5 ~ Part 2 ~ Steps 1 thru 6

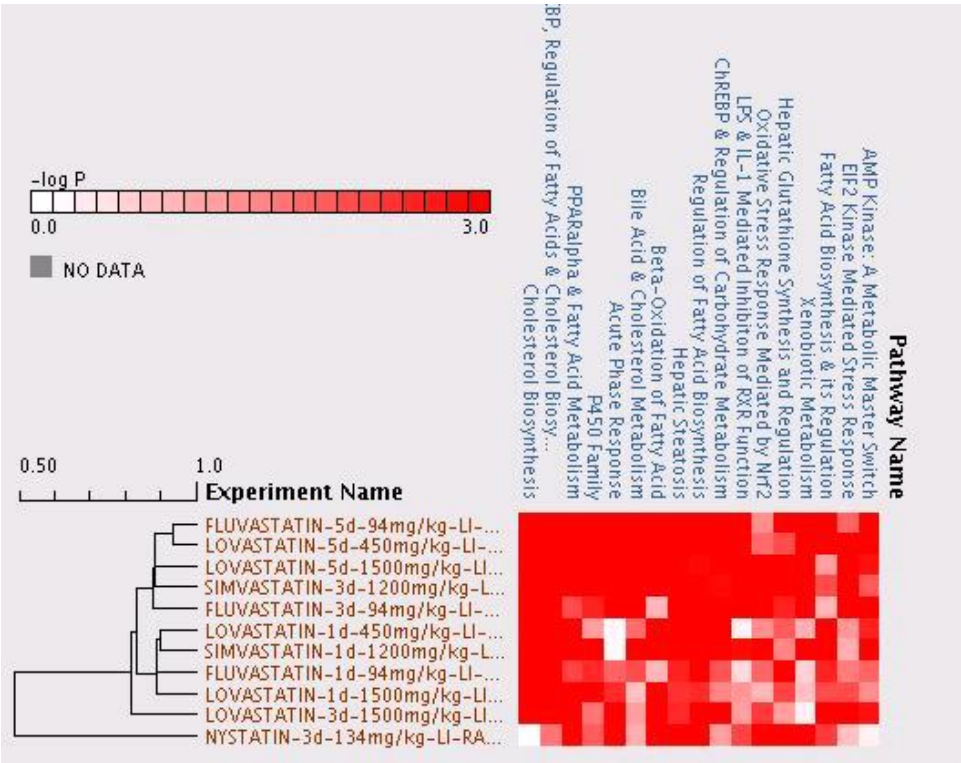


Step 7 Select the **WORK SPACE** tab and then open the toolbox and select the **PATHWAY IMPACT MATRIX** tool.
Step 8 Drag the **all pathways** list into the **PATHWAY LIST** box.
Step 9 Drag the **high dose statins in liver** list into the **EXPERIMENT LIST** box.
Step 10 Select options as indicated below.
Step 11 Click the **DISPLAY** button at the bottom of the dialog window.

Exercise 5 ~ Part 2 ~ Steps 8 thru 11



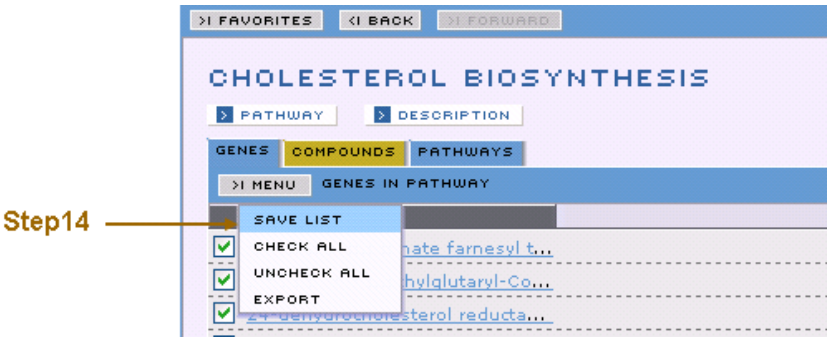
Exercise 5 ~ Part 2 ~ DATA VIEW for results from Steps 8 thru 11



This visualization clearly shows that the HMG CoA reductase inhibitor drugs cluster very tightly together, whereas the anti fungal drug nystatin clusters separately. The HMG-CoA reductase inhibitors strongly impact the Cholesterol Biosynthesis pathway, but nystatin does not.

Step 12 Click on the **PATHWAY** tab in the **PATHWAY IMPACT MATRIX** window. This shows a list of pathways being significantly impacted by the selected list of experiments.
Step 13 Click on the **Cholesterol Biosynthesis** pathway, which will open the pathway domain
Step 14 Under the **GENES** tab, select **SAVE LIST** from the drop-down **MENU** button to save the list of **Cholesterol Biosynthesis Genes**.

Exercise 5 ~ Part 2 ~ Step 14

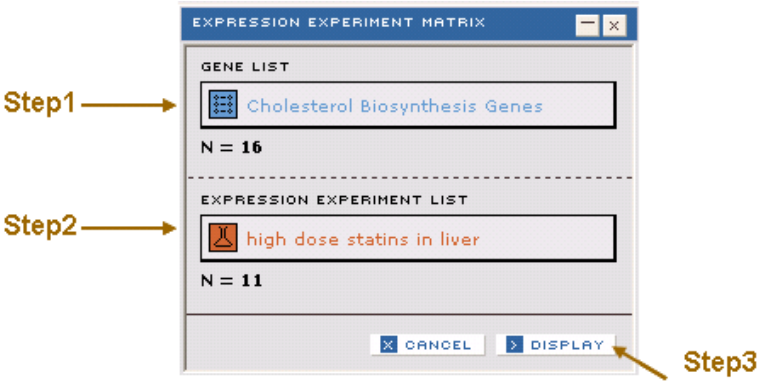


Exercise 5 ~ Part 3 ~ Expression Experiment Matrix tool

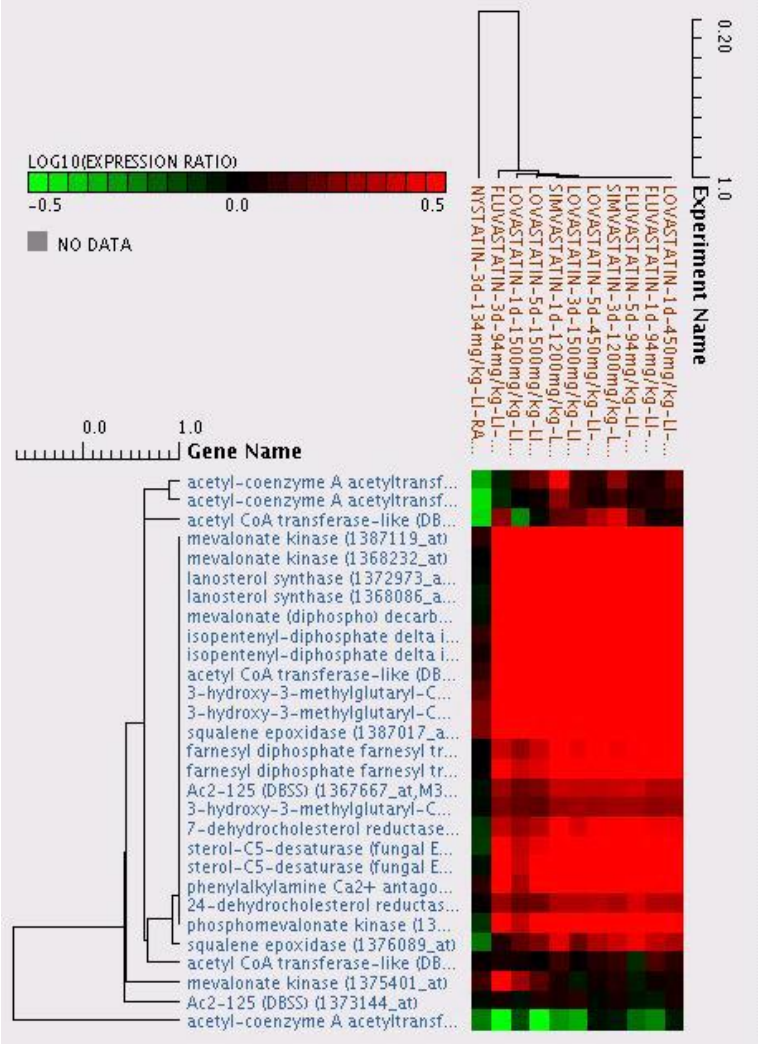
In this exercise, we will visualize the impact of treatment with statins on our list of pathway genes using the Expression Experiment Matrix tool.

Step 1 Drag the **Cholesterol Biosynthesis Genes** gene list to the **EXPRESSION EXPERIMENT MATRIX** tool.
Step 2 Drag the **high dose statins in liver** experiment list to the **EXPERIMENT LIST** field of the **EXPRESSION EXPERIMENT MATRIX** tool.
Step 3 Click on the **DISPLAY** button at the bottom of the **EXPRESSION EXPERIMENT MATRIX** dialog window.

Exercise 5 ~ Part 3 ~ Steps 1 thru 3



Expression Experiment Matrix output for Cholesterol Biosynthesis Genes gene list and high dose statins in liver experiment list



As expected, all the HMG-CoA reductase inhibitors induce genes involved in the cholesterol biosynthesis pathway, but nystatin does not.

Conclusions to Exercise 5

In this exercise you used the simple and advanced search features to construct compound lists, and used the **DATA TRANSLATION** feature to convert that compound list into an experiment list.
You used the **PATHWAY IMPACT MATRIX** tool to determine pathways of potential interest and the **EXPRESSION EXPERIMENT MATRIX** tool to explore similarities and differences at the individual gene expression level between a group of compound treatments.

Exercise 6: Evaluation of a Literature-derived Multi-gene Biomarker for Non-genotoxic Carcinogenicity Using the Pattern Tool

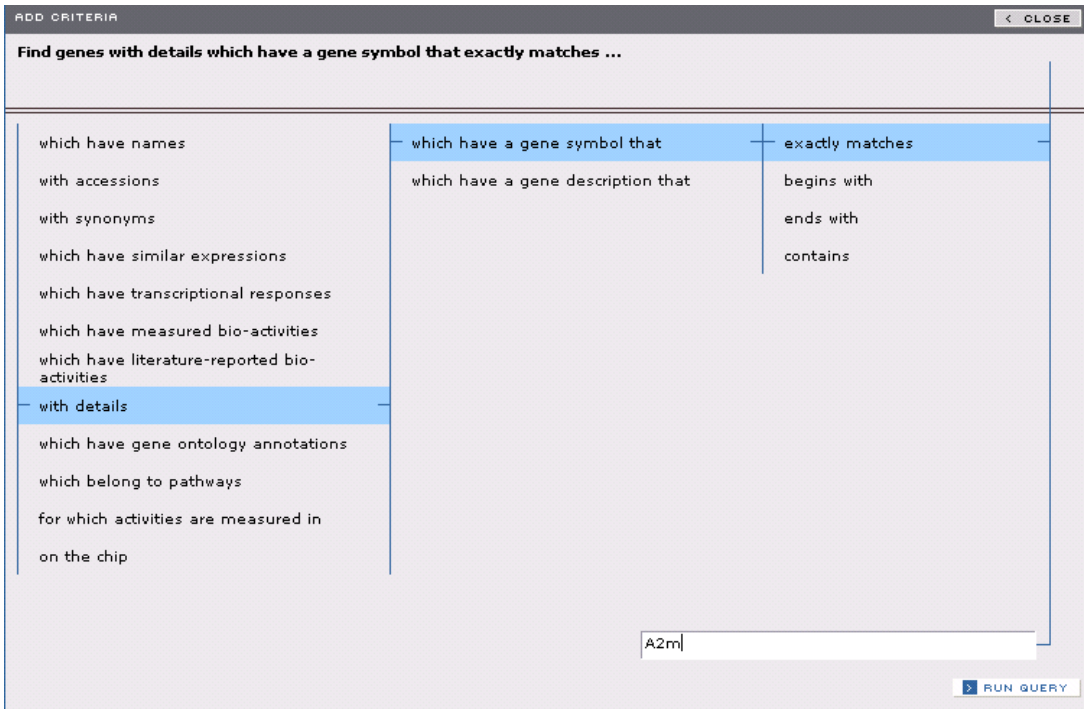
The DrugMatrix database contains a vast amount of compound annotation information and gene expression data and has proven to be very useful for evaluation of putative biomarkers. The following exercise will illustrate how to construct a literature-derived multi-gene biomarker for non-genotoxic carcinogens using the pattern export and import tool and then evaluate the performance of this literature-based biomarker using the pattern search tool.

Exercise 6 ~ Part 1 ~ Construct a Literature-derived Gene List for Non-genotoxic Carcinogens

Numerous papers have been published in recent years that have identified biomarkers for non-genotoxic carcinogens, with the hope of eventually replacing the industry and FDA standard (but very expensive) two year rodent bioassay. In this exercise we will use six different genes (biomarkers) from published studies to populate a gene list along with the reported direction of change in response to non-genotoxic carcinogens. The six genes with reported direction of change are:
A2m - α-2 macroglobulin (DOWN)
Cdc2A – cell division cycle 2 homolog A (UP)
Gsr – glutathione reductase (UP)
Myc – myelocytomatosis viral oncogene homolog (UP)
Tgfb1i4 – transforming growth factor beta 1 induced transcript 4 (DOWN)
Tp53 – tumor protein p53 (DOWN)

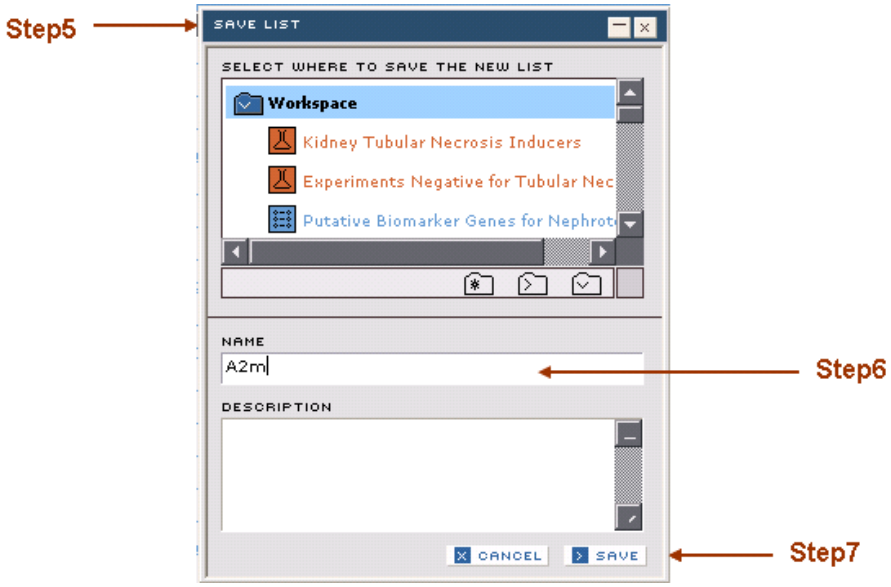
To find each gene in DrugMatrix we will do an advanced query:
Step 1 Click **ADVANCED** tab, and select **GENE** in the pull-down menu.
Step 2 Click the **ADD CRITERIA** button and select criteria indicated below.

Exercise 6 ~ Part 1 ~ Step 2



Step 3 Click **RUN QUERY**.
Step 4 Click on the **DISPLAY LIST** button to see the list of results.
This list returns the α-2 macroglobulin gene.
Step 5 Select **SAVE LIST** from the **MENU** button on the **LIST DISPLAY** panel.
Step 6 Name the list in the **SAVE LIST** dialog window as **A2m**.
Step 7 Click the **SAVE** button at the bottom of the **SAVE LIST** dialog window.

Exercise 6 ~ Part 1 ~ Steps 5 thru 7



To search for and save the other genes on the list:

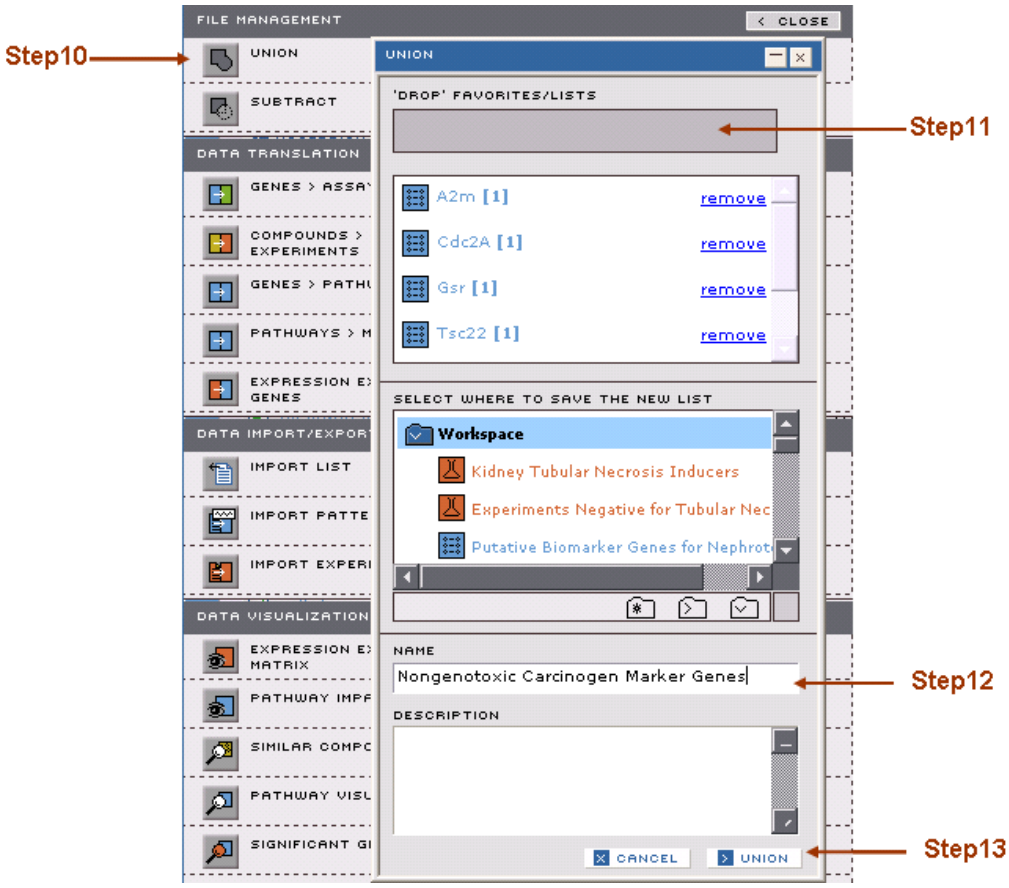
Step 8 Repeat Steps 1 to 7 for

-Cdc2A
-Gsr
-Myc
-Tgfb1i4
-Tp53

To combine all six genes into a gene list:

Step 9 Click the **TOOL BOX** in the **WORKSPACE**.
Step 10 Click to open the **UNION** dialog window under **FILE MANAGEMENT**.
Step 11 Drag all six genes to the **DROP FAVORITES/LISTS** field.
Step 12 Name the list as **Nongenotoxic Carcinogen Marker Genes**.
Step 13 Click on the **UNION** button at the bottom of the **UNION** dialog window.

Exercise 6 ~ Part 1 ~ Steps 10 thru 13



Exercise 6 ~ Part 2 ~ Create a Gene Expression Pattern for Non-genotoxic Carcinogens

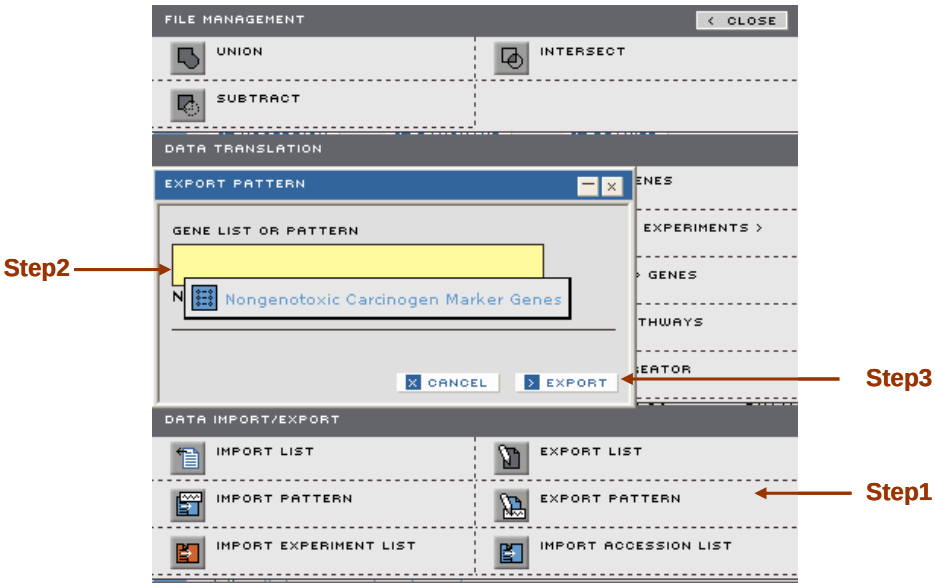
Once you have the gene list, you can manually create a gene expression pattern using the Pattern Import and Export tool.
First we need to export the Gene List to Excel in order to add the up or down regulation information.

Step 1 Click to open the **EXPORT PATTERN** dialog window under **DATA IMPORT/EXPORT.**

Step 2 Drag the Nongenotoxic Carcinogen Marker Genes list to the **GENE LIST OR PATTERN** field.

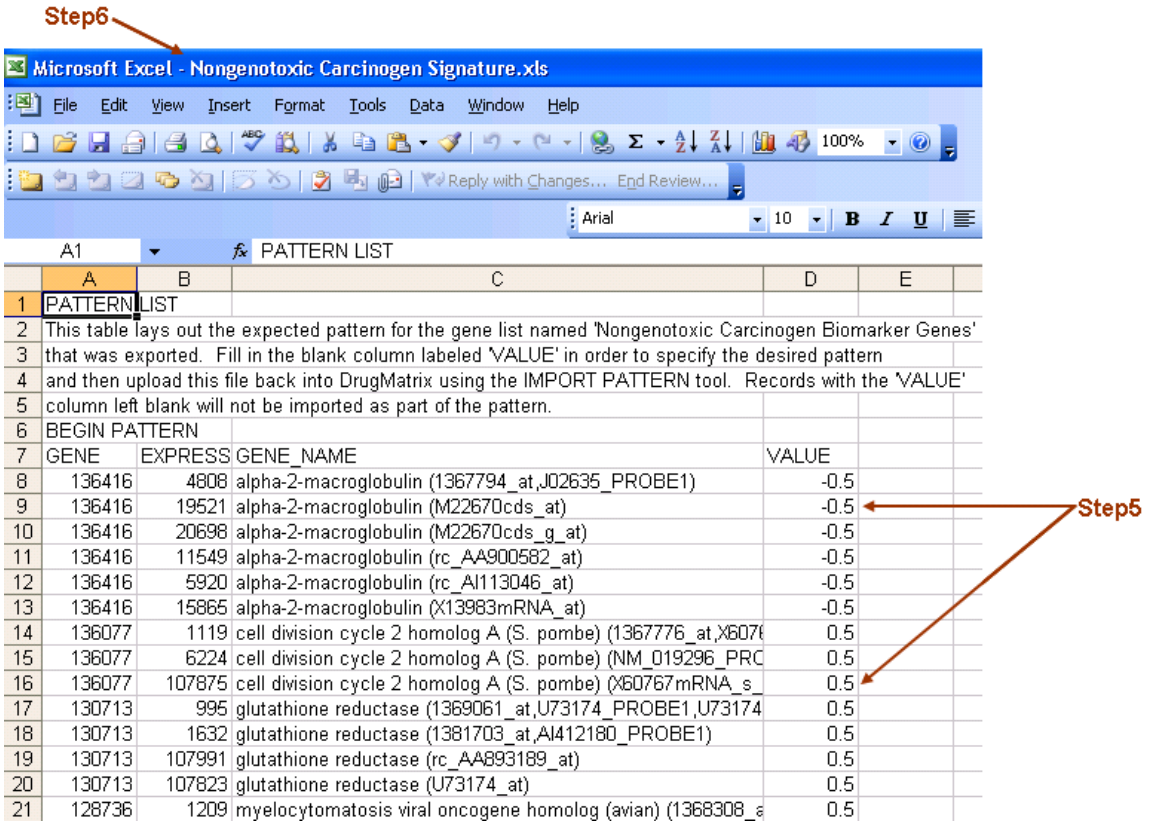
Step 3 Click on **EXPORT** button.

Exercise 6 ~ Part 2 ~ Steps 1 thru 3



Step 4 Save the file on your computer.
Step 5 Open the file in Excel. Following the instruction in the Excel file, enter under the **VALUE** column **"0.5"** for UP genes and **"-0.5"** for DOWN genes. The numbers entered here should be the best estimate of log ratios reflecting the magnitude and direction of gene changes derived from the literature. Note that genes may have more than one probe associated with them, and therefore more than one row.
Step 6 Save the file as a tab-delimited txt file with the name **Nongenotoxic Carcinogen Signature**.

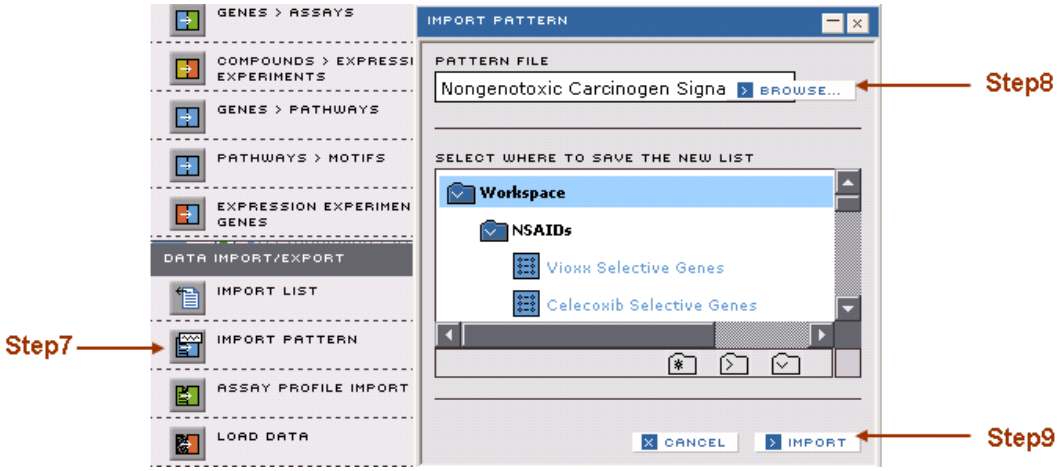
Exercise 6 ~ Part 2 ~ Steps 5 and 6



	A	B	C	D	E
1	PATTERN LIST				
2	This table lays out the expected pattern for the gene list named 'Nongenotoxic Carcinogen Biomarker Genes'				
3	that was exported. Fill in the blank column labeled 'VALUE' in order to specify the desired pattern				
4	and then upload this file back into DrugMatrix using the IMPORT PATTERN tool. Records with the 'VALUE'				
5	column left blank will not be imported as part of the pattern.				
6	BEGIN PATTERN				
7	GENE	EXPRESS	GENE_NAME	VALUE	
8	136416	4808	alpha-2-macroglobulin (1367794_at,J02635_PROBE1)	-0.5	
9	136416	19521	alpha-2-macroglobulin (M22670cds_at)	-0.5	
10	136416	20698	alpha-2-macroglobulin (M22670cds_g_at)	-0.5	
11	136416	11549	alpha-2-macroglobulin (rc_AA900582_at)	-0.5	
12	136416	5920	alpha-2-macroglobulin (rc_AI113046_at)	-0.5	
13	136416	15865	alpha-2-macroglobulin (X13983mRNA_at)	-0.5	
14	136077	1119	cell division cycle 2 homolog A (S. pombe) (1367776_at,X607	0.5	
15	136077	6224	cell division cycle 2 homolog A (S. pombe) (NM_019296_PRC	0.5	
16	136077	107875	cell division cycle 2 homolog A (S. pombe) (X60767mRNA_s_	0.5	
17	130713	995	glutathione reductase (1369061_at,U73174_PROBE1,U73174	0.5	
18	130713	1632	glutathione reductase (1381703_at,AI412180_PROBE1)	0.5	
19	130713	107991	glutathione reductase (rc_AA893189_at)	0.5	
20	130713	107823	glutathione reductase (U73174_at)	0.5	
21	128736	1209	myelocytomatosis viral oncogene homolog (avian) (1368308_a	0.5	

Step 7 Click to open the **IMPORT PATTERN** dialog window under **DATA IMPORT/EXPORT.**
Step 8 Click on **BROWSE** button to search and open the **Nongenotoxic Carcinogen Signature** file to the **IMPORT LIST** field.
Step 9 Click on **IMPORT** button.

Exercise 6 ~ Part 2 ~ Steps 7 thru 9

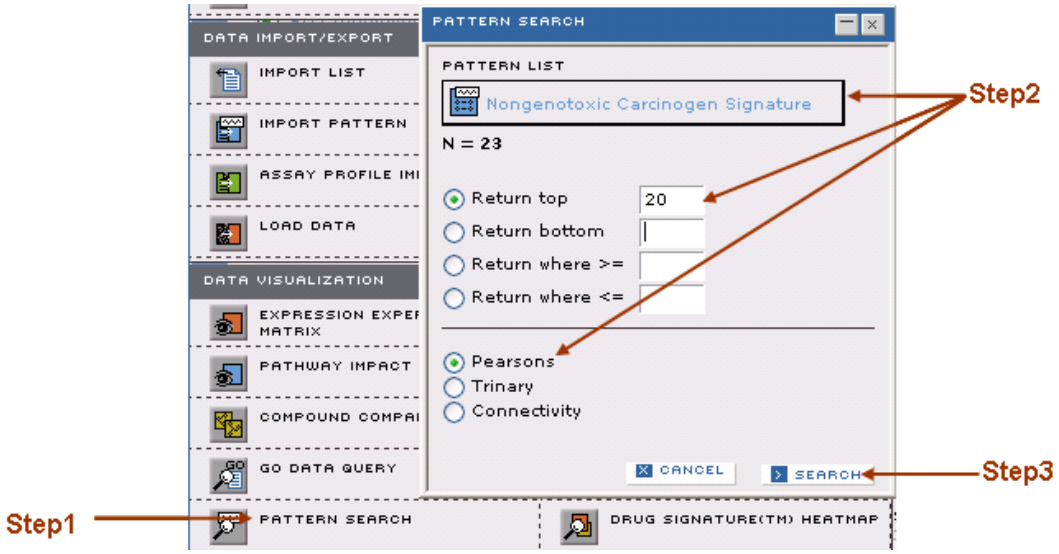


Exercise 6 ~ Part 3 ~ Identify Experiments that Match the Pattern Using Pattern Search Tool

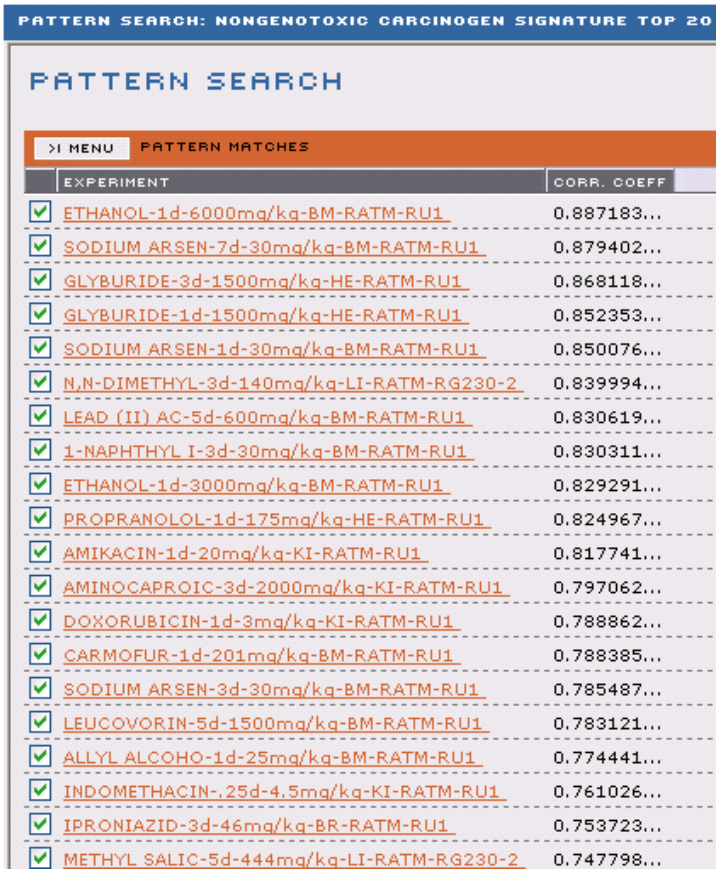
In this section we will take the **Nongenotoxic Carcinogen Signature** pattern to search in DrugMatrix database for treatments that modulate these genes similarly to known non-genotoxic carcinogens that were reported in the literature.

Step 1 Click on the **PATTERN SEARCH** tool in the **TOOL BOX** to open the **PATTERN SEARCH** dialog window.
Step 2 Drag the **Nongenotoxic Carcinogen Signature** into the **PATTERN LIST** entry box. Select the Return Top option and change the "Return top" number from 50 to 20. Select the "Pearsons" option as the method to calculate the match to the pattern.
Step 3 Click the **SEARCH** button at the bottom of the **PATTERN SEARCH** dialog window.

Exercise 6 ~ Part 3 ~ Steps 1 thru 3

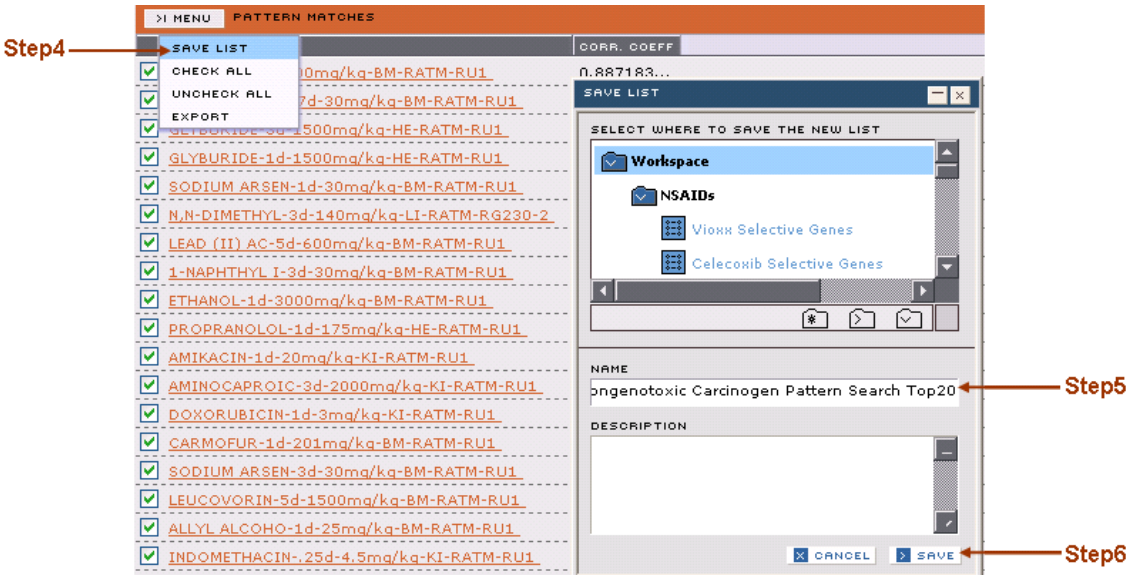


Exercise 6 ~ Part 3 ~ DATA VIEW on results from Steps 1 thru 3



Step 4 Select **SAVE LIST** from the drop-down **MENU** in the **PATTERN SEARCH** result window.
Step 5 Enter **Nongenotoxic Carcinogen Pattern Search Top 20** into the **NAME** entry box in the **SAVE LIST** dialog window.
Step 6 Click the **SAVE** button at the bottom of the **SAVE LIST** dialog window.

Exercise 6 ~ Part 3 ~ Steps 4 thru 6



Conclusions to Exercise 6

In this exercise, you learned to use the Pattern Import/Export tool to create pattern of gene expression data based on literature derived gene list, and use it to search for experiments that impact these marker genes in a similar pattern as the known non-genotoxic carcinogens.

Exercise 7 Evaluation of a Putative Biomarker of Nephrotoxicity

This exercise will show you how to apply the simple text search and advanced query tool to extract information from the database and how to construct lists of experiments in order to identify gene markers for nephrotoxicity.

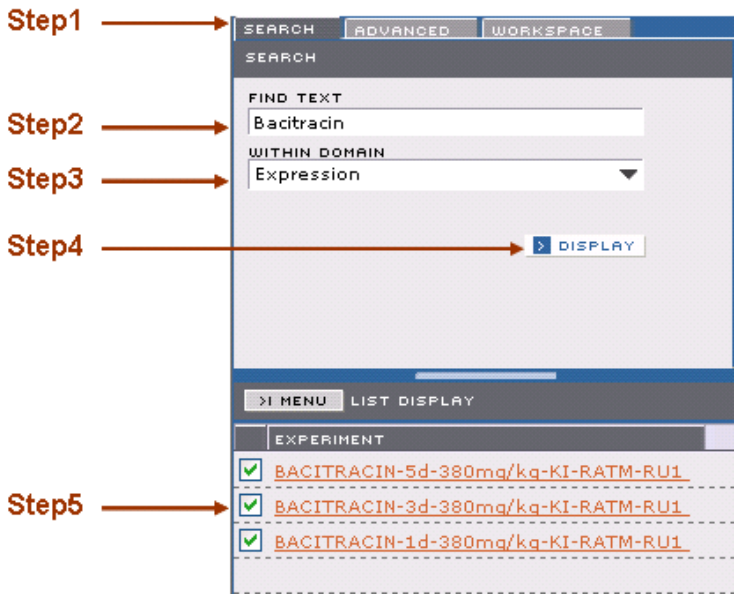
You will learn to:
• Formulate a simple query and obtain background information about a compound.
• Extract and save lists of positive and negative class experiments for nephrotoxicity using the advanced query tools.
• Identify biomarker genes for nephrotoxicity with these lists using the **SIGNIFICANT GENE FINDER** tool and evaluate their performance using the **EXPRESSION EXPERIMENT MATRIX** tool.

Exercise 7 ~ Part 1 ~ Simple Search of a Nephrotoxicant

Bacitracin is a known nephrotoxicant. This section will demonstrate how to extract information from the database on Bacitracin.

Begin with a simple search for information on the compound Bacitracin.
Step 1 Select the **SEARCH** tab.
Step 2 In the **FIND TEXT** field, enter **Bacitracin**.
Step 3 Use the **WITHIN DOMAIN** menu to select the **EXPRESSION** domain for the query.
Step 4 Click the **DISPLAY** button.
This will return three experiments on Bacitracin.
Step 5 Click on **BACITRACIN-3d-380mg/kg-KI-RATM-RU1** to populate the Expression Domain Report with all the information available in DrugMatrix about this drug.

Exercise 7 ~ Part 1 ~ Steps 1 thru 5



The **SIMILAR** panel is the first panel displayed in this report. It provides a list of experiments similar in transcriptional response to the queried experiment based on Pearson correlation.

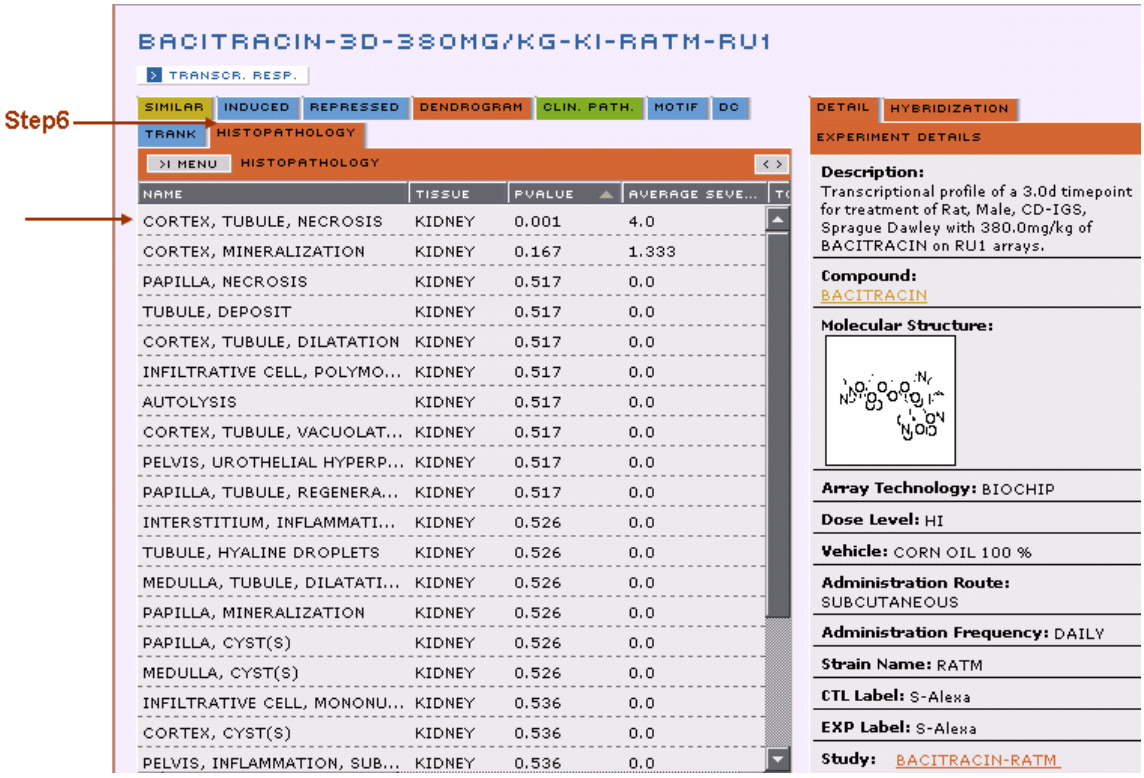
Step 6 Click on the **HISTOPATHOLOGY** panel to display a list of pathology findings associated with this experiment. Notice the name **CORTEX TUBULE NECROSIS** appears on top of this list with an average severity of 4.0 and a p-value of 0.001.

Step 7 Click on **CLIN. PATH.** panel to display clinical chemistry data collected on this experiment.

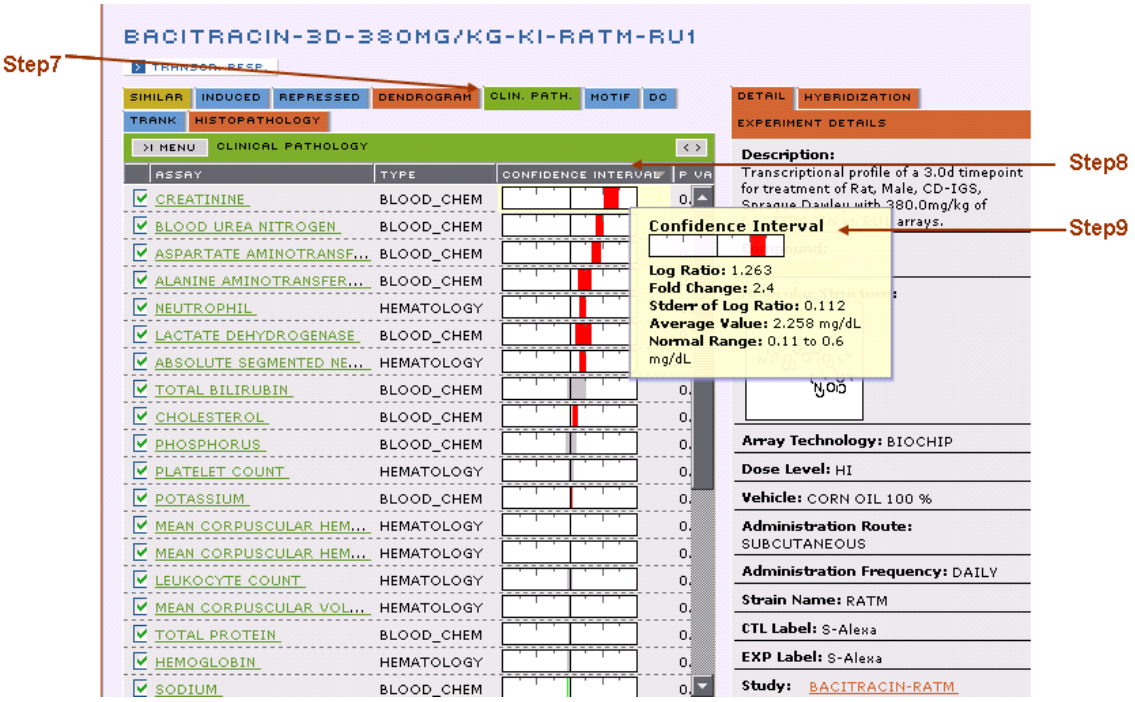
Step 8 Click on the **CONFIDENCE INTERVAL** tab to sort clinical chemistry data by fold change. A second click will sort from highest to lowest value.

Step 9 Mouse over the confidence interval symbol to display fold change, average value, etc. Notice that **CREATININE** level has been increased by 2.4-fold by this treatment.

Exercise 7 ~ Part 1 ~ Step 6



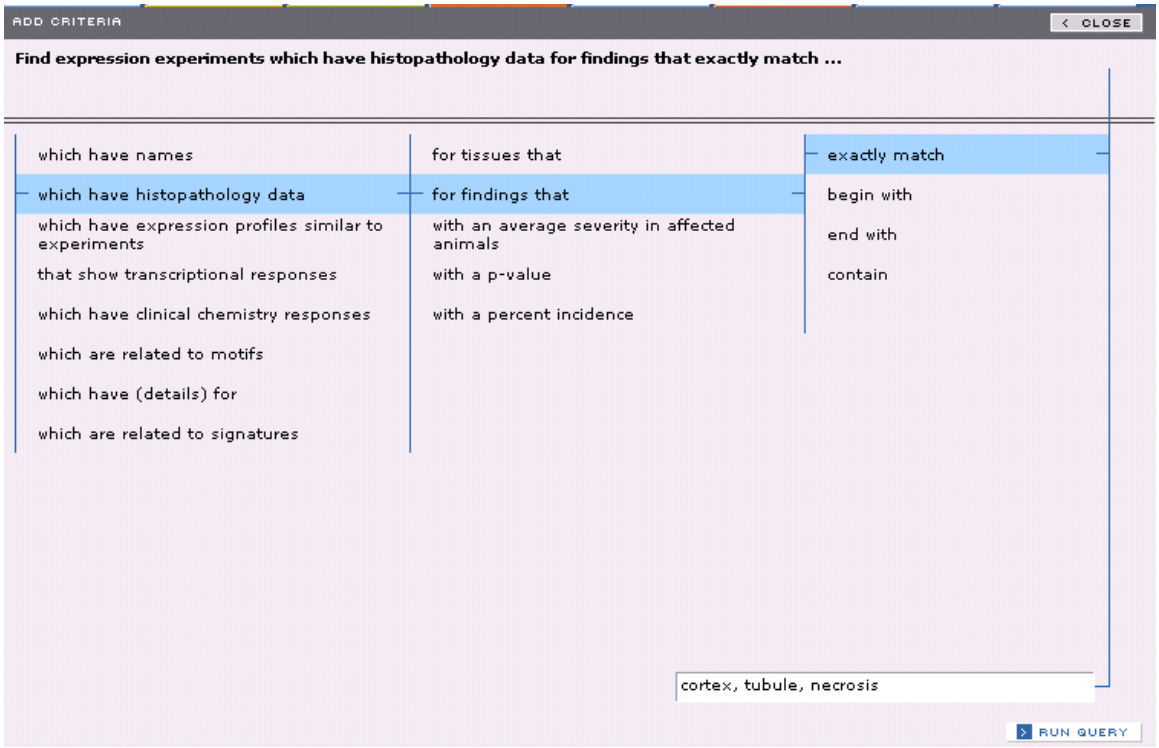
Exercise 7 ~ Part 1 ~ Steps 7 thru 9



Exercise 7 ~ Part 2 ~ Build an Expression Experiment Advanced Query

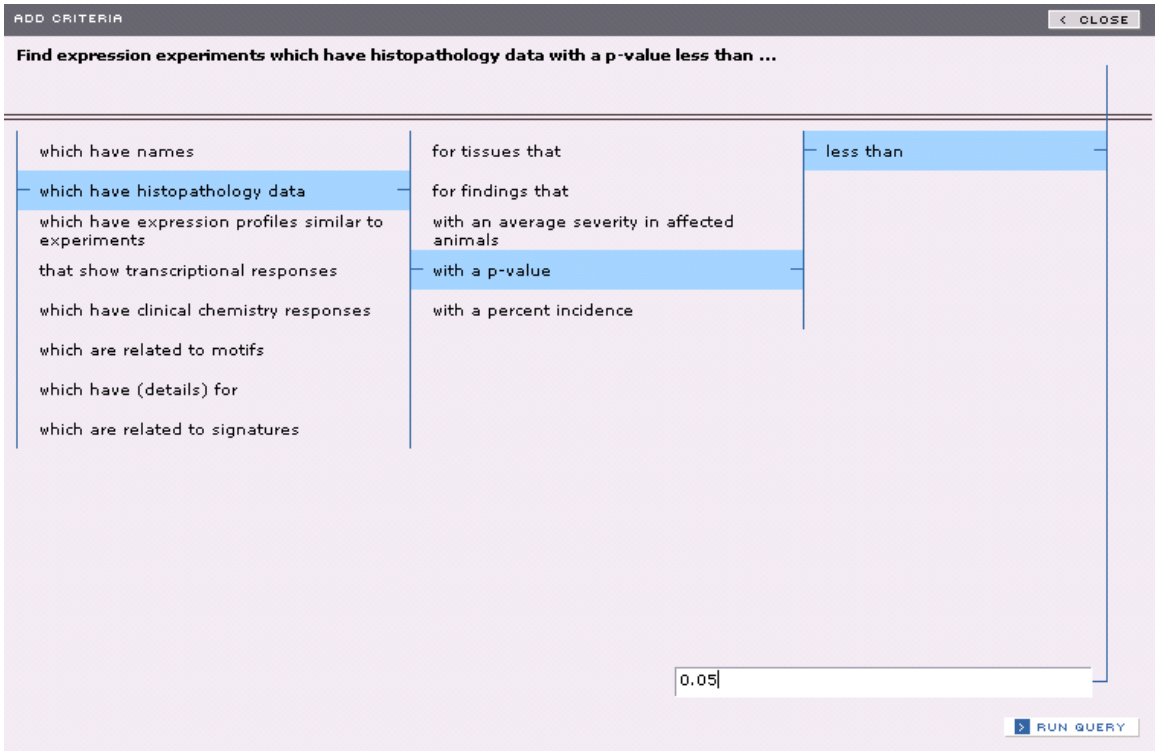
In this exercise you will use the **ADVANCED** search tool to identify kidney experiments in DrugMatrix that induce cortex tubular necrosis.
Step 1 Click on the **ADVANCED** search tab and click the **CLEAR ALL** button to remove all previous criteria.
Step 2 Select **EXPRESSION** from the pull-down menu.
Step 3 Click on **ADD CRITERIA**.
Step 4 In the **ADD CRITERIA** panel build the query by clicking on the criteria indicated below. Then type the finding “cortex, tubule, necrosis” into the text box. Note that the syntax must match the syntax from the Histopathology Panel in Part 1 Step 6 above.

Exercise 7 ~ Part 2 ~ Step 4



The number of experiments that fit these criteria is shown within the Advanced Search Frame. In this case the search returns N=422. We can now narrow the search in several ways by adding criteria to the query. To limit the search to experiments that induce cortex tubule necrosis above a specific statistical threshold:
Step 5 Click **ADD CRITERIA**.
Step 6 Add the criteria indicated in cyan below.
Step 7 Type **0.05** into the entry field.
Step 8 Click **RUN QUERY**.

Exercise 7 ~ Part 2 ~ Steps 6 thru 8

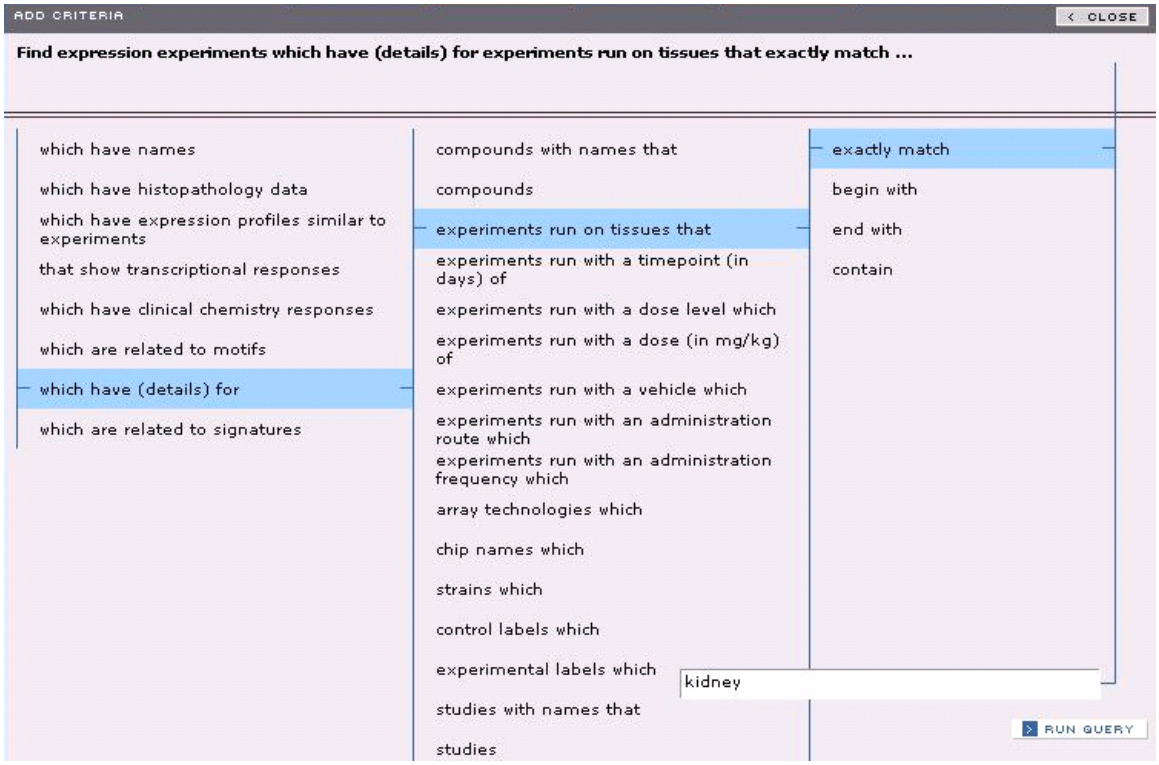


These additional criteria have reduced the returned experiments to N=28.

We will now further restrict the experiments to only those that have been run in the kidney.

Step 9 Click **ADD CRITERIA**.
Step 10 Add the criteria indicated in cyan below and click **RUN QUERY**.

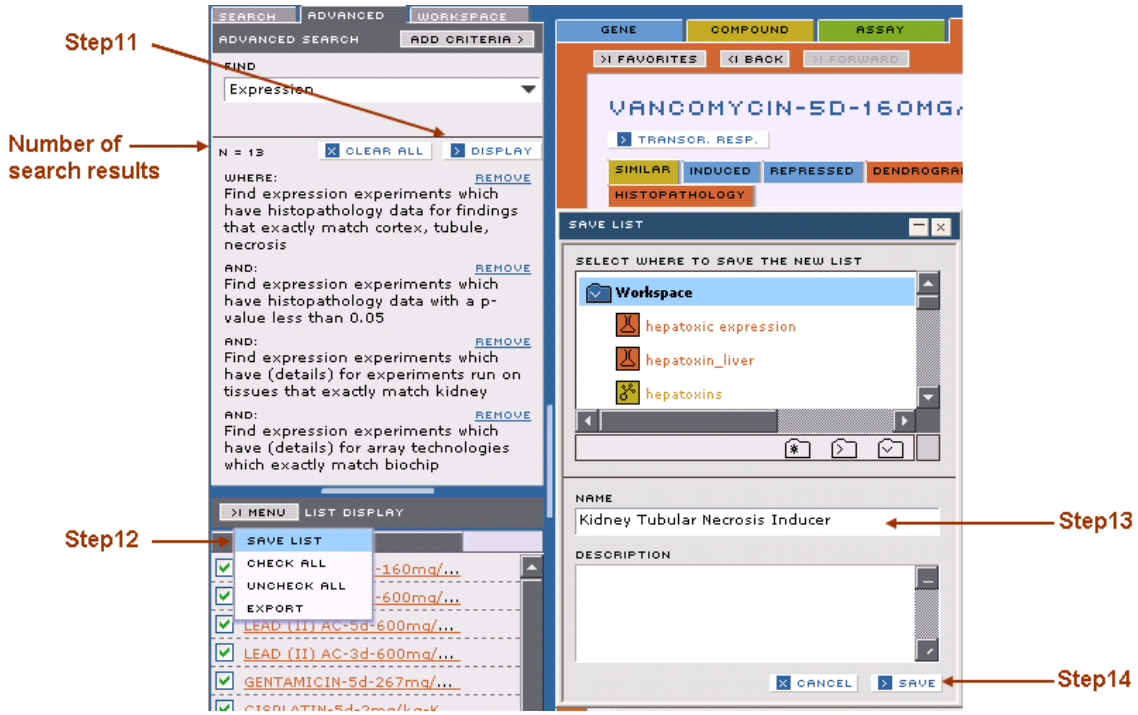
Exercise 7 ~ Part 2 ~ Steps 9 and 10



When the search is complete, the **ADVANCED SEARCH** panel shows all of the criteria by which you have searched. You can remove any criteria by click on the **REMOVE** button. It also indicates the number of expression experiments you have identified using these criteria. In this case N=13. Now that we have refined the query, we will display the results in the **LIST DISPLAY** panel and save the results to a list:

Step 11 Click the **DISPLAY** button to display this list of experiments below the panel.
Step 12 Select **SAVE LIST** from the **MENU** button on the **LIST DISPLAY** panel.
Step 13 Name the list in the **SAVE LIST** dialog window as **Kidney Tubular Necrosis Inducers**. Note that additional information can be added in the **DESCRIPTION** Box
Step 14 Click the **SAVE** button at the bottom of the **SAVE LIST** dialog window.

Exercise 7 ~ Part 2 ~ Steps 11 thru 14

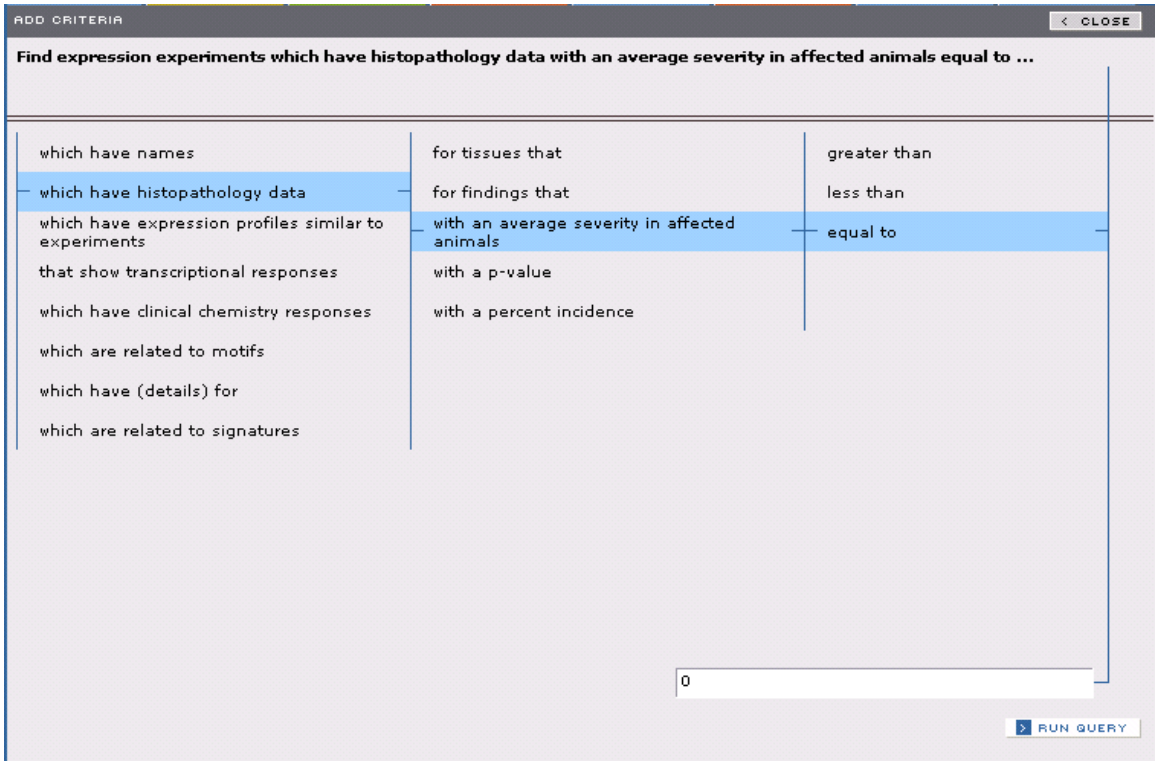


We have just created the positive test set of experiments and we now need to select a set of experiments that are negative for tubular necrosis as reference set in order to identify biomarkers of nephrotoxicity. To do this, we just need to modify the current search criteria by changing criteria number 2.

Step 15 Click **REMOVE** button for criteria number 2.

Step 16 Click **ADD CRITERIA**, type 0 as the average severity score in the text box as indicated below and click **RUN QUERY**.

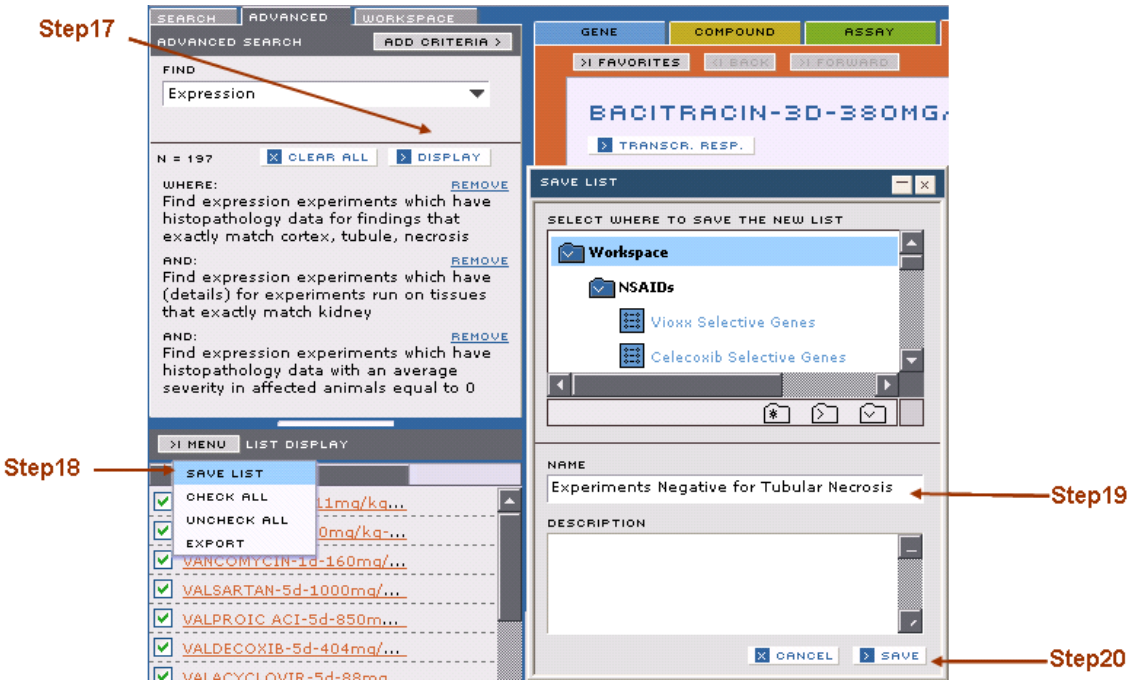
Exercise 7 ~ Part 2 ~ Steps 15 and 16



The Advanced Search Box should now display N=197 reflecting the experiments run in kidney with an average severity score=0 (i.e. no histopathological sign of tubular necrosis).

Step 17 Click the **DISPLAY** button to display this list of experiments below the panel.
Step 18 Select **SAVE LIST** from the **MENU** button on the **LIST DISPLAY** panel.
Step 19 Name the list in the **SAVE LIST** dialog window as **Experiments Negative for Tubular Necrosis**.
Step 20 Click the **SAVE** button at the bottom of the **SAVE LIST** dialog window.

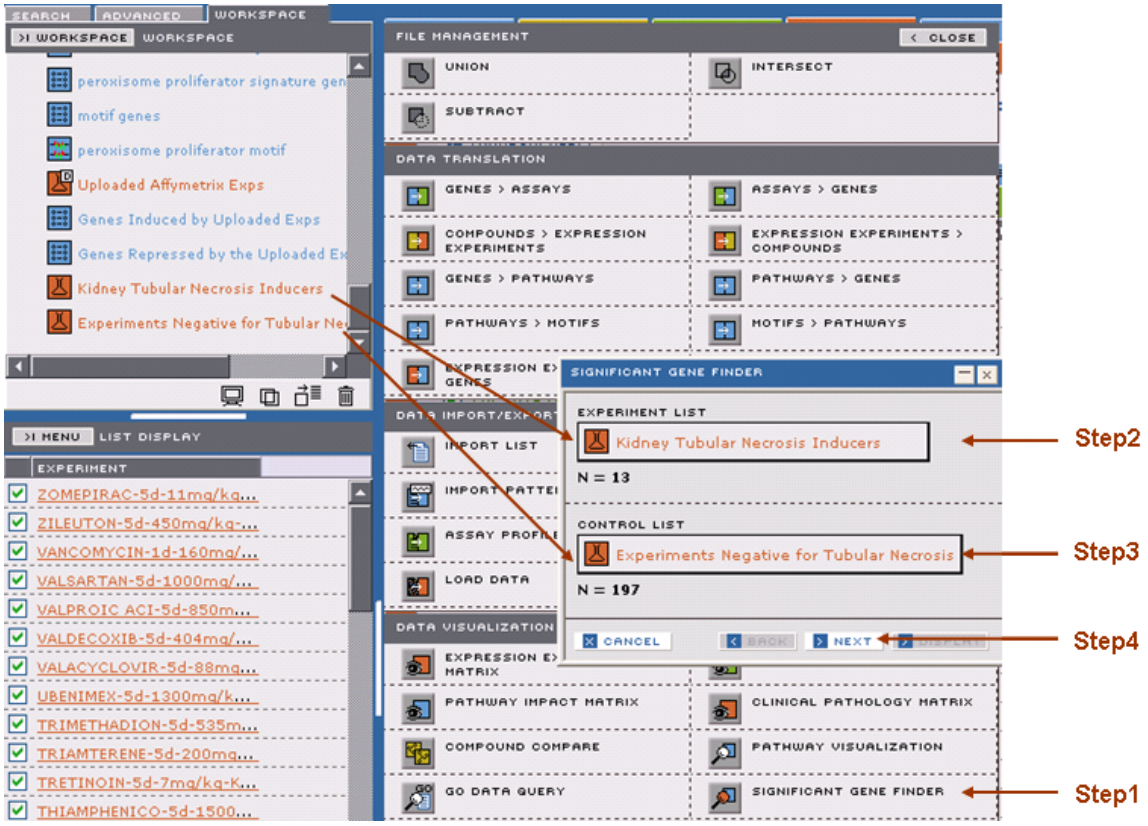
Exercise 7 ~ Part 2 ~ Steps 17 thru 20



Exercise 7 ~ Part 3 ~ Find Significant Genes in an Experiment List

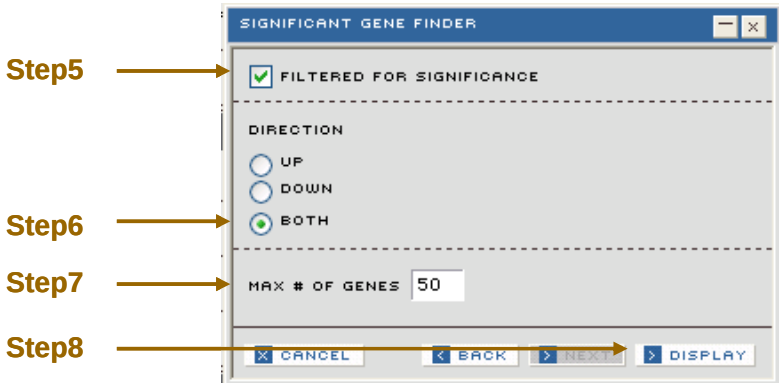
We will now compare these two lists of experiments in order to identify putative biomarker genes that are significantly changed only in the positive class of experiments.
We will use the **SIGNIFICANT GENE FINDER** tool to identify biomarker genes specific for the test set of experiments that induced kidney tubular necrosis.
Step 1 Click on the **SIGNIFICANT GENE FINDER** tool in the **TOOLBOX**.
Step 2 Drag the experiment list **Kidney Tubular Necrosis Inducers** into the **EXPERIMENT LIST** field of the **SIGNIFICANT GENE FINDER** dialog window.
Step 3 Drag the experiment list **Experiments Negative for Tubular Necrosis** into the **CONTROL LIST** field of the **SIGNIFICANT GENE FINDER** dialog.
Step 4 Click on the **NEXT** button at the bottom of the **SIGNIFICANT GENE FINDER** dialog window.

Exercise 7 ~ Part 3 ~ Steps 1 thru 4



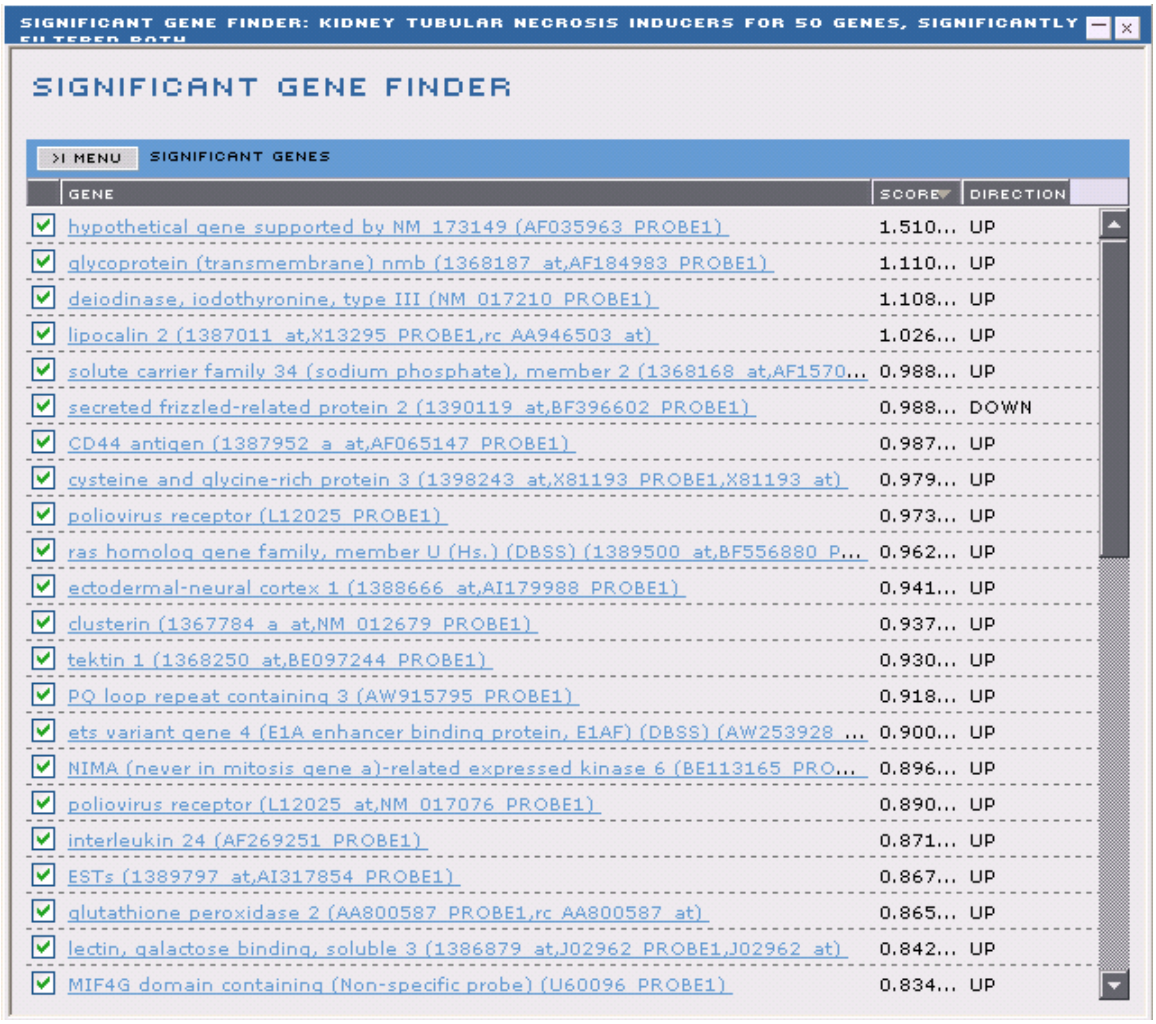
Step 5 Leave the **FILTERED FOR SIGNIFICANCE** check box checked, so that only significantly modulated genes are included in the calculation.
Step 6 You can choose **UP**, **DOWN**, or **BOTH** to find genes that are either induced, repressed or both. (For this exercise select **BOTH**).
Step 7 Leave the **MAX # OF GENES** at 50. This limits the display to 50 genes.
Step 8 Click the **DISPLAY** button on the bottom of this dialog window.

Exercise 7 ~ Part 3 ~ Steps 5 thru 8



The result is a list of genes with associated score and direction of change. The score indicates how consistently the expression of the gene differs between the two set of experiments.
Step 9 Click on the **SCORE** tab to sort the genes by their associated scores from the highest to lowest.

Exercise 7 ~ Part 3 ~ DATAVIEW resulting from Steps 1 thru 9



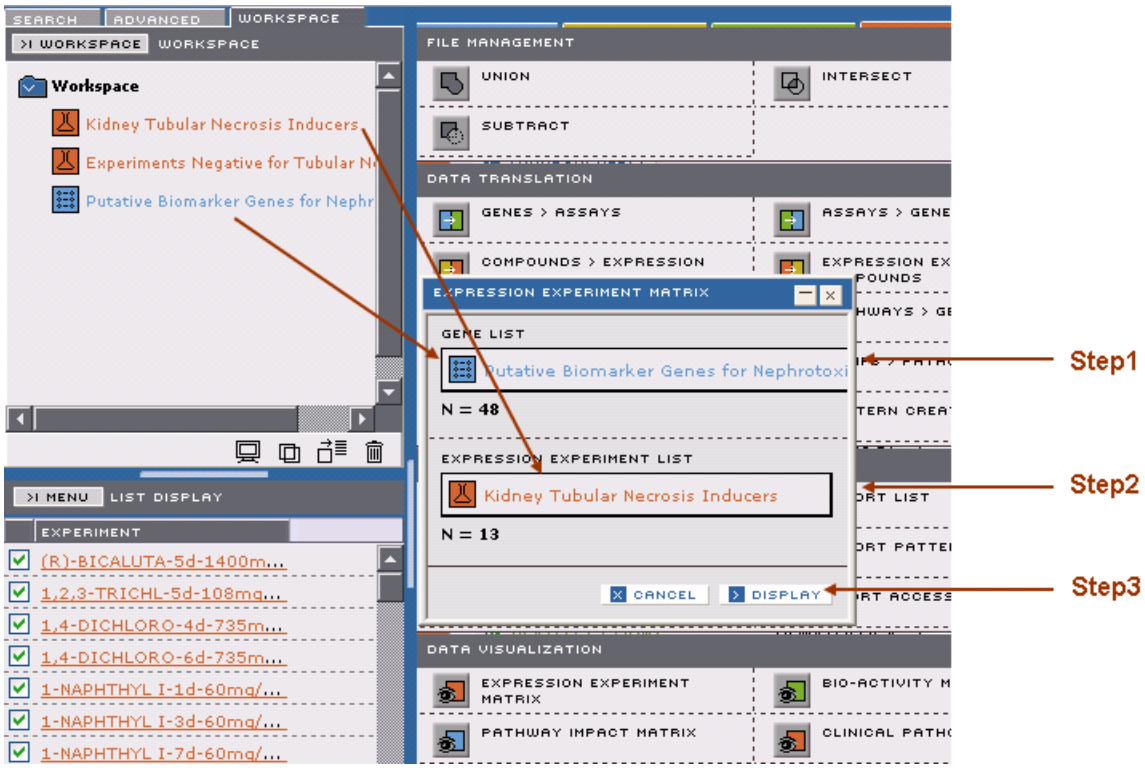
In order to work further with this list of putative biomarker genes, we need to save it as a list.
Step 9 Select **SAVE LIST** from the **MENU** button on the **LIST DISPLAY** panel.
Step 10 Name the list in the **SAVE LIST** dialog window as **Putative Biomarker Genes for Nephrotoxicity**.
Step 11 Click the **SAVE** button at the bottom of the **SAVE LIST** dialog window.

Exercise 7 ~ Part 4 ~ Evaluate Biomarker Genes Using Expression Experiment Matrix tool

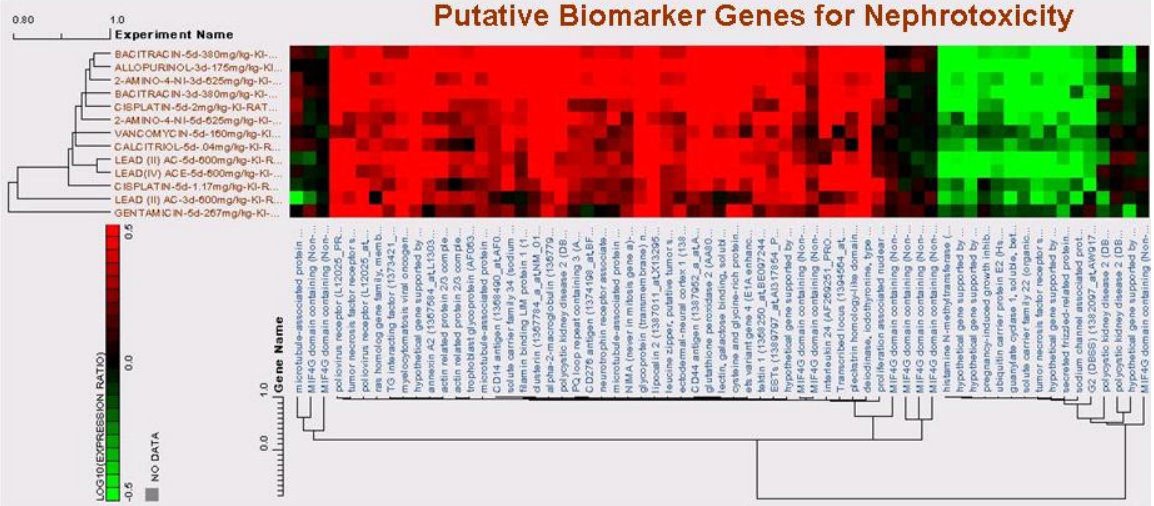
Now that we have identified the putative biomarker genes for nephrotoxicity, we can evaluate the performance of these genes as biomarkers by using the Expression Experiment Matrix tool to generate a 2-dimensional hierarchical clustering matrix that will display the impact of treatments with compounds positive for kidney tubular necrosis on those selected genes.

Step 1 Drag the **Putative Biomarker Genes for Nephrotoxicity** list to the **EXPRESSION EXPERIMENT MATRIX** tool.
Step 2 Drag the **Kidney Tubular Necrosis Inducers** experiment list to the **EXPERIMENT LIST** field of the **EXPRESSION EXPERIMENT MATRIX** tool.
Step 3 Click on the **DISPLAY** button at the bottom of the **EXPRESSION EXPERIMENT MATRIX** dialog window.

Exercise 7 ~ Part 4 ~ Steps 1 thru 3



Expression Experiment Matrix output for Putative Biomarker Genes for Nephrotoxicity gene list and Kidney Tubular Necrosis Inducers experiment list

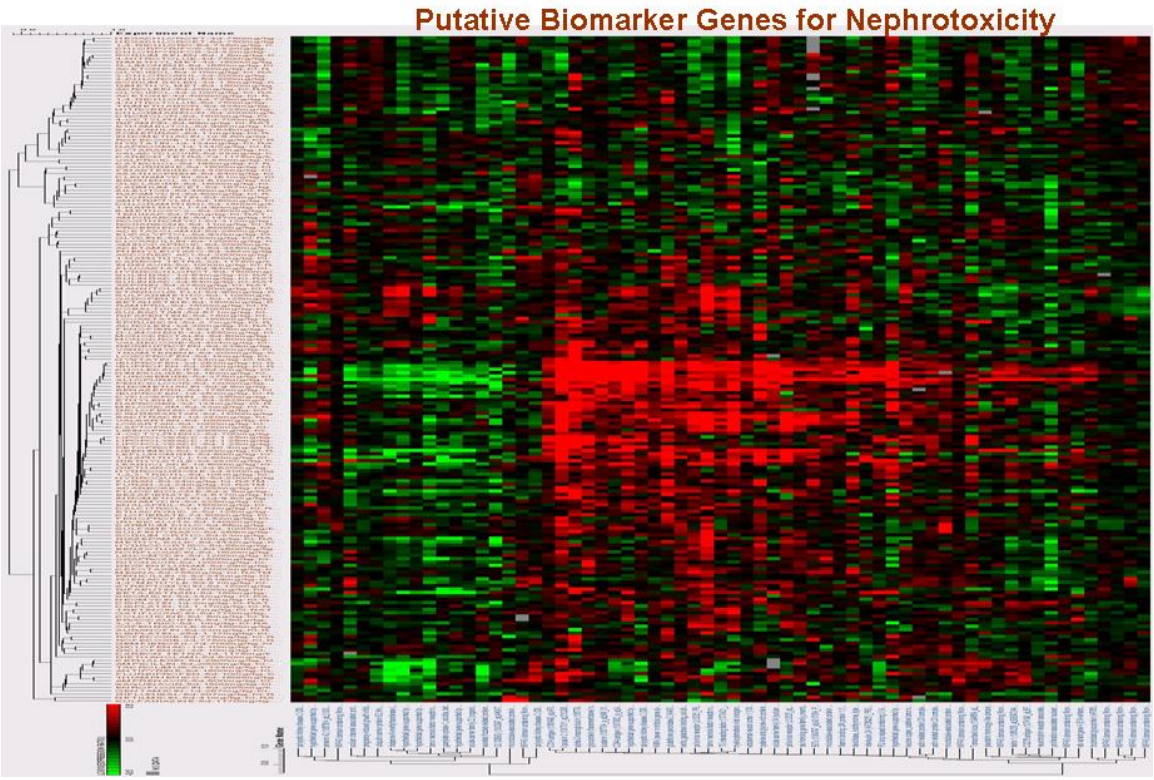


As expected, compounds that induce kidney tubular necrosis strongly and consistently induced or repressed these biomarker genes.

We can then create a hierarchical clustering analysis that compares induction or repression of these same 50 biomarker genes with treatments in the negative reference class of experiments that we created.

Step 4 Drag the **Putative Biomarker Genes for Nephrotoxicity** list to the **EXPRESSION EXPERIMENT MATRIX** tool.
Step 5 Drag the **Experiments Negative for Tubular Necrosis** experiment list to the **EXPERIMENT LIST** field of the **EXPRESSION EXPERIMENT MATRIX** tool.
Step 6 Click on the **DISPLAY** button at the bottom of the **EXPRESSION EXPERIMENT MATRIX** dialog window.

Expression Experiment Matrix output for Putative Biomarker Genes for Nephrotoxicity gene list and Experiments Negative for Tubular Necrosis experiment list



The resulting heat map illustrates that the experiments that are negative for kidney tubular necrosis impact the 50 biomarker genes in a manner that appears random and inconsistent compared to the positive class.

Conclusions to Exercise 7

In this exercise you used the simple and advanced search features to populate experiment lists, and used the **SIGNIFICANT GENE FINDER** feature to identify putative biomarker genes for nephrotoxicity.
You then used the **EXPRESSION EXPERIMENT MATRIX** tool to visualize the performance of the identified biomarker genes across positive and negative class treatments.

Exercise 8 Compare Fibrates and Statins Using Drug Signatures

A Drug Signature is a pattern of gene expression responses for a set of genes that can be used to classify a set of experiments reliably compared to the experiment population, or a defined set of control experiments.
Signatures can be used to find compounds that are similar to an unknown, more reliably than by using the full gene expression pattern. The content of a signature can also be used to provide understanding of the role key genes play in characterizing of the experiments in the set. In this exercise, we will explore Drug Signatures in the liver and compare signature matches between fibrates and statins.

You will learn to:
• Use an **ADVANCE** query to extract appropriate Drug Signatures.
• Use the **DRUG SIGNATURE HISTOGRAM and DRUG SIGNATURE HEATMAP** tools to compare the matches of fibrates and statins to Drug Signatures in liver.
• Use the **EXPRESSION EXPERIMENT MATRIX** tool to examine the impact of selected experiments on Drug Signature genes.
• Use the **MOTIF** technology to dissect biological meaning for the Drug Signature matches.

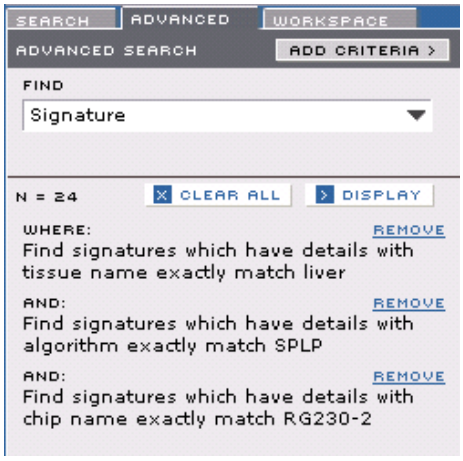
Exercise 8 ~ Part 1 ~ Construct Appropriate Drug Signature List

The following exercise will show you how to build a signature list using the advance query tool containing Drug Signatures in liver, derived from Affymetrix rat whole genome RG230-2 array data and using SPLP algorithm.

Step 1 Click on the **ADVANCED SEARCH** tab and click the **CLEAR ALL** button to remove all previous criteria.
Step 2 Select **SIGNATURE** from the drop down menu.
Step 3 Click **ADD CRITERIA** and select
-which have details with
-tissue name
-exactly match
-[liver]
Step 4 Click the **RUN QUERY** button at the bottom of the **ADD CRITERIA** panel
Step 5 Click on **ADD CRITERIA** (this is equivalent to the Boolean operator "**AND**") and select
-which have details with
-algorithm
-exactly match
-[SPLP]
Step 6 Click the **RUN QUERY** button at the bottom of the **ADD CRITERIA** panel
Step 7 Click on **ADD CRITERIA**
-which have details with
-chip name

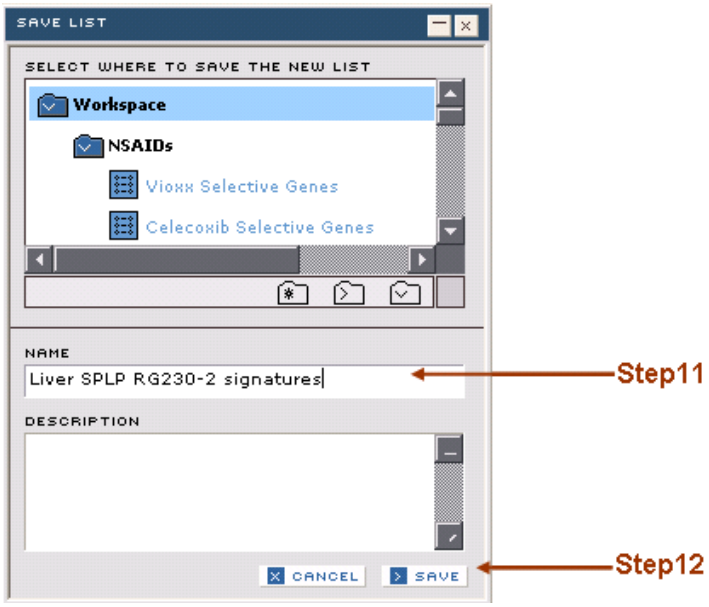
-exactly match
-[RG230-2]
Step 8 Click the **RUN QUERY** button
This search results in a total of 24 Drug Signatures.

Search results from Exercise 8 ~ Part 1 ~ Steps 1 thru 8



Step 9 Click on the **DISPLAY** button.
Step 10 Select **SAVE LIST** from the **MENU** button on the **LIST DISPLAY** panel.
Step 11 Name the list in the **SAVE LIST** dialog window, for example as
Liver SPLP RG230-2 signatures
Step 12 Click the **SAVE** button at the bottom of the **SAVE LIST** dialog window.

Exercise 8 ~ Part 1 ~ Steps 11 and 12



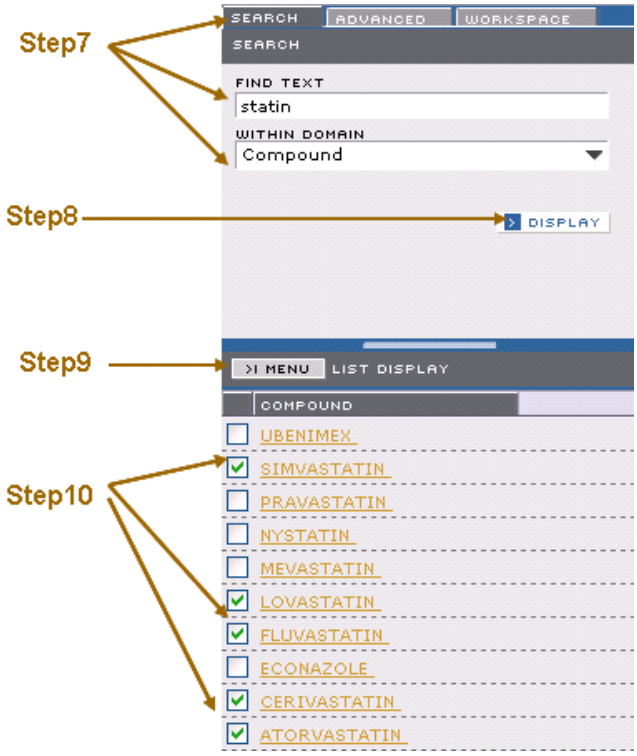
Exercise 8 ~ Part 2 ~ Create a Compound List

Step 1 Select the **SEARCH** tab in the **SEARCH - WORKSPACE** area.
Step 2 Type the letters **fibr** in the **FIND TEXT** field.
Step 3 Choose **COMPOUND** from the **WITHIN DOMAIN** drop-down menu.
Step 4 Click on **DISPLAY** button

Exercise 8 ~ Part 2 ~ Steps 1 thru 4

Step 5 Choose **SAVE LIST** from the **MENU** button on the **LIST DISPLAY** panel
Step 6 Save the list as **Fibrates**
Step 7 Next do a **SIMPLE SEARCH** in the **COMPOUND** domain for **statin**
Step 8 Click the **DISPLAY** button
Step 9 Choose **UNCHECK ALL** from the **MENU** button on the **LIST DISPLAY** panel
Step 10 Select the five compounds shown selected below

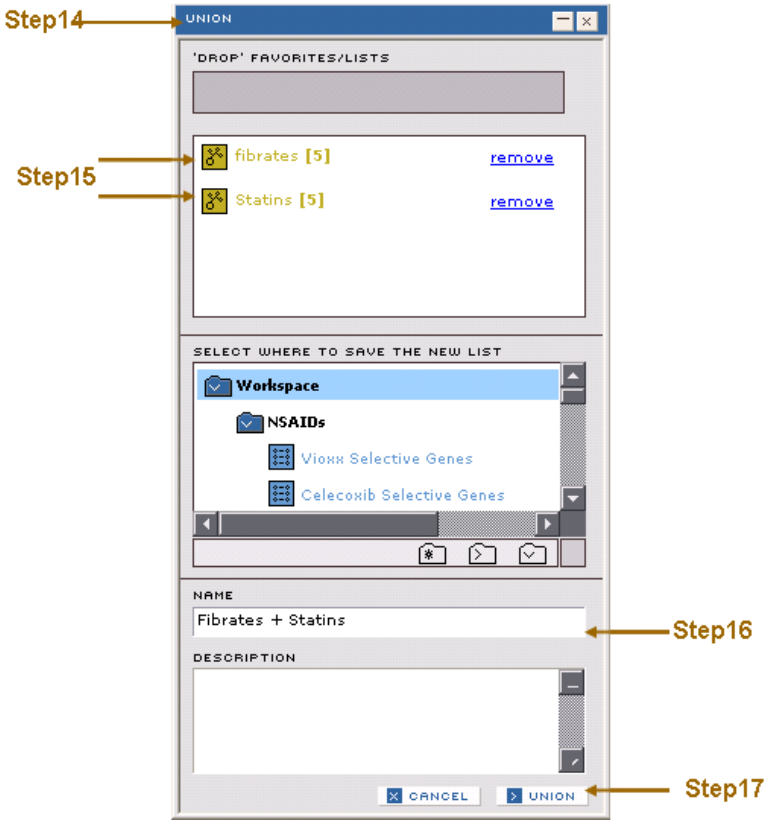
Exercise 8 ~ Part 2 ~ Steps 7 thru 10



Step 11 Choose **SAVE LIST** from the **MENU** button on the **LIST DISPLAY** panel
Step 12 In the **SAVE LIST** dialog, save the list as **Statins**

Next, use the **UNION** tool to combine the two compound lists you have just created.
Step 13 Click on the **WORKSPACE** tab and click on the **TOOLBOX** icon
Step 14 Click on the **UNION** tool in the **LIST EDIT** area of the **TOOLBOX**
Step 15 Drag each of the two lists **Fibrates** and **Statins** into the **'DROP' FAVORITE/LISTS** box
Step 16 Name the list **Fibrates + Statins**
Step 17 Click the **UNION** button at the bottom of the **UNION** dialog window

Exercise 8 ~ Part 2 ~ Steps 14 thru 17



At this stage we have the list **Fibrates + Statins** containing two classes of compounds with different mechanistic actions which are used therapeutically to lower cholesterol. The fibrates do this by increasing fatty acid oxidation, while the statins do this by blocking cholesterol production.

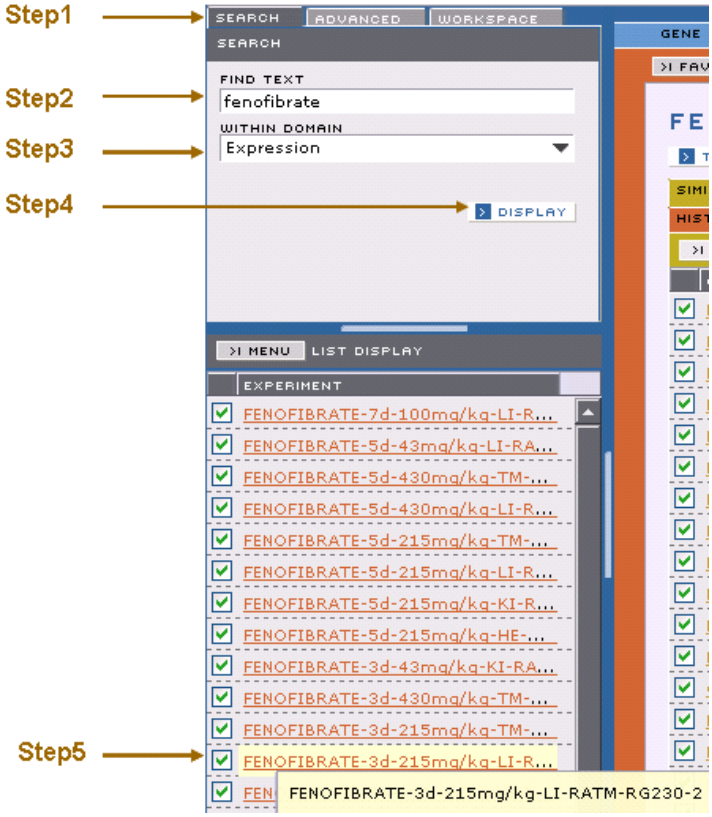
Exercise 8 ~ Part 3 ~ Generate a Drug Signature Histogram

The previous exercise showed you how to generate a simple "T-rank" signature. The resulting signature could be viewed to understand the genes and their expression pattern.
This next exercise will show you how to analyze an expression experiment using Drug Signatures derived using the high performance **SPLP** classification algorithm.

Step 1 Click on the **SEARCH** tab.
Step 2 Type **fenofibrate** in the **FIND TEXT** field.
Step 3 Choose **EXPRESSION** from the **WITHIN DOMAIN** select menu.

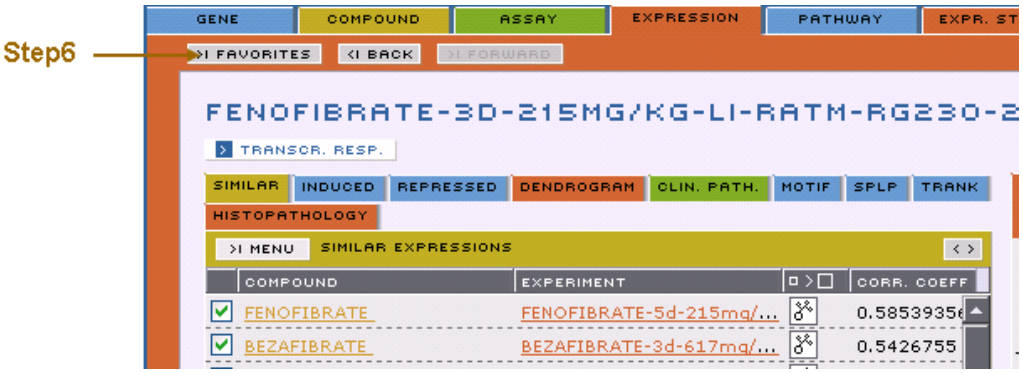
Step 4 Click on the **DISPLAY** button.
Step 5 Scroll down and click on **FENOFIBRATE-3D-215MG/KG-LI-RATM-RG230-2** in the **EXPERIMENT** list.

Exercise 8 ~ Part 3 ~ Steps 1 thru 5



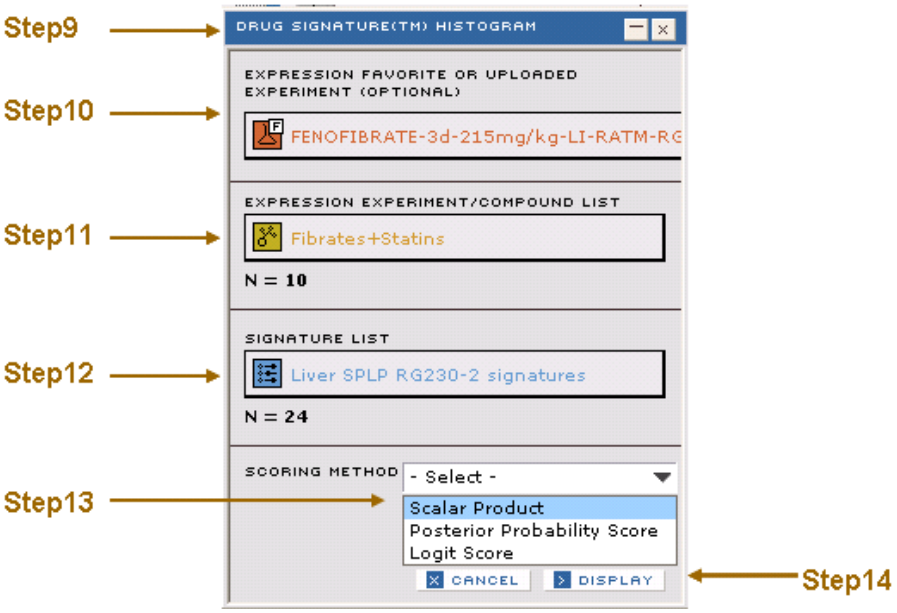
This will populate an **EXPRESSION** Domain Report with information about the experiment **FENOFIBRATE-3D-215MG/KG-LI-RATM-RG230-2**.
Step 6 On the **EXPRESSION** Domain Report panel, click on the **FAVORITES** button and choose **ADD TO FAVORITES** from the drop-down menu.
This will add **FENOFIBRATE-3D-215MG/KG-LI-RATM-RG230-2** to your **FAVORITES** list and it will also be available from the **WORKSPACE** area.

Exercise 8 ~ Part 3 ~ Step 6



Step 7 Click the **WORKSPACE** tab.
Step 8 Click on the **TOOLBOX** icon to open the **TOOLBOX**.
Step 9 Click the **DRUG SIGNATURE(TM) HISTOGRAM** tool from the **TOOLBOX**.
Step 10 Drag the list **FENOFIBRATE-3D-215MG/KG-LI-RATM-RG230-2** from the **WORKSPACE** area into the **EXPRESSION EXPERIMENT DATA/FAVORITE** field of the **DRUG SIGNATURE(TM) HISTOGRAM** dialog window.
Step 11 Drag the list **Fibrates + Statins** into the **COMPOUND/EXPRESSION EXPERIMENT LIST** field of the **DRUG SIGNATURE(TM) HISTORGRAM** dialog window.
Step 12 Drag the list **Liver SPLP RG230-2 signatures** into the **SIGNATURE LIST** field of the **DRUG SIGNATURE(TM) HISTOGRAM** dialog window.
Step 13 Select **SCALAR PRODUCT** from the **SCORING METHOD** drop-down menu.
Step 14 Click on the **DISPLAY** button at the bottom of this dialog window.

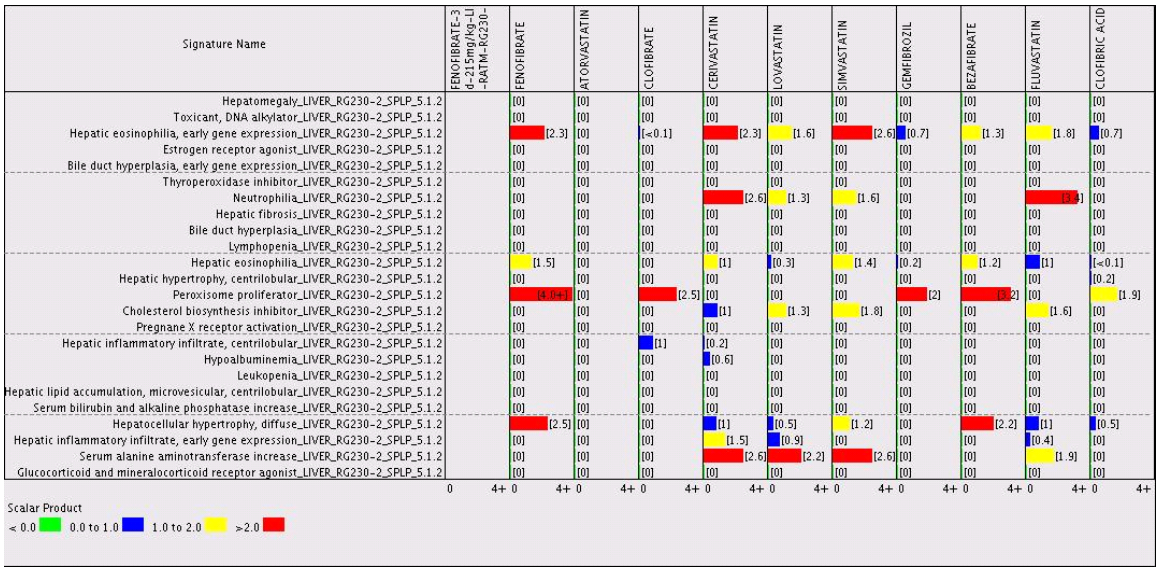
Exercise 8 ~ Part 3 ~ Steps 9 thru 14



This will display a table with Drug Signature names in the first row and compound names in the first column. The matches to Drug Signatures are represented by

the scalar product score as indicated by the color scale. Scalar product score less than 0 indicates not matching the signature, and scalar product score higher than 0 but less than 1 is considered weak match. Compound strongly match a Drug Signature will have a scalar product score higher than 1, which indicates a high probability that the test compound has the feature exhibited in the positive class of the training set.

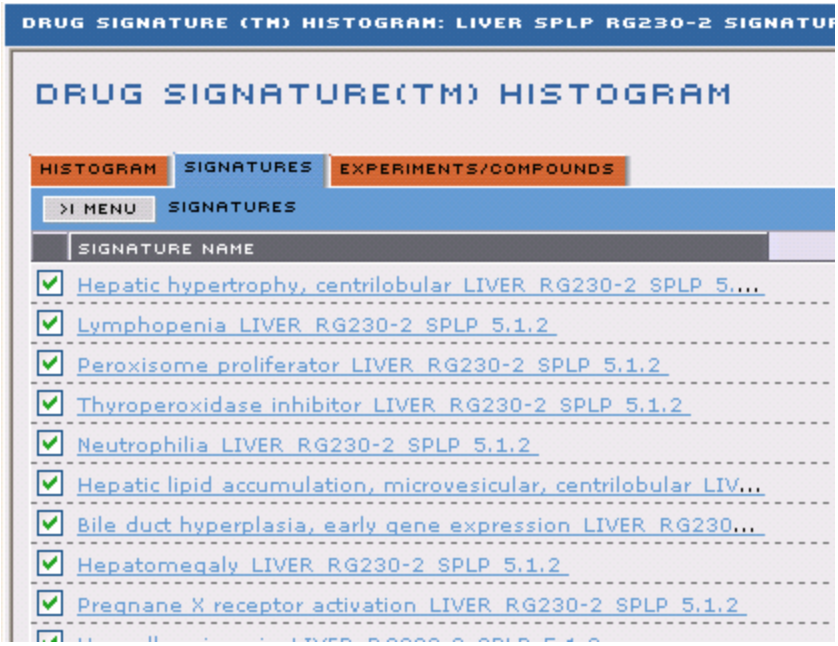
Exercise 8 ~ Part 3 ~ DATAVIEW of Drug Signature **HISTOGRAM** panel



The analysis shows that the fibrate compounds match the peroxisome proliferator signature, and the statins match the cholesterol biosynthesis inhibitor signature as we would expect.

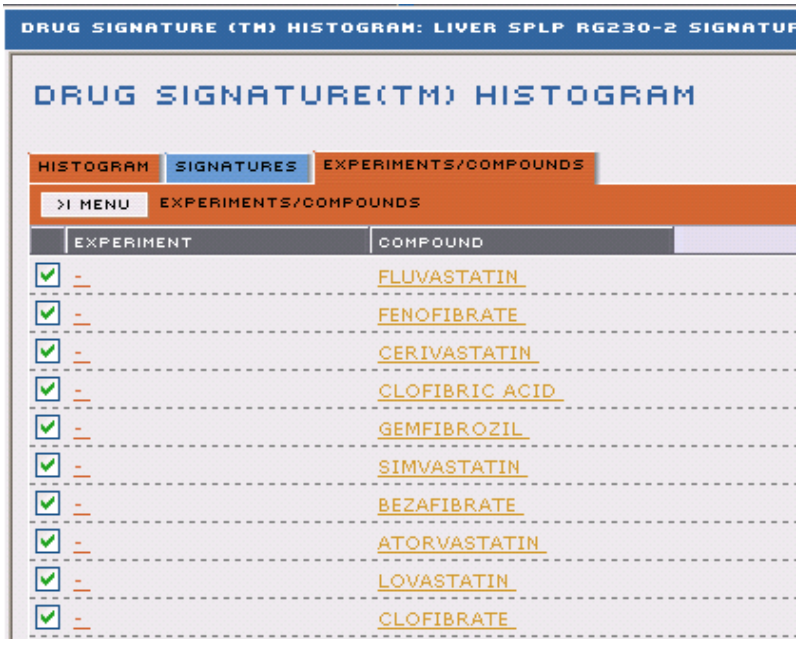
Step 15 You can now select signatures that are relevant by selecting the **SIGNATURE** tab in the **SIGNATURE HISTOGRAM** panel and saving the list.

Exercise 8 ~ Part 3 ~ Step 15 Drug Signature **SIGNATURE** panel



Step 16 You can also refine your lists of compounds or experiments by selecting the **EXPERIMENTS** tab in the **SIGNATURE HISTOGRAM** panel and saving the appropriate list.

Exercise 8 ~ Part 3 ~ Step 16 Drug Signature EXPERIMENTS/COMPOUNDS panel



You can save signature list or experiment/compound lists using the **MENU** button on any of the **SIGNATURE HISTOGRAM** tabs.

Exercise 8 ~ Part 4 ~ Generate a Drug Signature ™ Heatmap

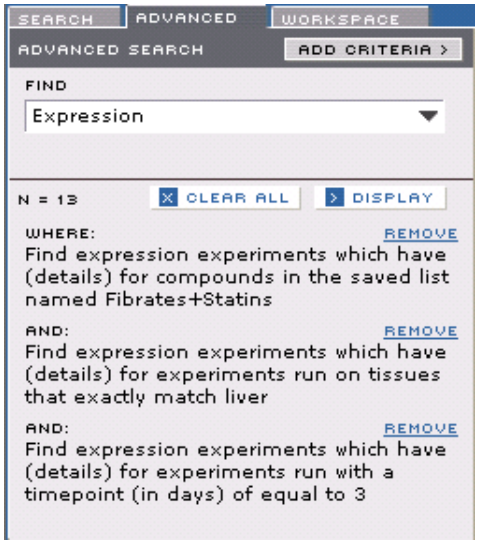
This exercise will show you how to use a different display to explore your signature matches. The visualization in this section uses a heatmap to display the signature matches and a 2-D hierarchical clustering to show the relative relationship among signatures or experiments.

To do this, we need to first populate a list of liver experiments for the five fibrates and five statins.

Step 1 Click on the **ADVANCED SEARCH** tab.
Step 2 Select **EXPRESSION** from the drop down menu.
Step 3 Click **CLEAR ALL** then click **ADD CRITERIA** and select
-which have (details) for
-compounds
-in the saved list named
-[Fibrates+Statins]
Step 4 Click the **RUN QUERY** button at the bottom of the **ADD CRITERIA** panel
Step 5 Click on **ADD CRITERIA** and select
-which have (details) for
-experiments run on tissues that
-exactly match
-[liver]
Step 6 Click the **RUN QUERY** button at the bottom of the **ADD CRITERIA** panel
This search results in a total of 48 expression experiments. We will further refine the list to only experiments performed with a time-point of 3 days.
Step 7 Click on **ADD CRITERIA** and select
-which have (details) for
-experiments run with a timepoint (in days) of
-equal to
-[3]
Step 8 Click on **ADD CRITERIA**
-which have (details) for
-chip names which
-exactly match
-[RG230-2]

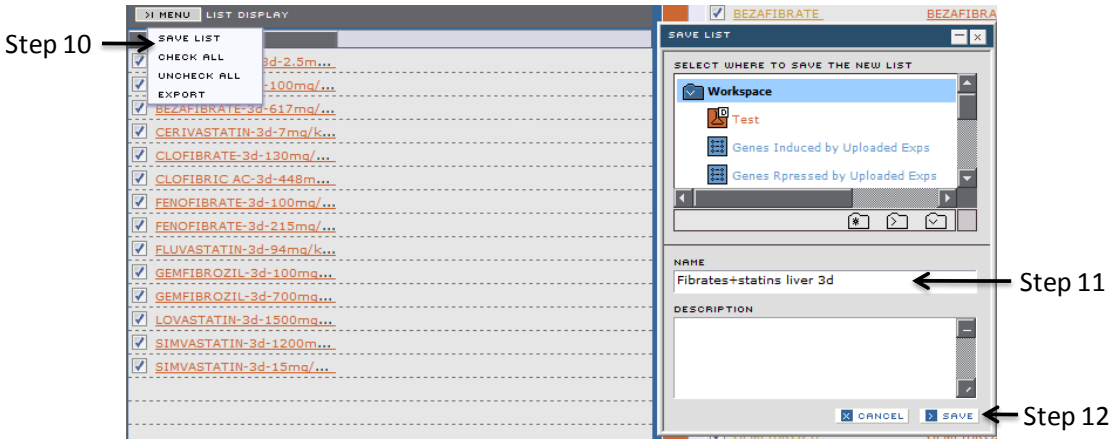
Step 8 Click the **RUN QUERY** button

Search results from Exercise 8 ~ Part 4 ~ Steps 1 thru 8



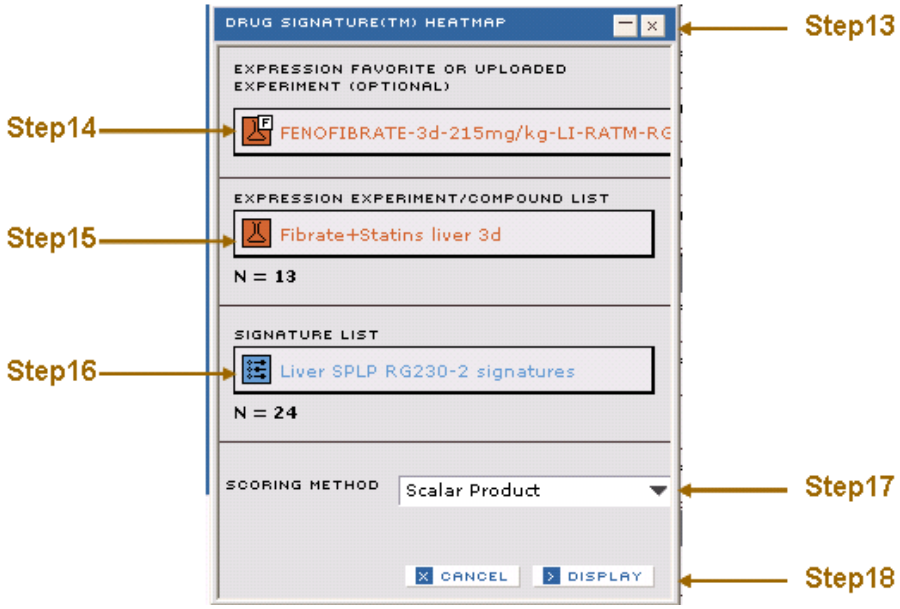
Step 9 Click on the **DISPLAY** button.
Notice that the number of experiment is down to 14.
Step 10 Select **SAVE LIST** from the **MENU** button on the **LIST DISPLAY** panel.
Step 11 Name the list in the **SAVE LIST** dialog window, for example as **Fibrates+Statins liver 3d**.
Step 12 Click the **SAVE** button at the bottom of the **SAVE LIST** dialog window.

Exercise 8 ~ Part 4 ~ Steps 10 thru 12

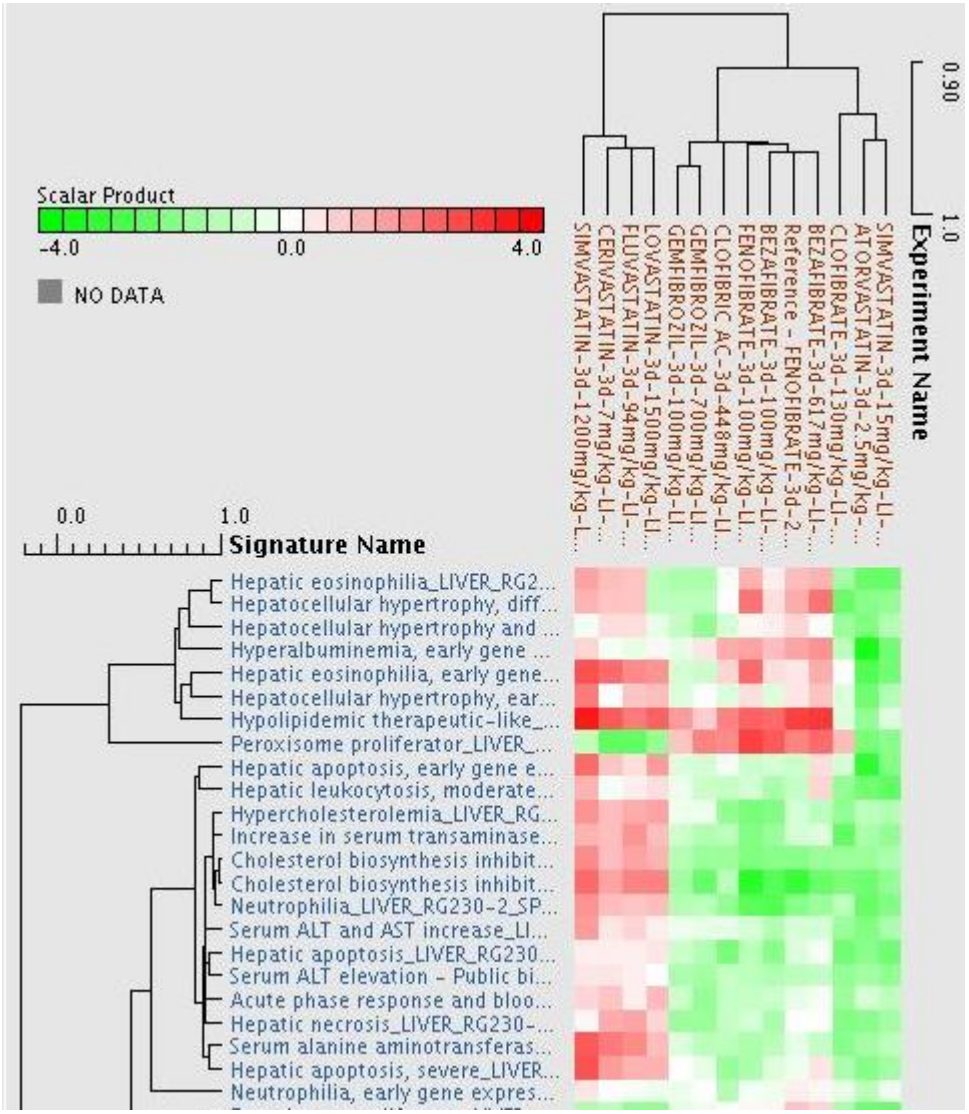


Step 13 Open the **DRUG SIGNATURE(TM) HEATMAP** tool from the **DATA VISUALIZATION** section of the **TOOLBOX.**
Step 14 Drag the list **FENOFIBRATE-3D-215MG/KG-LI-RATM-RG230-2** that you created in **Exercise 9 ~ Part 3** into the **EXPRESSION EXPERIMENT DATA/FAVORITE** field of the **DRUG SIGNATURE(TM) HEATMAP** dialog window.
Step 15 Drag the list **Fibrates+Statins liver 3d** into the **COMPOUND/EXPRESSION EXPERIMENT LIST** field of the **DRUG SIGNATURE(TM) HEATMAP** dialog window.
Step 16 Drag the list **Liver SPLP RG230-2 signature** into the **SIGNATURE LIST** field of the **DRUG SIGNATURE(TM) HEATMAP** dialog window.
Step 17 Select **SCALAR PRODUCT** from the **SCORING METHOD** drop-down manu.
Step 18 Click on the **DISPLAY** button at the bottom of this dialog window.

Exercise 8 ~ Part 4 ~ Steps 13 thru 18



Exercise 8 ~ Part 4 ~ DATAVIEW of Drug Signature Performance HEATMAP panel



The 13 experiments in the list are separated into two clusters, the fibrates and the statins, based on the matches to Drug Signatures. The fibrate compounds match the peroxisome proliferator signature, while the statins match the cholesterol biosynthesis inhibitor signature, a result you would expect given their different mechanism of action.
The **EXPERIMENTS/COMPOUNDS** tab in the display also allows you view or save the list of experiments.

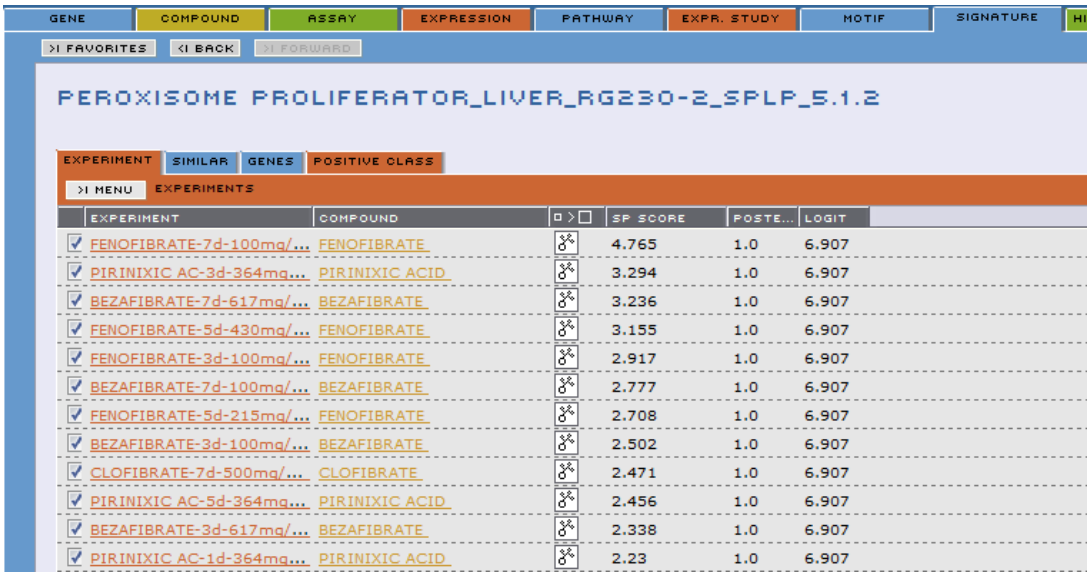
Exercise 8 ~ Part 5 ~ Dissect Biological Meaning in a Drug Signature

Each Drug Signature is composed of a list of genes with associated weight for best separation between the positive and negative training class compounds. This part of the exercise will show you how to visualize the impact of expression experiments on signature genes.

In DrugMatrix, it is also possible to identify biological pathways that can be used to explain the match to a Drug Signature, which is achieved by using the innovative **MOTIF** technology. A motif is derived very much like a Drug Signature, except that only genes associated with the 137 annotated biological pathways in Drug Matrix are used as gene universe. So instead of using genes, such as ESTs, that have no apparent biological annotation, motif uses genes with well characterized physiological and molecular function. This can facilitate the interpretation of Drug Signature matches in a biologically meaningful way.

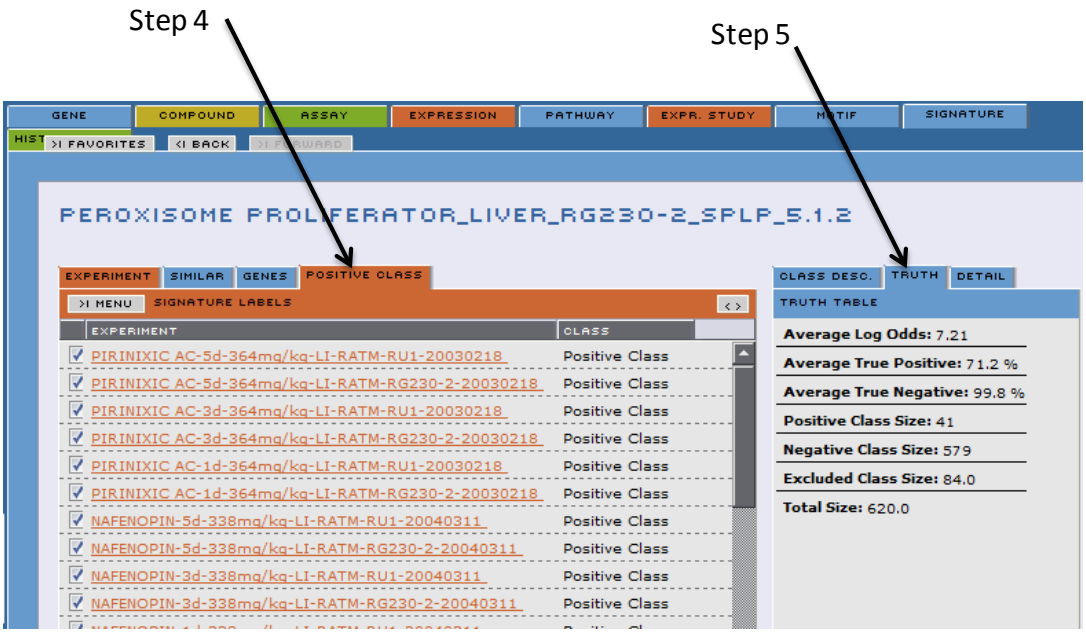
We will use the **peroxisome proliferator** Drug Signature as an example.
Step 1 Click on the **SIGNATURES** tab in the **DRUG SIGNATURE(TM) HEATMAP** panel.
Step 2 Select the **peroxisome proliferator** signature from the list.
Step 3 The hyperlink will take you to the **SIGNATURE DOMAIN** panel.

Exercise 8 ~ Part 5 ~ DATA VIEW of the SIGNATURE DOMAIN



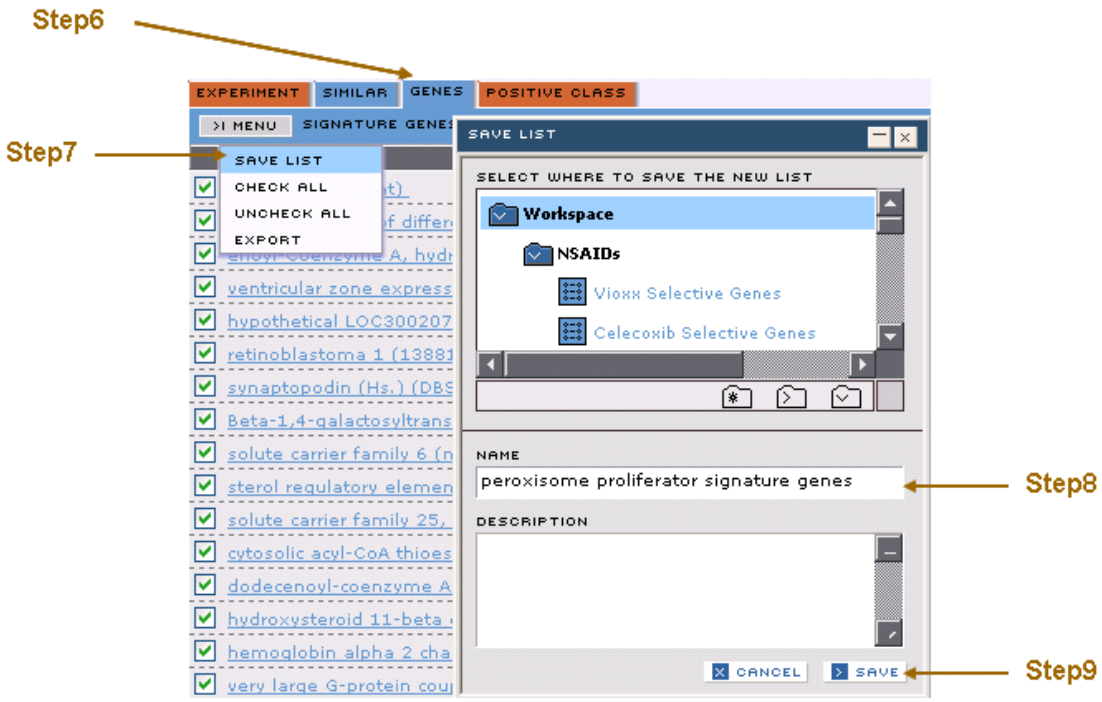
The first window shown is the **EXPERIMENT** panel (a list of experiments that match this signature) on the left and the **CLASS DESCRIPTION** panel (the description of the positive, negative and excluded classes used to derive the signature) on the right.
Step 4 Click on the **POSITIVE CLASS** tab in the left panel, this shows a list of experiments used in the positive training class for derivation of this signature.
Step 5 Click on the **TRUTH** tab in the right panel, this shows a list of signature performance criteria, including log odd ratio, percent of true positive and true negative, the size of positive, negative and exclude classes, etc.

Exercise 8 ~ Part 5 ~ Steps 4 and 5



Step 6 Click on the **GENES** tab in the left panel, this shows a list of genes used in this signature to classify compounds.
Step 7 Select **SAVE LIST** from the drop-down **MENU**.
Step 8 Name the list in the **SAVE LIST** dialog window, for example as **peroxisome proliferator signature genes**.
Step 9 Click the **SAVE** button at the bottom of the **SAVE LIST** dialog window.

Exercise 8 ~ Part 5 ~ Steps 6 thru 9



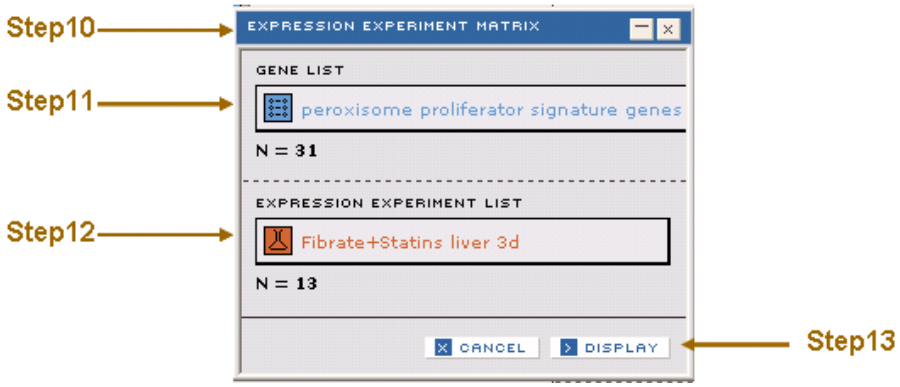
Step 10 Open the **EXPRESSION EXPERIMENT MATRIX** tool from the **DATA VISUALIZATION** section of the **TOOLBOX.**

Step 11 Drag the list **peroxisome proliferator signature genes** into the **GENE LIST** field of the **EXPRESSION EXPERIMENT MATRIX** dialog window.

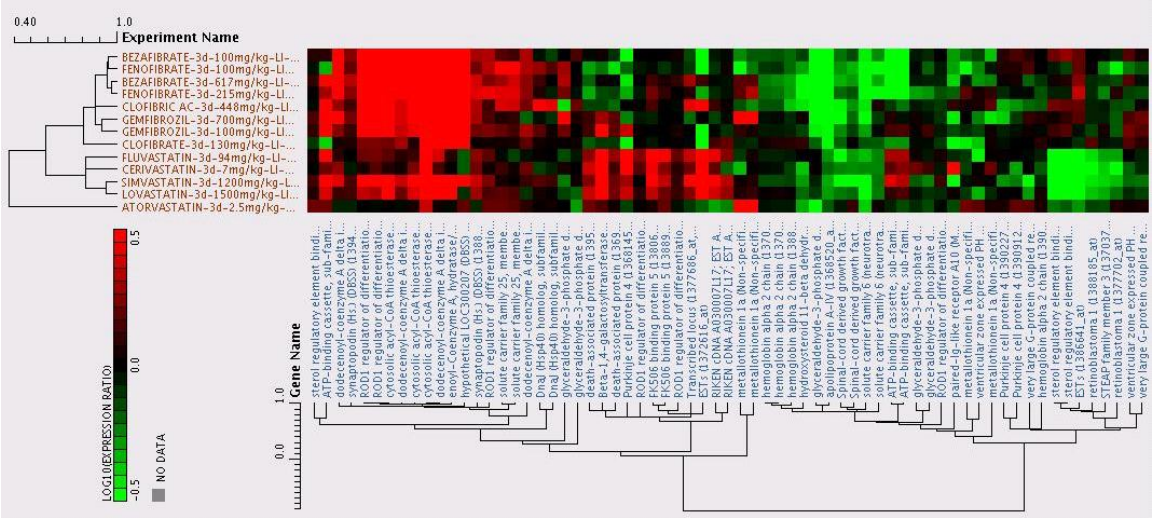
Step 12 Drag the list **Fibrates+Statins liver 3d** into the **EXPRESSION EXPERIMENT LIST** field.

Step 13 Click on the **DISPLAY** button at the bottom of this dialog window.

Exercise 8 ~ Part 5 ~ Steps 10 thru 13



Expression Experiment Matrix output for peroxisome proliferator signature genes gene list and Fibrates+Statins liver 3d experiment list

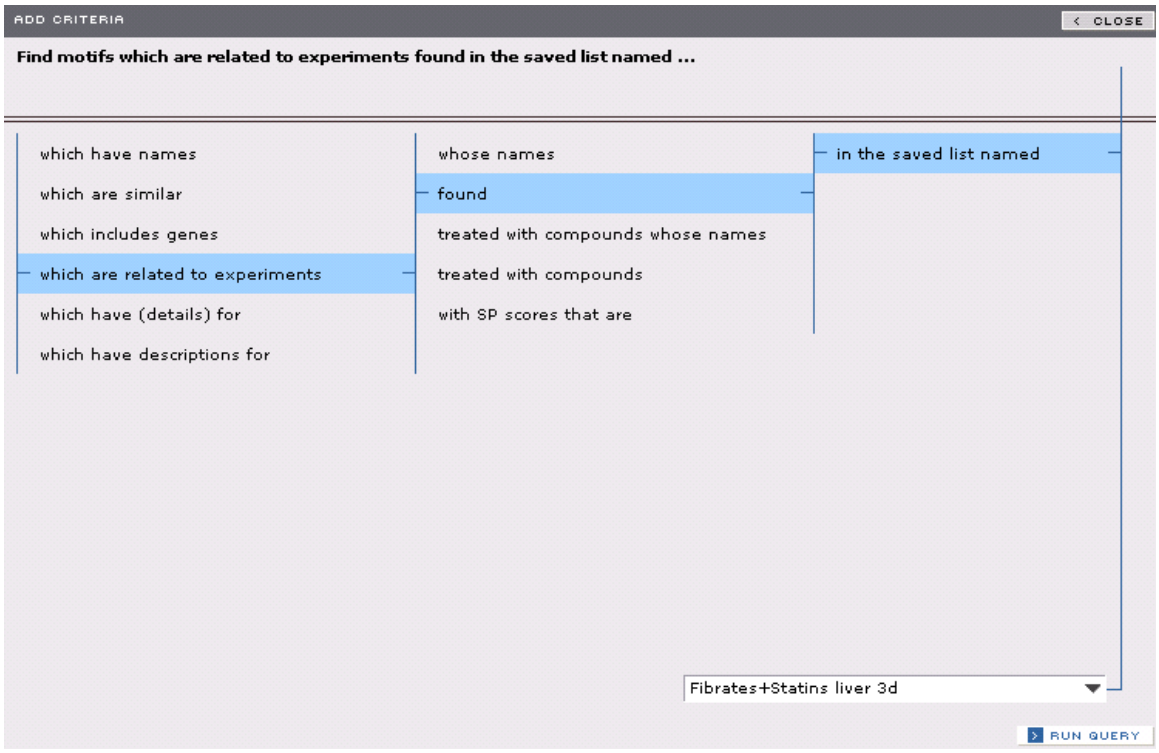


As you can see from the expression experiment matrix, the fibrates and statins are clustered into two distinct groups, indicating distinct pattern of impact on the signature genes.

Next, we will look at motifs associated with the peroxisome proliferator signature and identify pathway(s) that can help explain the signature match.

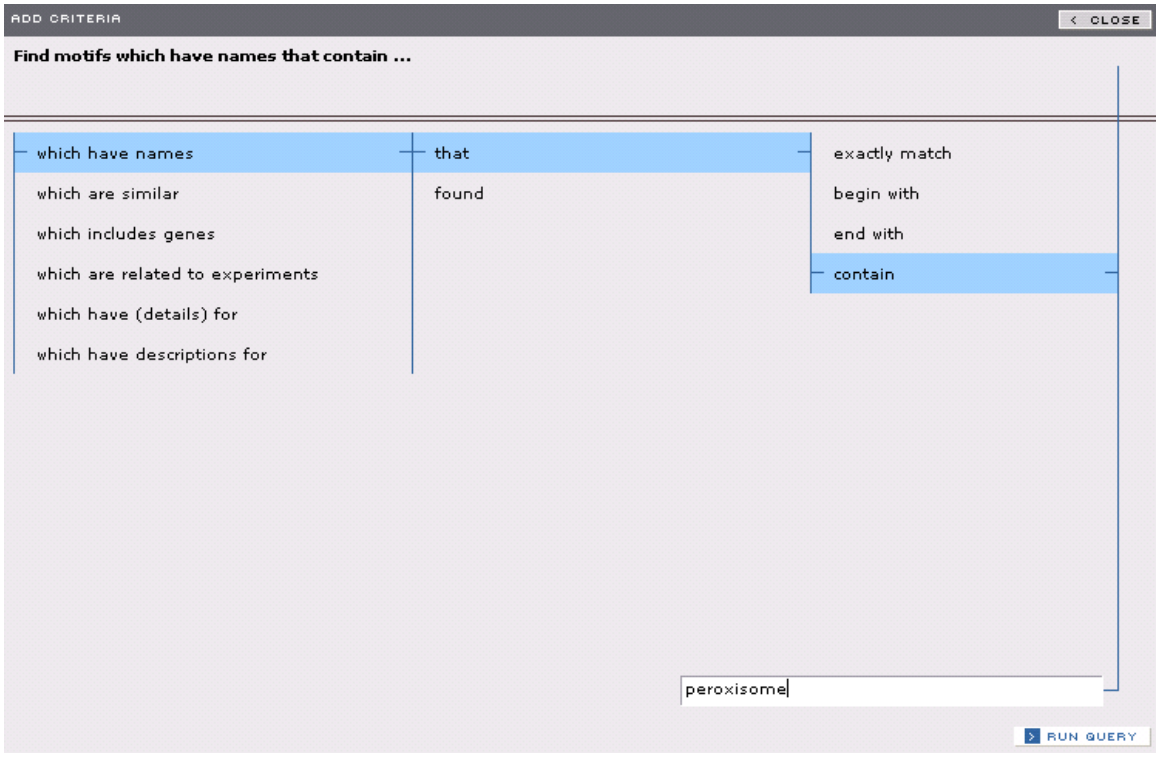
Step 14 Click on the **ADVANCE SEARCH** tool
Step 15 Select **MOTIF** from the drop down manu
Step 16 Click **ADD CRITERIA** and select
-which are related to experiments
-found
-in the saved list named
-[Fibrates+Statins liver 3d]
Step 17 Click the **RUN QUERY** button at the bottom of the **ADD CRITERIA** panel

Exercise 8 ~ Part 5 ~ Steps 16 and 17



Step 18 Click **ADD CRITERIA** and select
-which have names
-that
-contain
-[peroxisome]
Step 19 Click the **RUN QUERY** button at the bottom of the **ADD CRITERIA** panel

Exercise 8 ~ Part 5 ~ Steps 18 and 19

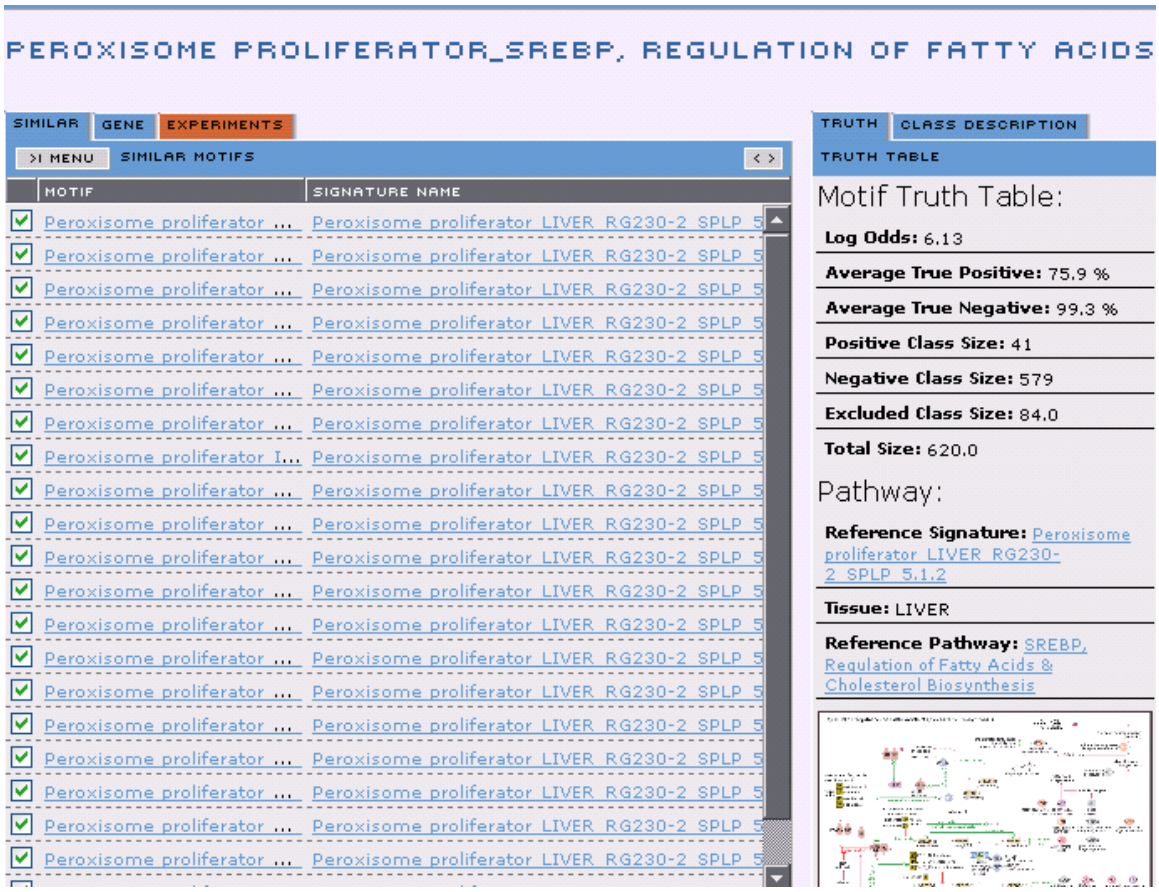


Step 20 Click the **RUN QUERY** button at the bottom of the **ADVANCE SEARCH** panel
Step 21 Click the **DISPLAY** button on the **ADVANCE SEARCH** panel.
A total of 32 motifs are found with the above search. A quick scan of the motif names indicates that the regulation of fatty acid metabolism pathways are associated with the peroxisome proliferator signature, and can at least partially explain the match to the peroxisome proliferator Drug Signature.

Exercise 8 ~ Part 5 ~ Step 21

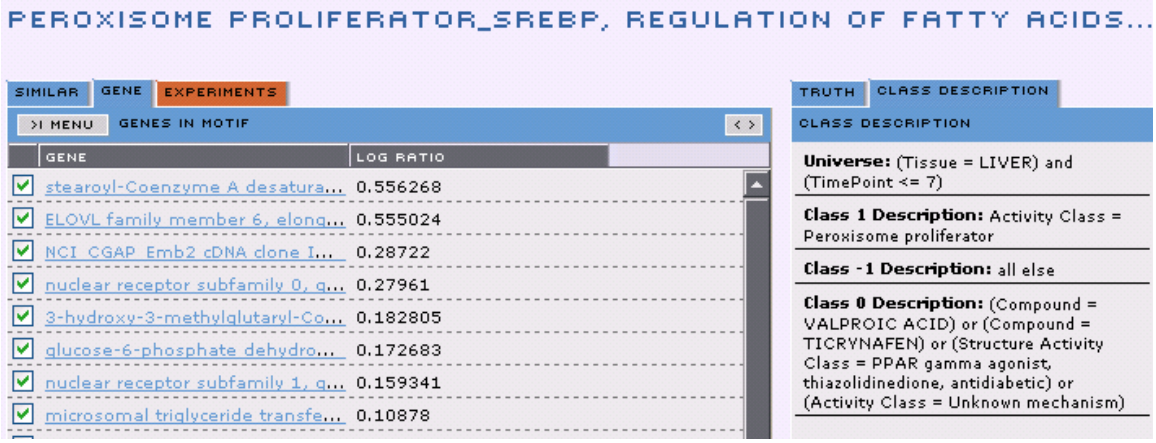
Step 22 Click the **Peroxisome proliferator SREBP, Regulation of Fatty Acid & Cholesterol Biosynthesis** motif.

Exercise 8 ~ Part 5 ~ DATA VIEW of the MOTIF panel



The first window shown is the **SIMILAR** tab (a list of motifs that are similar to the **Peroxisome proliferator_SREBP, fatty acid & cholesterol biosynthesis** motif) on the left panel and the **TRUTH** tab (a list of signature performance criteria, including log odd ratio, percent of true positive and true negative, the size of positive, negative and exclude classes, etc.) on the right panel. The reference signature and reference pathway names are shown at the bottom of the **TRUTH TABLE** as hyperlink.

Step 23 Click on the **GENE** tab in the left panel, this shows a list of genes used in this motif for classifying compounds.
Step 24 Click on the **CLASS DESCRIPTION** tab in the right panel, this shows the description of the universe, the positive, negative and exclude class compounds.

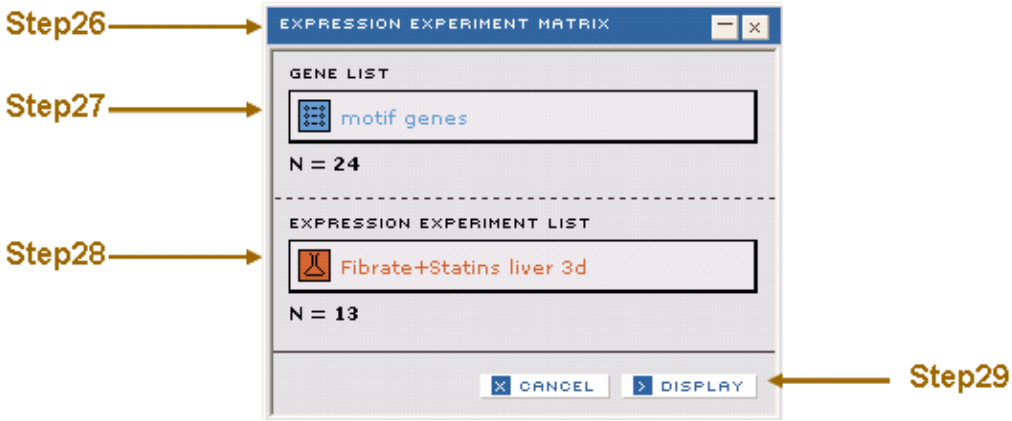


Step 25 Save the list of genes as **Motif genes** by selecting **SAVE LIST** from the drop-down **MENU**.

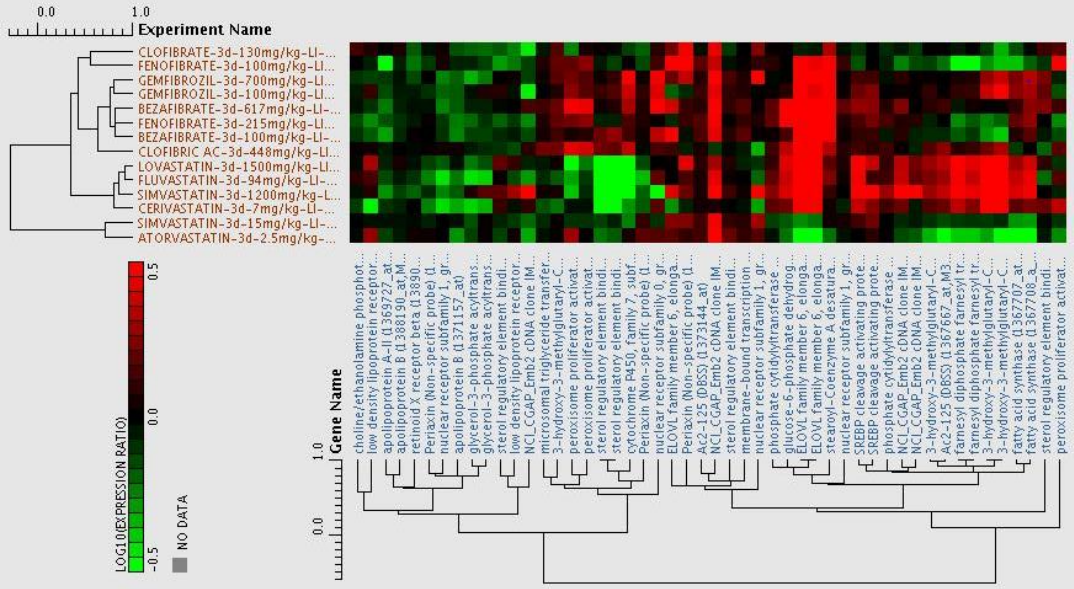
You can also click on the **EXPERIMENT** tab in the motif panel, which will show a list of experiments that match to the motif.

Step 26 Open the **EXPRESSION EXPERIMENT MATRIX** tool from the **DATA VISUALIZATION** section of the **TOOLBOX.**
Step 27 Drag the list **Motif genes** into the **GENE LIST** field of the **EXPRESSION EXPERIMENT MATRIX** dialog window.
Step 28 Drag the list **Fibrates+Statins liver 3d** into the **EXPRESSION EXPERIMENT LIST** field.
Step 29 Click on the **DISPLAY** button at the bottom of this dialog window.

Exercise 8 ~ Part 5 ~ Steps 26 thru 29



Expression Experiment Matrix output for Motif Genes gene list and Fibrates+Statins liver 3d experiment list



As you can see from the expression experiment matrix, the fibrates and statins are clustered into two distinct groups, indicating distinct pattern of impact on the motif genes. The fibrates strongly induce genes involved in fatty acid biosynthesis, whereas statins do not.

Conclusions to Exercise 8

In this exercise you have learned how to extract appropriate sets of Drug Signatures using the **ADVANCE** query tool. You used the **DRUG SIGNATURE HISTOGRAM and DRUG SIGNATURE HEATMAP** tools to visually examine similarities and differences between a group of fibrates and statins. You compared the impact of the different fibrates and statins on peroxisome proliferator signature genes using the **EXPRESSION EXPERIMENT MATRIX** tool. Finally, you identified biological pathways that could help explain the match to the peroxisome proliferator signature using the Iconix **MOTIF** technology. These tools allowed you to compare and contrast the two groups of experiments using the Iconix library of validated signatures and motifs.

Exercise 9: Relating Clinical Chemistry and Molecular Pharmacology Findings to Gene Expression Data

This exercise demonstrates how to explore multiple domains to elucidate compound similarities and differences all the way to the relevant pathway. The exercise investigates cholesterol lowering compounds, what are they and what are their possible mechanisms, and relates clinical chemistry and molecular pharmacology data with gene expression.

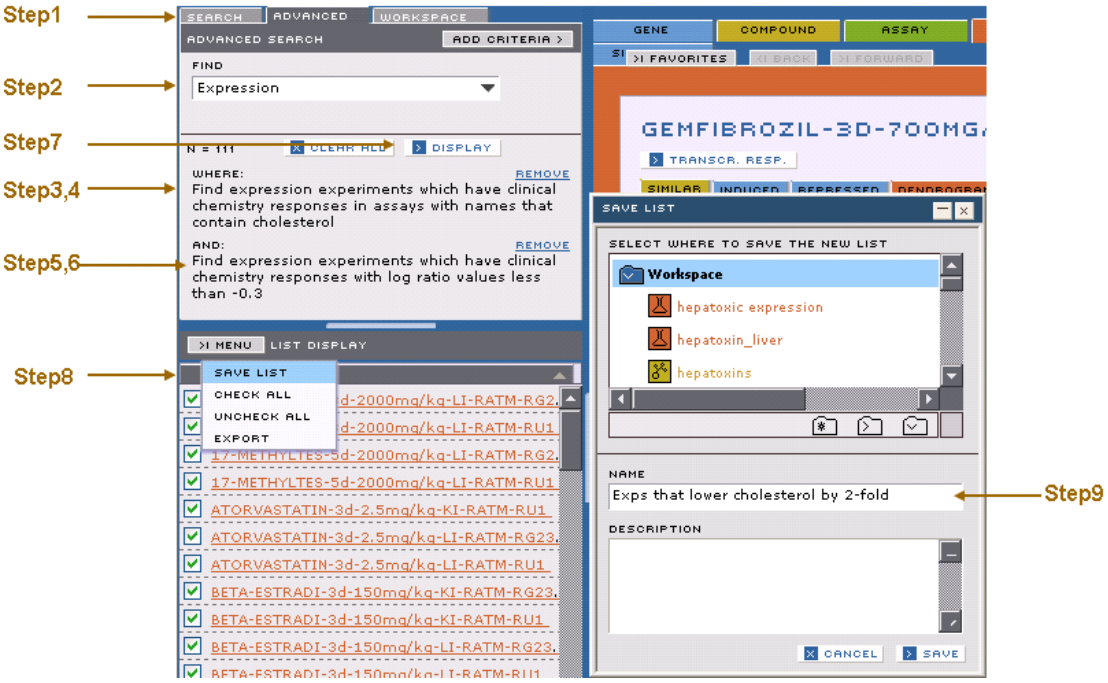
You will learn to:
• Use an advanced search to extract experiments that have certain clinical chemistry response, assays that meet certain criteria.
• Use the **EXPRESSION EXPERIMENT > COMPOUNDS** translation tool to convert an experiment list to a compound list and vice versa.
• Use the **COMPOUND COMPARE** tool to get a quick overview of a list of compounds.
• Use the **BIO-ACTIVITY MATRIX** tool to visualize the clustering of compounds based on assays.
• Use a simple search to extract a desired pathway and save the pathway genes.
• Use the **EXPRESSION EXPERIMENT MATRIX** tool to visualize the clustering of experiments based on gene expression.
• Use the **PATHWAY VISUALIZATION** tool to visualize the effect of compounds on pathways.

Exercise 9 ~ Part 1 ~ Search for Cholesterol-lowering Experiments

In this part of the exercise, we will conduct an advance query to extract experiments that lower cholesterol by at least 2 fold.
Step 1 Click on the **ADVANCED** search tab.
Step 2 Select **EXPRESSION** from the select menu.
Step 3 Click on **ADD CRITERIA**, and select the following criteria
-which have clinical chemistry responses
-in assays with names that
-contain
-[cholesterol]
Step 4 Click the **RUN QUERY** button.
This query selects all experiments with cholesterol level measured in clinical chemistry study.
To further limit the experiments...
Step 5 Click the **ADD CRITERIA** button, and select
-which have clinical chemistry responses
-with log ratio values
-less than
-[-0.3]
Step 6 Click on the **RUN QUERY** button.

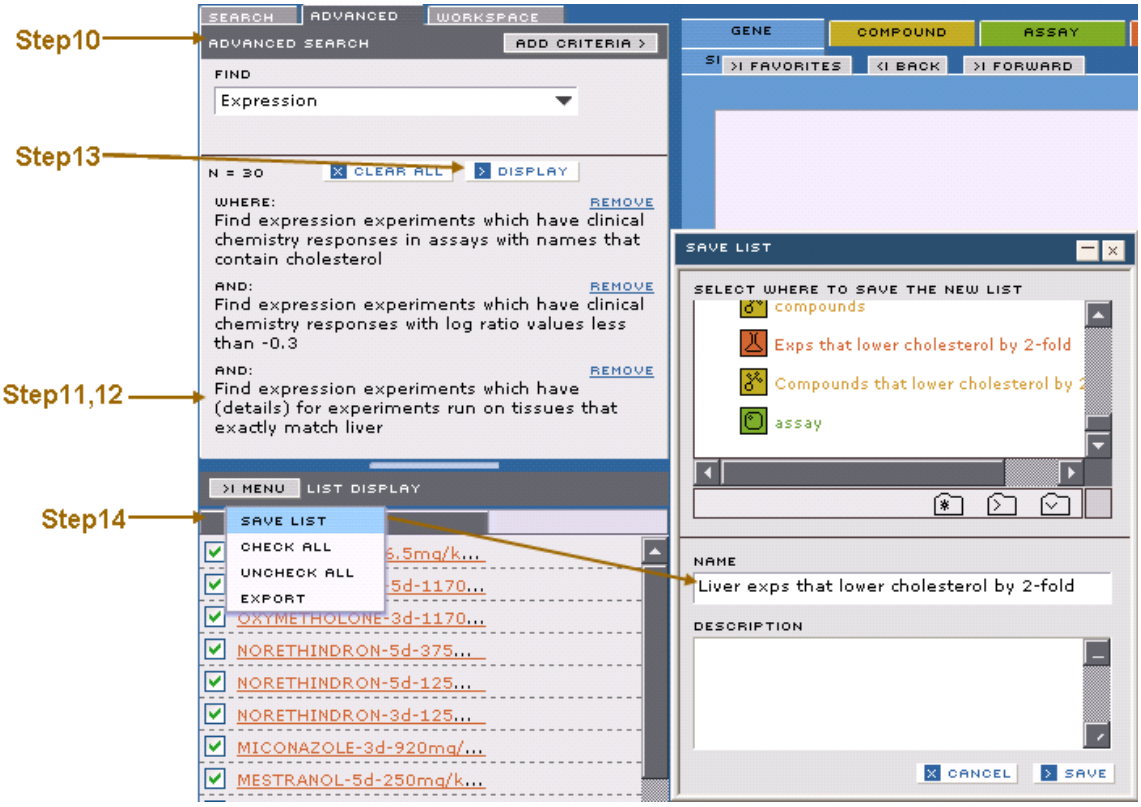
This query selects all experiments with any clinical chemistry response lowered by at least 2 fold. Combined with the first query, it selects all experiments that lower cholesterol level by 2 fold.
Step 7 Click on the **DISPLAY** button to display the experiment list in the **LIST DISPLAY** area.
Step 8 Select **SAVE LIST** from the drop-down **MENU** button
Step 9 Save the list of experiments as **exps that lower cholesterol by 2 fold**

Exercise 9 ~ Part 1 ~ Steps 1 thru 9



To further narrow the experiments to only those done in liver:
Step 10 Click the **ADVANCED** tab make sure that the previous search results are still displayed in the **SEARCH** panel.
Step 11 Click the **ADD CRITERIA** button, and select
-which have (details) for
-experiments run on tissues that
-exactly match
-[liver]
Step 12 Click on the **RUN QUERY** button
Step 13 Click on the **DISPLAY** button to display the experiment list, there should be a total of 30 experiments
Step 14 Click on the **MENU** button and save the list of experiments as **Liver exps that lower cholesterol by 2 fold**

Exercise 9 ~ Part 1 ~ Steps 10 thru 14

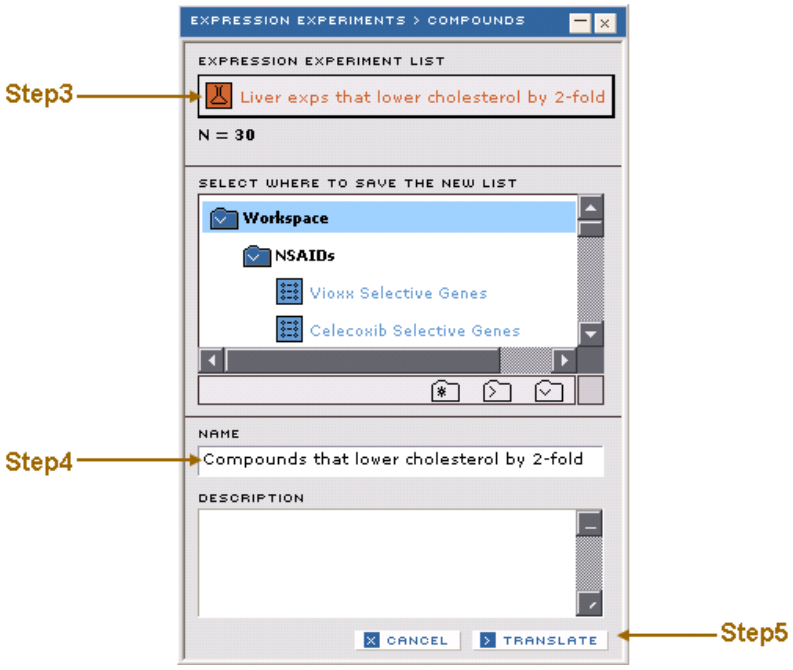


Exercise 9 ~ Part 2 ~ Convert the Experiment List to a Compound list

Next, we will convert the experiment list to a compounds list.

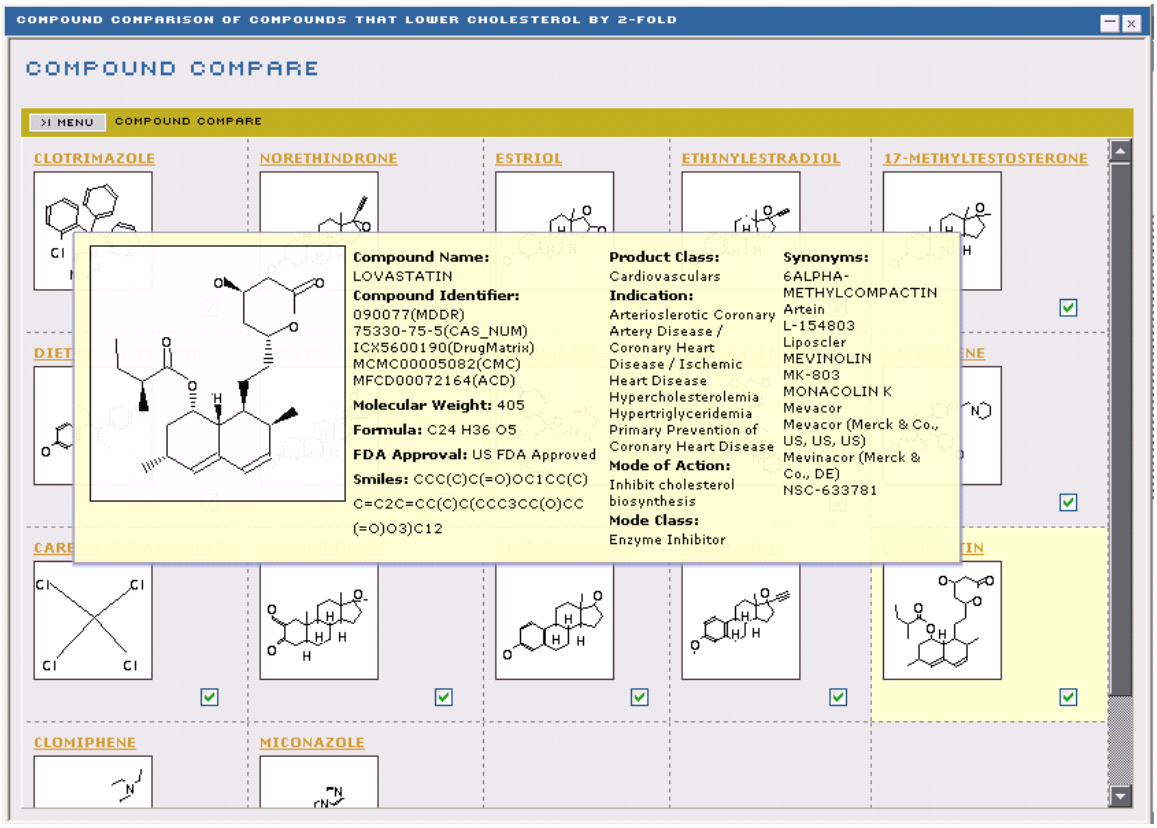
Step 1 Click on the **WORKSPACE** tab and open the **TOOLBOX**
Step 2 Choose **EXPRESSION EXPERIMENTS > COMPOUNDS** tool from the **DATA TRANSLATION** section.
Step 3 Drag the **Liver exps that lower cholesterol by 2 fold** list into the **EXPRESSION EXPERIMENT LIST** field
Step 4 Name the new list **comps that lower cholesterol by 2 fold**
Step 5 Click the **TRANSLATE** button.

Exercise 9 ~ Part 2 ~ Steps 3 thru 5



To get a quick overview of these compounds...
Step 6 Open the **TOOLBOX** and choose the **COMPOUND COMPARE** tool from the **DATA VISUALIZATION** section
Step 7 Drag the **comps that lower cholesterol by 2 fold** list into the **COMPOUND LIST** field.
Step 8 Click the **DISPLAY** button
Step 9 Mouse over the chemical structures for detailed information about the displayed compounds.

Exercise 9 ~ Part 2 ~ DATA VIEW of results from Steps 6 thru 9



Notice that the compounds include the statin class of drugs, sex steroids and some anti-cancer agents.

Exercise 9 ~ Part 3 ~ Extract Specific Assays for Compounds

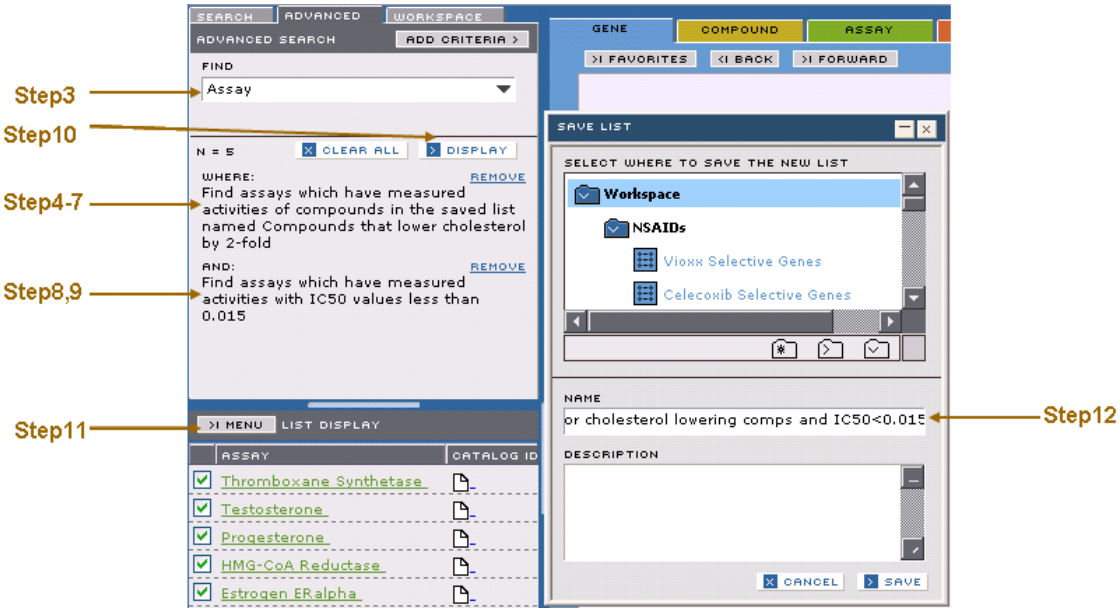
We will conduct advance queries on assays that have been performed for the list of cholesterol lowering compounds you have previously extracted from the data base.

Step 1 Click on the **ADVANCED** search tab
Step 2 Click the **CLEAR ALL** button to remove all previous criteria
Step 3 Select **ASSAY** from the select menu.
Step 4 Click the **ADD CRITERIA** button
Step 5 Select the following criteria from the **ADD CRITERIA** panels for **ASSAYS...**
-which have measured activities
-of compounds
-in the saved list named
Step 6 Choose **comps that lower cholesterol by 2 fold** from the select menu
Step 7 Click on the **RUN QUERY** button
This query selects all assays done for the selected compound list.

To limit the results to more specific assays...
Step 8 Click **ADD CRITERIA** and select the following criteria
-which have measured activities
-with IC50 values
-less than
-[0.015]
Step 9 Click the **RUN QUERY** button

This query selects all assays with IC50 < 0.015μm. Combined with the first query, it selects all assays done for the specified compound list with IC50 < 0.015μm.
Step 10 Click on **DISPLAY** to display the assay list in the **LIST DISPLAY** area
Step 11 Select **SAVE LIST** from the **MENU** button on the **LIST DISPLAY** panel
Step 12 In the **SAVE LIST** dialog window, save the list of assays as **assays for cholesterol lowering comps and IC50<0.015um**

Exercise 9 ~ Part 3 ~ Steps 3 thru 12

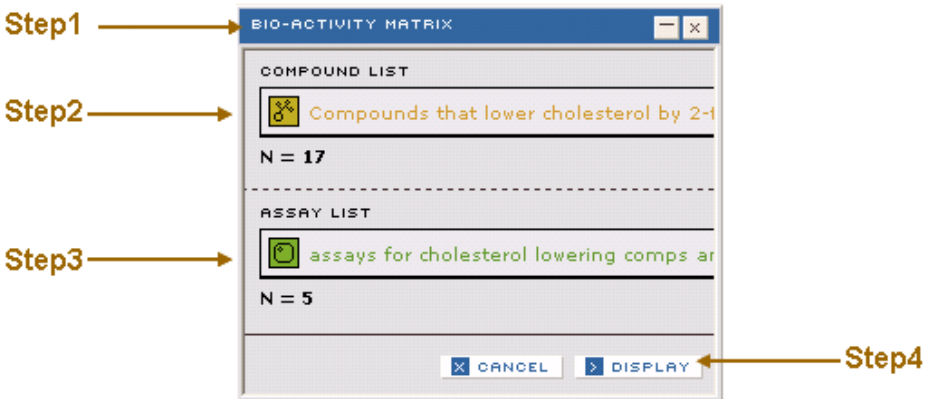


Exercise 9 ~ Part 4 ~ Visualize the Clustering of Compounds by Assays

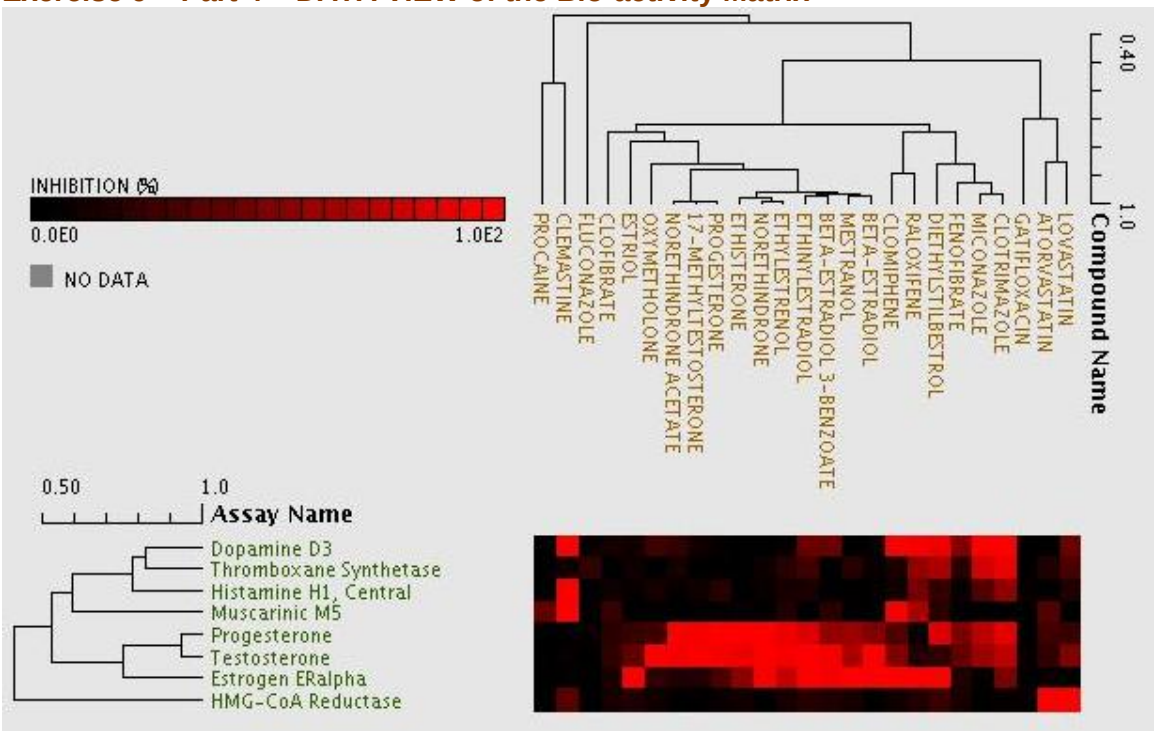
We will use the **BIO-ACTIVITY MATRIX** tool to create a visualization of the clustering of compounds based on assays.
Step 1 Open the **TOOLBOX** and choose the **BIO-ACTIVITY MATRIX** tool from the **DATA VISUALIZATION** section.
Step 2 Drag the **comps that lower cholesterol by 2 fold** list into the **COMPOUND LIST** field.
Step 3 Drag the **assays for cholesterol lowering comps and IC50<0.015um** list into the **ASSAY LIST** field

Step 4 Click the **DISPLAY** button

Exercise 9 ~ Part 4 ~ Steps 1 thru 4



Exercise 9 ~ Part 4 ~ DATA VIEW of the Bio-activity Matrix



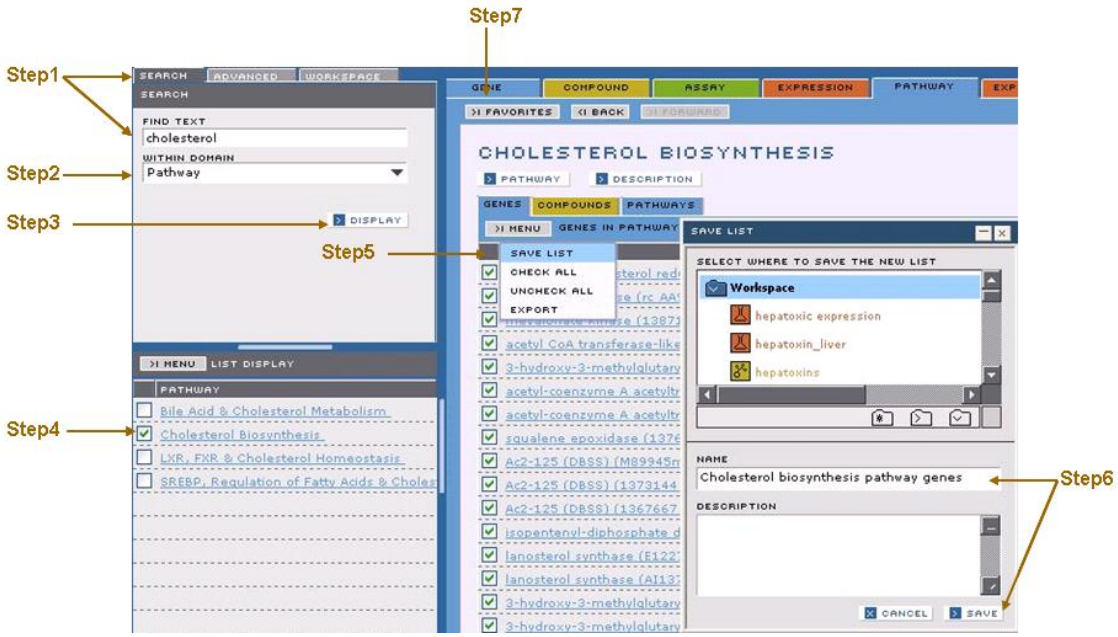
Notice how the assays cluster into consistent sets. The compounds show some similar natural clustering by class.

Exercise 9 ~ Part 5 ~ Visualize Clustering of Experiments Based on Gene Expression

In this section, we will use the **EXPRESSION EXPERIMENT MATRIX** tool to visualize the clustering of experiments based on gene expression

Step 1 Click on the **SEARCH** tab and type **cholesterol** in the **FIND TEXT** field
Step 2 Select **PATHWAY** from the **WITHIN DOMAIN** select menu
Step 3 Click on the **DISPLAY** button
Step 4 Click on **Cholesterol Biosynthesis** in the **LIST DISPLAY** area
This opens the **PATHWAY** domain report to the right.
Step 5 Select **SAVE LIST** from the **MENU** button on the **GENES** panel of the **PATHWAY** domain report.
Step 6 In the **SAVE LIST** dialog save the list of genes as **cholesterol biosynthesis genes**
Step 7 Click on the **FAVORITES** button on the top of the **PATHWAY** domain report and click on **ADD TO FAVORITES**.

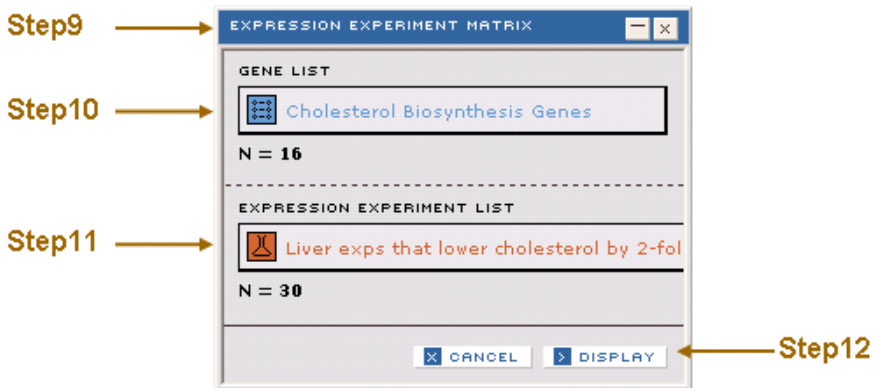
Exercise 9 ~ Part 5 ~ Steps 1 thru 7



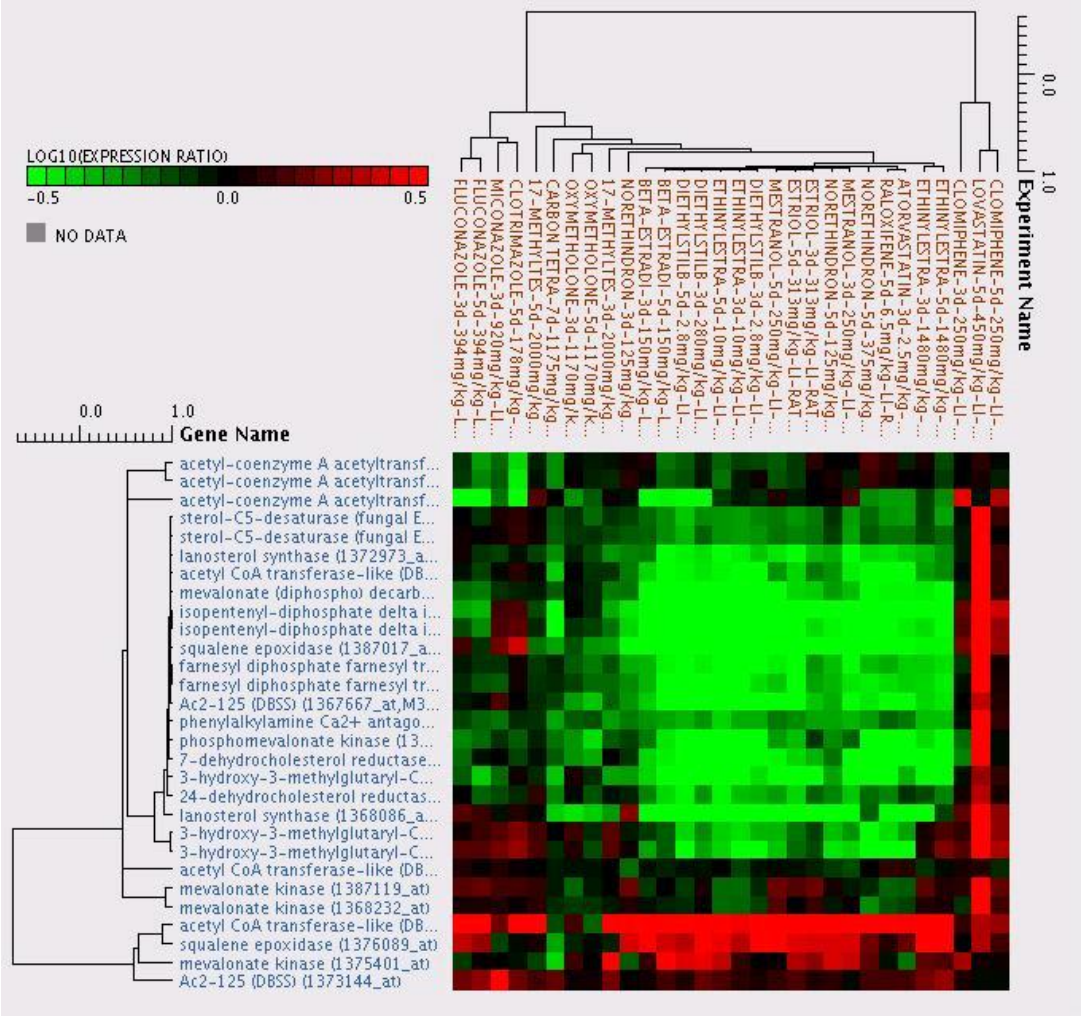
Step 8 Select the **WORKSPACE** tab.
Step 9 Open the **TOOLBOX** and select the **EXPRESSION EXPERIMENT MATRIX** tool from the **DATA VISUALIZATION** section.
Step 10 Drag the **cholesterol biosynthesis genes** list into the **GENE LIST** field.
Step 11 Drag the **Liver exps that lower cholesterol by 2 fold** list into the **EXPERIMENT LIST** field.

Step 12 Click the **DISPLAY** button.

Exercise 9 ~ Part 5 ~ Steps 9 thru 12



Expression Experiment Matrix output for cholesterol biosynthesis genes gene list and Liver exps that lower cholesterol by 2 fold



Notice how the genes and experiments cluster.
Note: The image presented here is derived from an older version of the Drugmatrix database that did not include Affymetrix data, hence the Expression Experiment Matrix displayed by the new version of the database will be much larger and include a number of additional experiments. Also note the greyed areas are where the Codelink arrays are missing probes for a given gene, hence there is no measurement.

Exercise 9 ~ Part 6 ~ Visualize the Effects of Compounds on Pathways

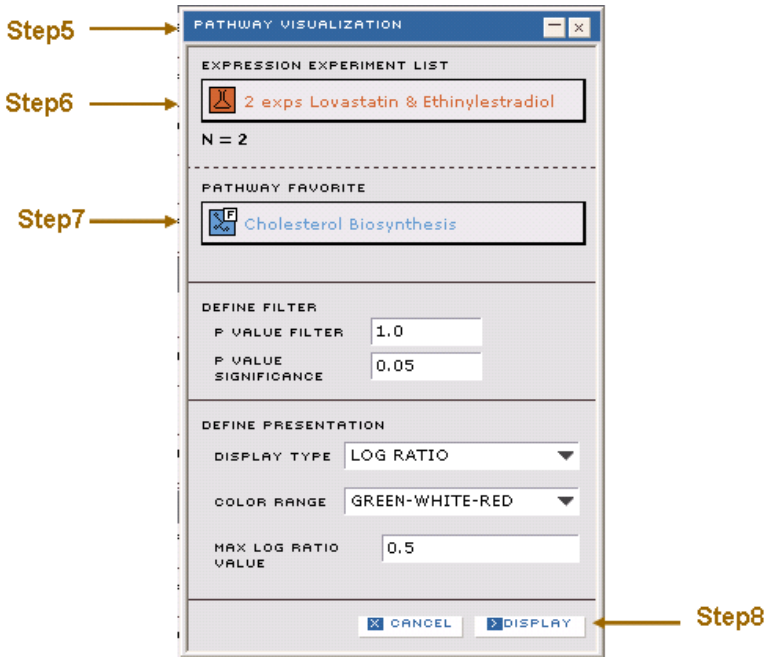
We will now use the **PATHWAY VISUALIZATION** tool to visualize the effect of compounds on pathways.

Step 1 Click on the **EXPERIMENTS** tab in the **EXPRESSION EXPERIMENT MATRIX DATAVIEW** visualization panel.
Step 2 Select **UNCHECK ALL** from the **MENU** button on that panel.
Step 3 Check the experiments **LOVASTATIN-5d-450mg/kg** and **ETHINYLESTRA-5d-10mg/kg**

Exercise 9 ~ Part 6 ~ Steps 1 thru 3

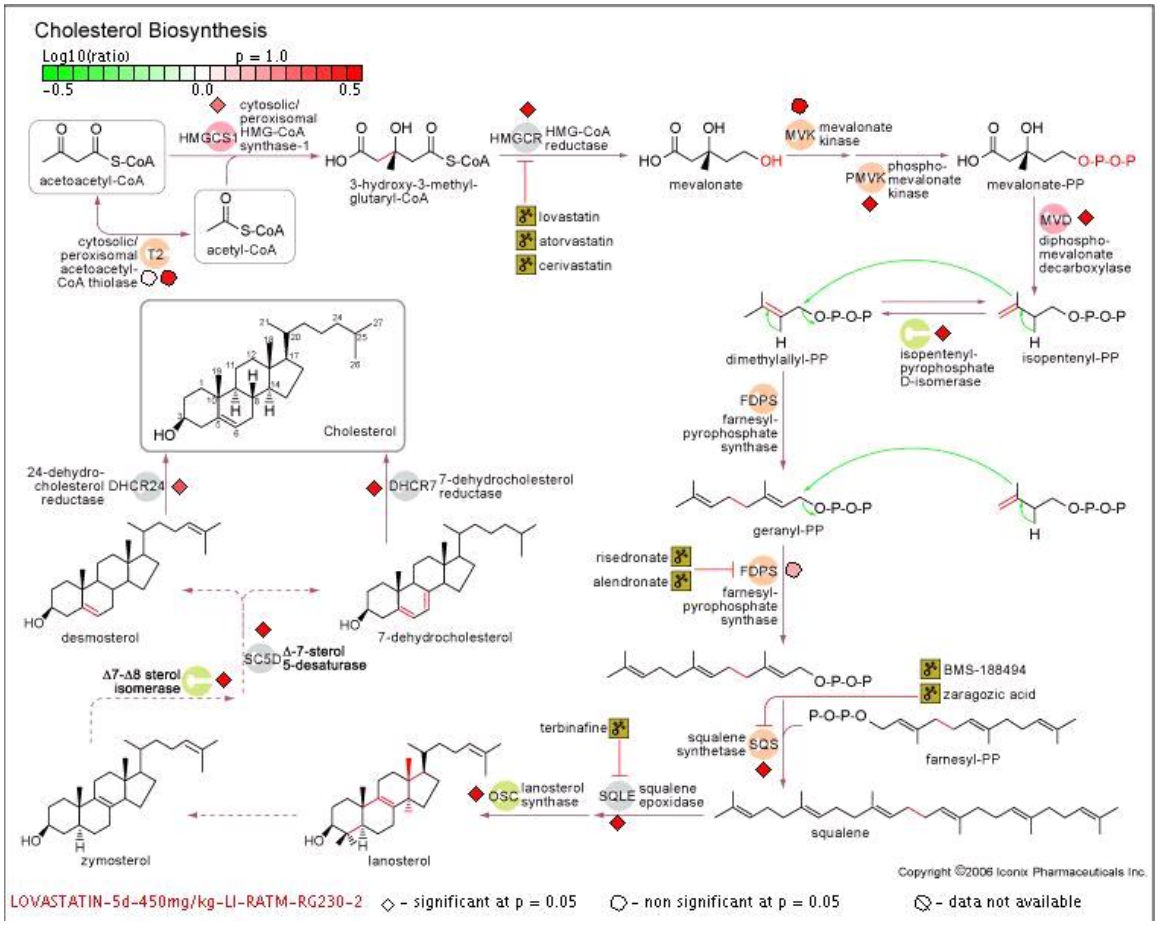
Step 4 Save the experiment list as **2 exps lovastatin & ethinylestradiol**
Step 5 Open the **TOOLBOX** and select the **PATHWAY VISUALIZATION** tool from the **DATA VISUALIZATION** section.
Step 6 Drag the **2 exps lovastatin & ethinylestradiol** list from the **WORKSPACE** into the **EXPRESSION EXPERIMENT LIST** field.
Step 7 Drag the **cholesterol biosynthesis** pathway **favorites** list from the **WORKSPACE** into the **PATHWAY FAVORITE** field.
Step 8 Click the **DISPLAY** button.

Exercise 9 ~ Part 6 ~ Steps 5 thru 8

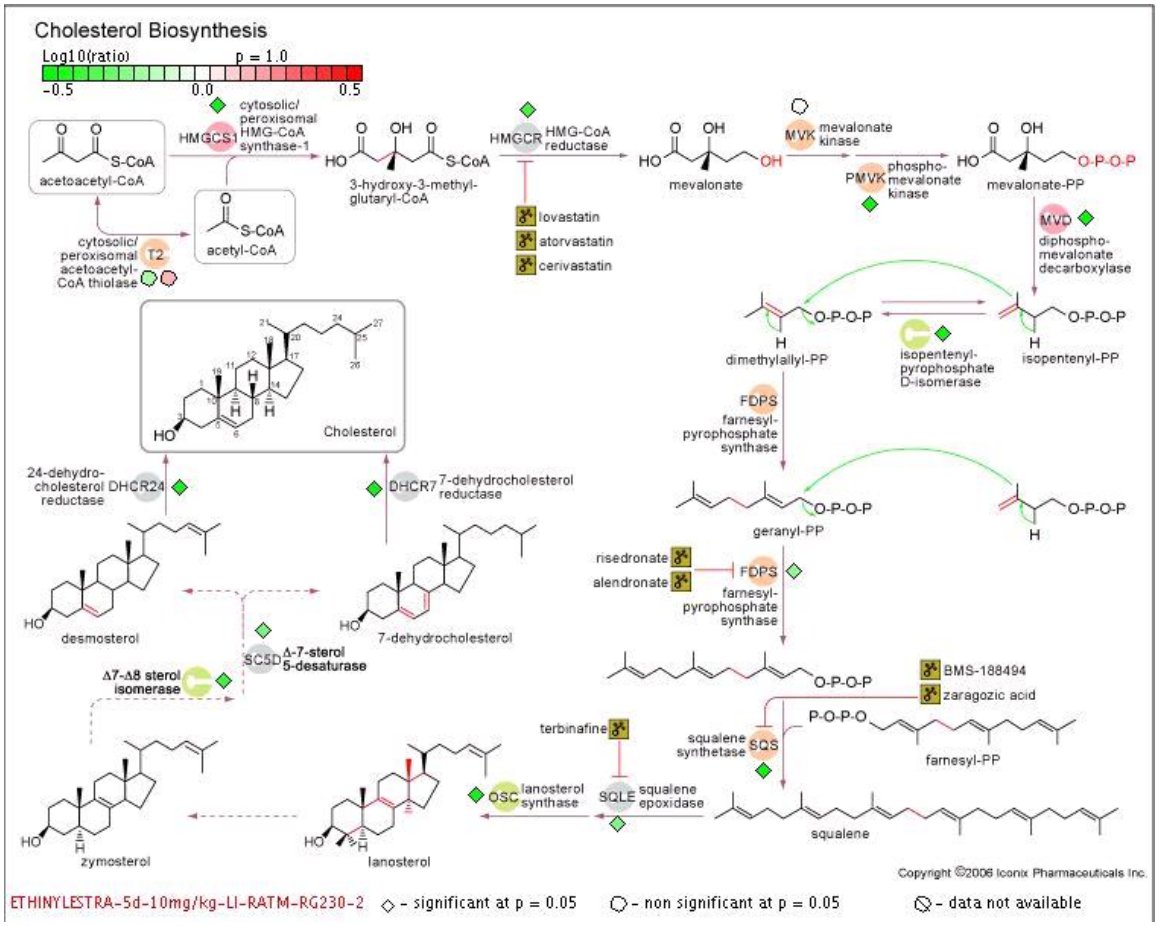


Flip through **LOVASTATIN** and **ETHINYLESTRADIOL** experiments, see the different effects of the two compounds on cholesterol biosynthesis pathway.

Exercise 9 ~ Part 6 ~ Pathway Visualization DATA VIEW for LOVASTATIN



Exercise 9 ~ Part 6 ~ Pathway Visualization DATA VIEW for ETHINYLESTRADIOL



You will see every gene has been up-regulated in this pathway in response to the HMG-CoA reductase inhibitor Lovastatin, and down-regulated in response to ethinylestradiol. This suggests that the genes in this pathway are highly co-regulated in response to change in serum cholesterol levels and drug treatment.

Conclusions to Exercise 9

We have navigated through the **COMPOUND**, **ASSAY**, **EXPRESSION**, **PATHWAY** domains, used **ADVANCE QUERY** tool to extract the desired data, used some **DATA TRANSLATION**, **DATA VISUALIZATON** tools to explore and analyze the data. At the same time, we explored cholesterol lowering mechanisms.