Public Training: Fundamentals

October 8, 2010

Recombinant

Table of Contents

Logging In to tranSMART Training

For this training, you will use a training server that is isolated from the "real-world" tranSMART environment. The login credentials and login address that you will use for this training apply to the training server only.

Login Credentials

Login credentials are as follows:

- ID: **publicuser***xx*

  where *xx* is a 1- or 2-digit number that the instructor will assign you.

  For example, **publicuser5** or **publicuser21**.

- PWD: **training**

Credentials are case-sensitive.

Login Address

Please use the following address to log in to the training server:

http://75.101.162.195/transmart

Application and Data Differences

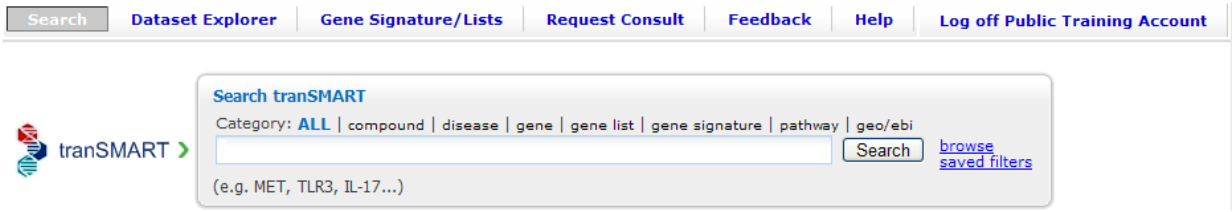
Due to periodic updates to the application and data on the training server, the figures and specific data references in this tutorial may be different than what you see on your screen.

> **Note:** tranSMART includes features and access to data that are not available with the training version of tranSMART.

Lesson 1

What You Can Do with tranSMART

Lesson Goal: Become acquainted with the tools on the tranSMART toolbar.



These are the high-level tasks you can perform using the following tranSMART tools:

- **Search** – Search across internal and external data sources for research data and literature related to one or more search terms that you specify. For example, if MET is a gene of interest to you, you can run a single query with MET as your search term. tranSMART will search studies and experiments, gene signatures, pathways, literature, and other areas to find instances ("hits") where the MET gene is a factor.

- **Dataset Explorer** – Test a hypothesis by analyzing study data generated for one or two study groups that you define, using criteria and points of comparison that you specify. For example, you could define cohorts in an asthma study – one who contracted asthma as children, and the other who contracted asthma as adults. You could then introduce a point of comparison – for example, a tendency towards other atopic allergies – to determine whether one cohort is more inclined towards atopic allergies than the other.

- **Gene Signature/Lists** – Generate a hypothesis by creating a gene signature. Instruction on creating gene signatures is included in the tranSMART Advanced training.

- **Request Consult** – Email a request for information that cannot be found in the data warehouse.

- **Feedback** – Email questions, problem reports, enhancement requests, or any other feedback about the tranSMART application.

- **Help** – Display links to the tranSMART documentation. For training purposes, **Help** links to the Basic and Advanced training guides only.

Lesson 2

Conducting a tranSMART Search

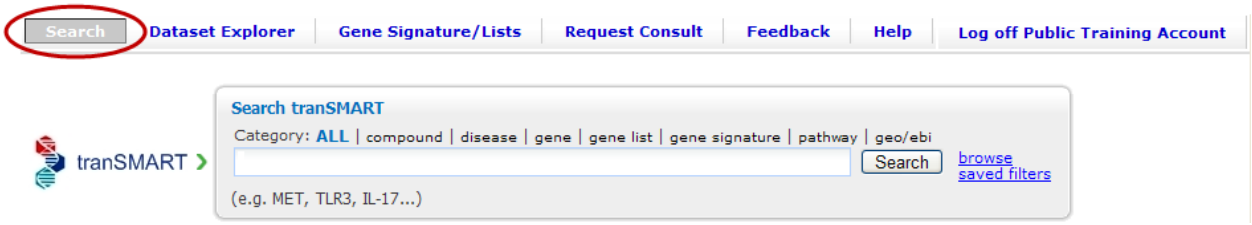
Lesson Goal: Learn how to define a search using the autocomplete, browse, and free-text search features. Also learn how to define, save, and delete a multi-filter search string.

Autocomplete

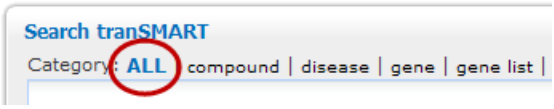
Description of the autocomplete feature: tranSMART finds pre-defined search terms based on a partial search term that you type into the search field. When you stop typing, tranSMART displays a list of up to 20 search terms that are possible matches for the characters you typed.

Scenario: You want information about pathways for histidine biosynthesis.

1. If the tranSMART Search tool is not displayed, as shown below, click the **Search** tab to display it:



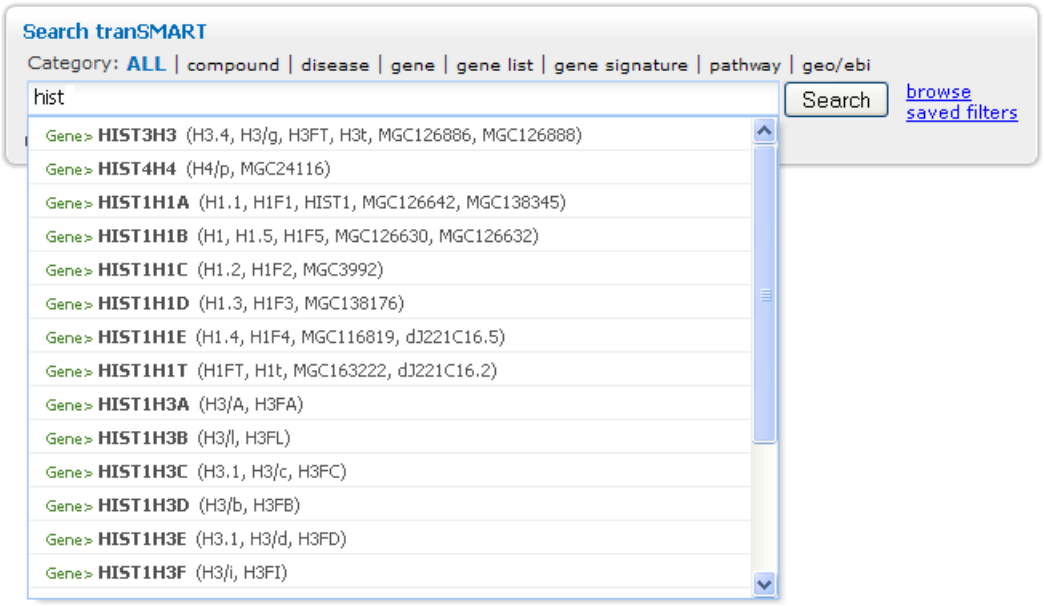
   When the search window is first displayed, the selected search category is **ALL**, meaning that tranSMART will search across all search categories (compound, disease, gene, etc.):



2. Type **hist** into the search field, then pause. *Do not click the Search button or press the Enter key*.

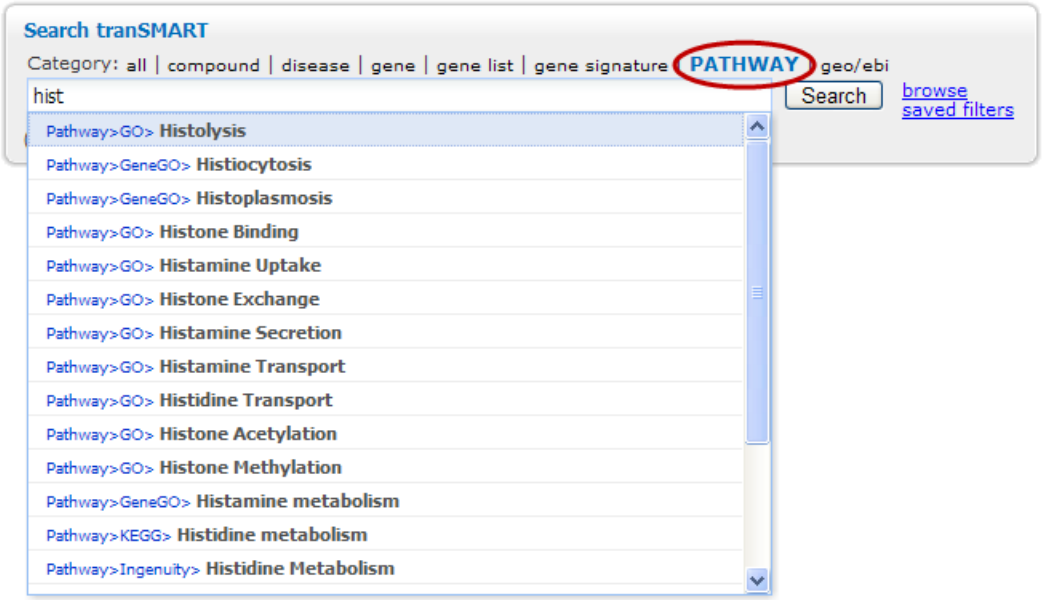
   When you stop typing, the autocomplete feature displays a dropdown list of possible matches for the characters you typed. The list contains up to 20 search terms (called *filters*). Each filter represents a biomedical concept (compound, disease, gene, gene list, gene signature, or pathway) or a European Bioinformatics Institute reference to a study.

When the list of filters appears, you see that the list does not contain a term like histidine biosynthesis:



3. To narrow your search to just pathways, click **pathway** above the search field.

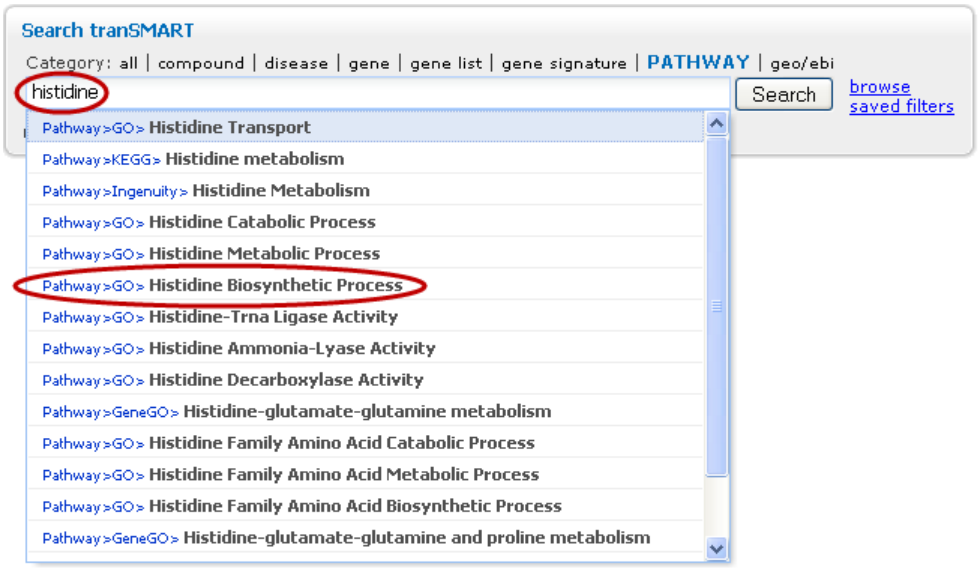
   The list now contains only pathways, but still has nothing like histidine biosynthesis:



   Because the dropdown list is limited to 20 search terms, it's possible that there are more matches for the text you typed than can appear in the dropdown list.

4. To reduce the number of hits, you can add more characters to the text you typed. Type the characters **idine** after the text in the search field, then wait for the dropdown list to appear.

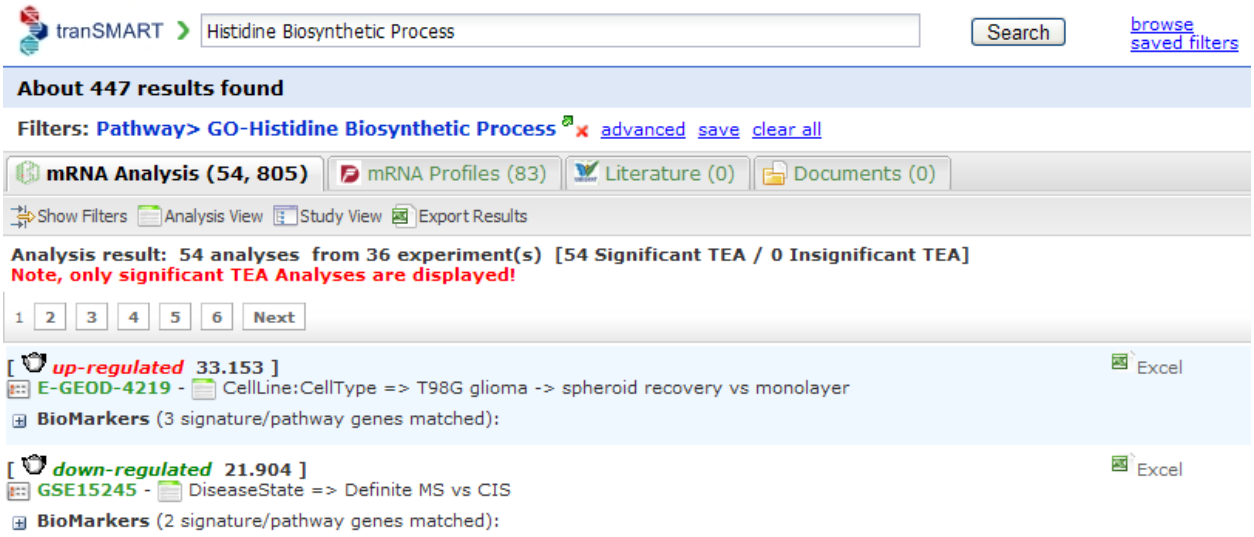
   The filter **Histidine Biosynthetic Process** now appears in the list:



5. Click **Pathway>GO> Histidine Biosynthetic Process**.

   Again, don't click the **Search** button or press **Enter**. After you click a filter in the list, tranSMART begins searching immediately, using the selected search filter, and returns a search result.

   Part of the result is shown below:

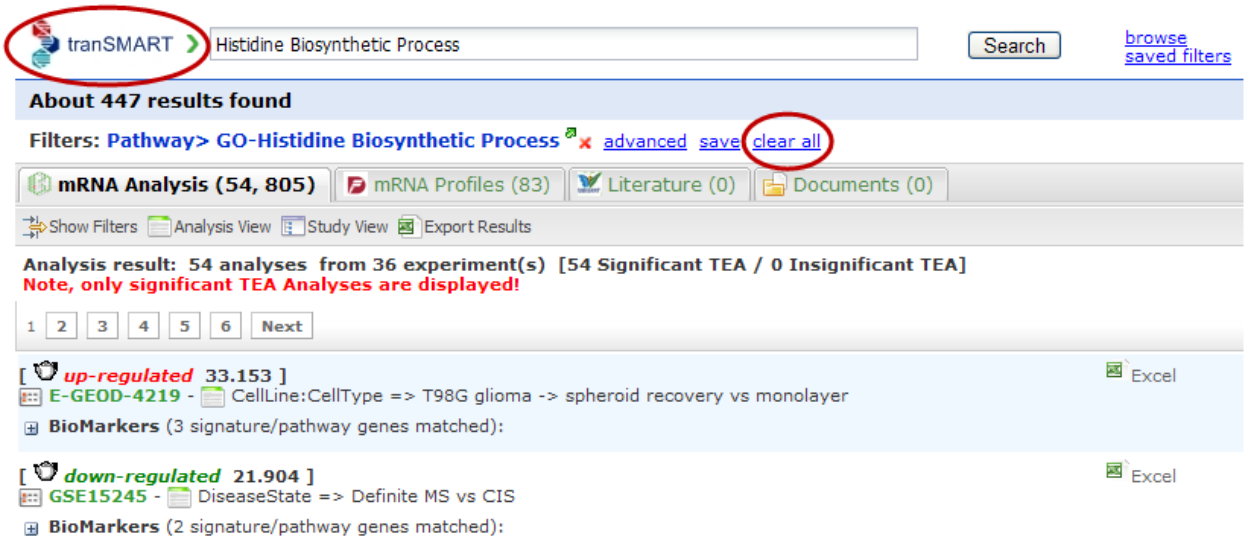


Browse

Description of the browse feature: tranSMART displays lists of pre-defined search terms to let you select the one you want.

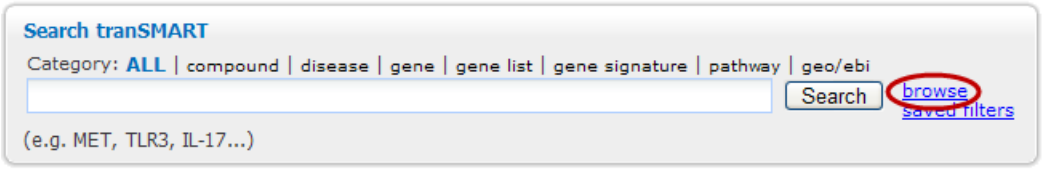
Scenario: You want information about pathways for histidine.

1. To remove the results from the last lesson and start a new search, click the **tranSMART** logo or the **clear all** button:



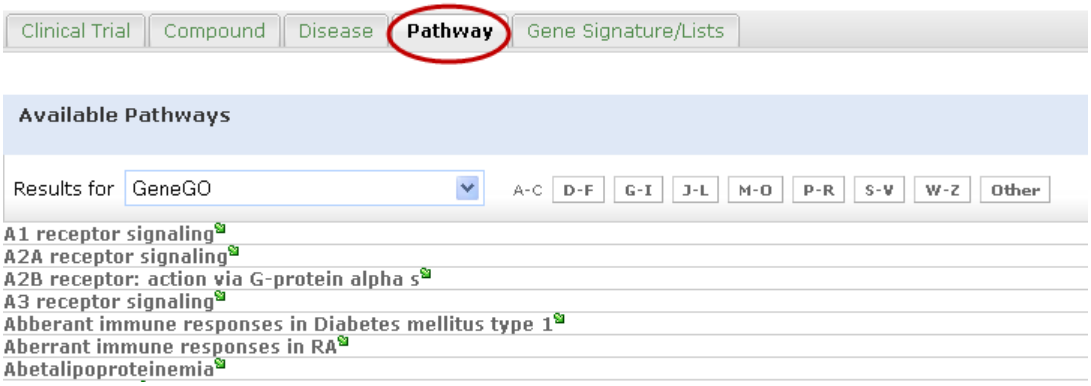
Note: Your training application may have the following logo in the place of the tranSMART logo shown above:

2. Click **browse** to the right of the search field:



Note: The search categories above the search field have no effect on browsing.

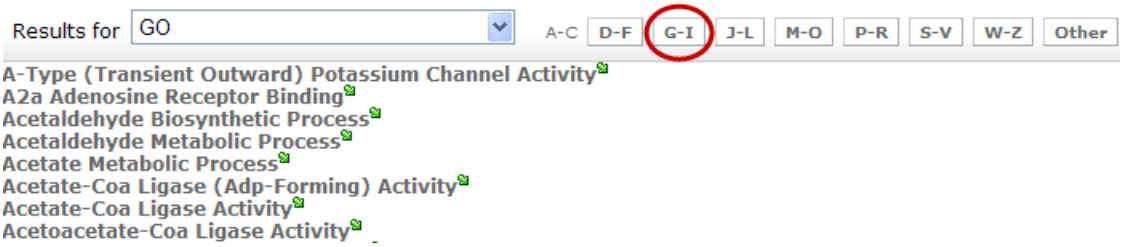
3. In the browse window, click the **Pathway** tab:



4. In the **Results for** dropdown, select **GO**.

   A list of available filters whose names begin with the letters A, B, or C appears in the GeneGO pathway scope:

5. Click the **G-I** button to display that alphabetical portion of the list:



6. Scroll down until **Histidine Biosynthetic Process** appears in the list.

7. Click the filter name **Histidine Biosynthetic Process** or the green arrow after it. The search begins immediately, and a search result is returned.

Free Text

Description of the free text feature: tranSMART searches through proprietary document repositories for the exact text you type in the search field (not case-sensitive).

> **Note:** On the training server, public documents are substituted for the proprietary documents.

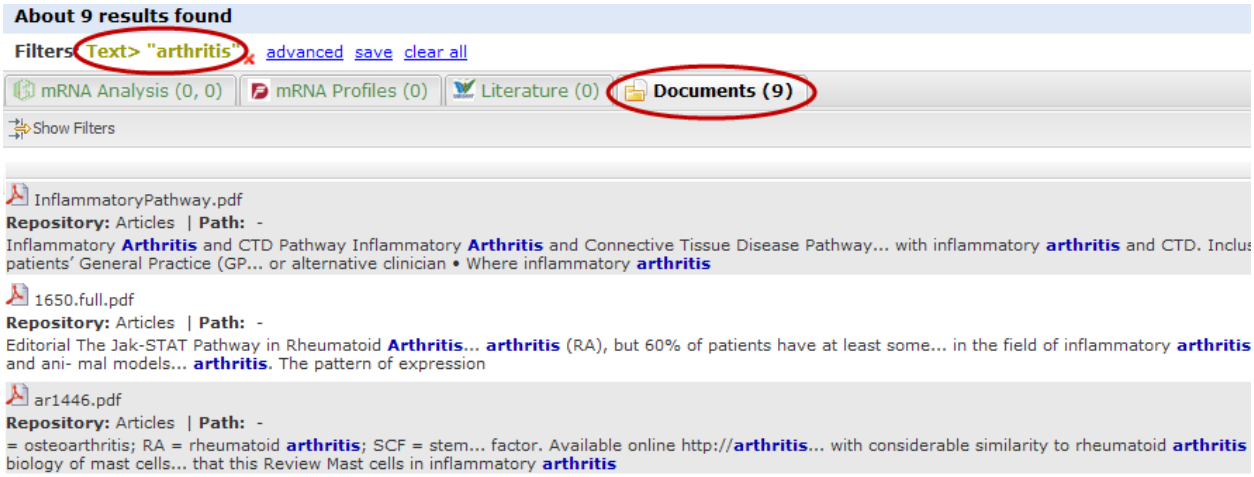
Scenario: You want to search the proprietary documents for all articles that contain the text **arthritis**.

1. Click the **tranSMART** logo to begin a new search.

2. Type **arthritis** in the search field.

   Type with care. The free text feature searches for the exact text you type.

3. Click the **Search** button. The free-text search begins immediately, and a search result is returned:

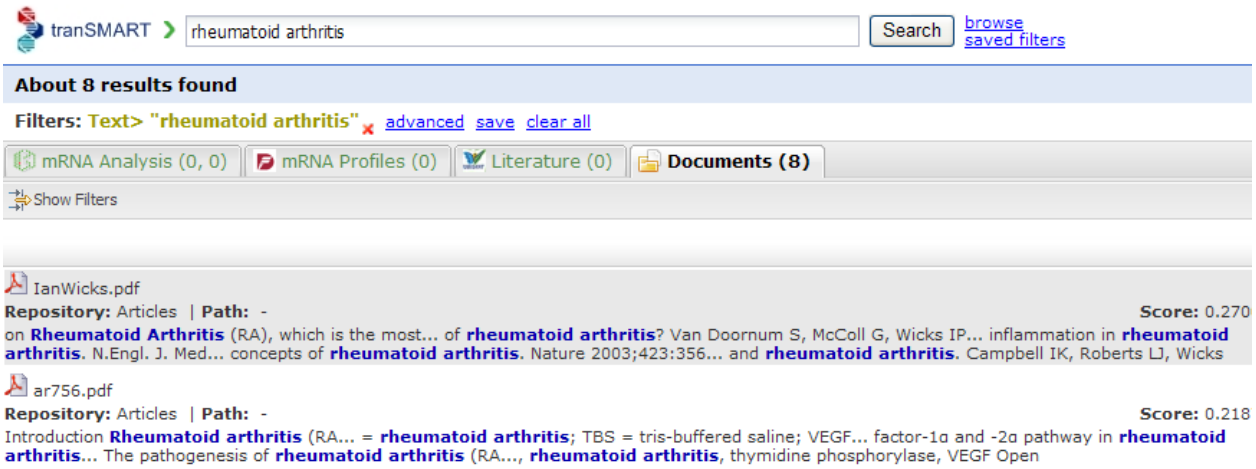


Note: Clicking the **Search** button limits the search to the proprietary document repositories only**.** Records involving mRNA analyses, mRNA profiles, and other publications available through the **Literature** tab are not searched.

The search term **arthritis** will return references to all proprietary documents that contain that word. However, your particular interest is rheumatoid arthritis.

4. In the search field, type **rheumatoid** and a space in front of the word **arthritis**.
5. Click the **Search** button.

The results now contain references to documents that contain the phrase **rheumatoid arthritis**:

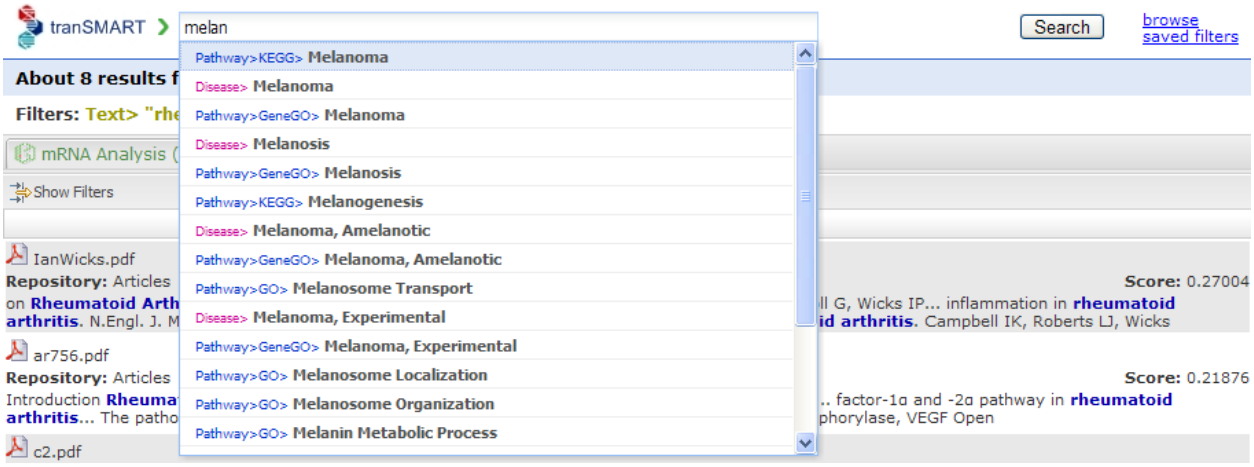


Note: There is no autocomplete feature with free-text searches. When you click the **Search** button or press the **Enter** key, tranSMART searches for the exact text in the search field.

Multi-Filter Searches

Scenario: You want information about associations between the KEGG melanoma pathway and melanoma or lung cancer.

1. Type **melan** into the search field, overwriting all other text that may be there from the previous lesson:



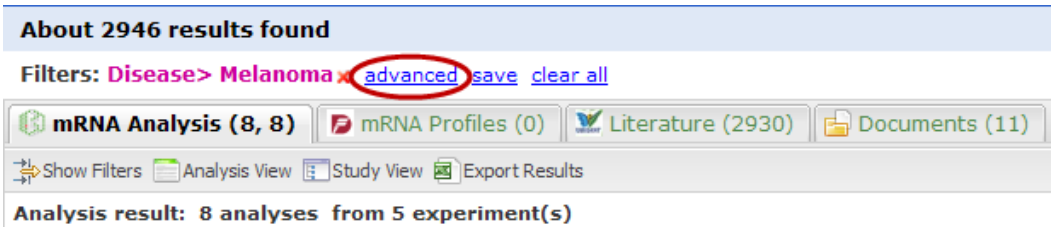
2. Click the search filter **Disease> Melanoma**.

   The previous search results are removed. The new search begins immediately, and a search result is returned.

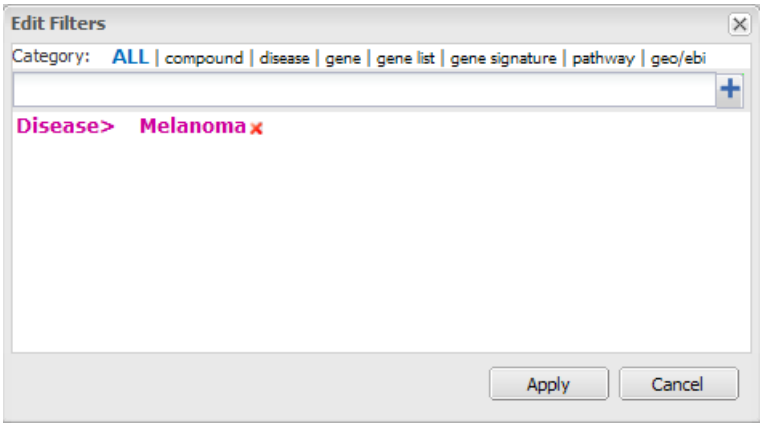
   You now want to add a second filter, lung cancer, to the search string.

> **Note:** In a multi-filter search, only the first filter can be specified in the search field of the search window. If you specify another filter in the field and then run the search, a new search begins, using only the most recently specified filter. The following steps show the correct way to add additional filters for a multi-filter search.

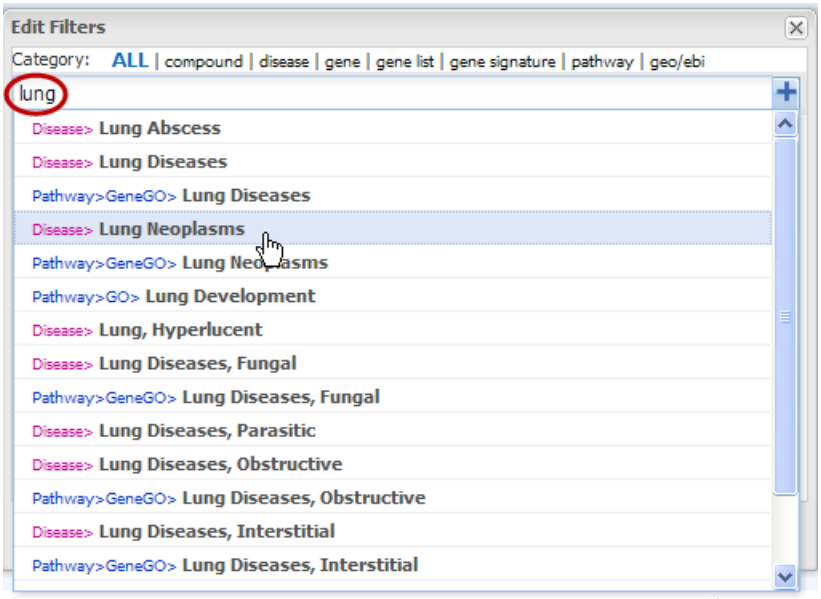
3. Click the **advanced** button above the search results:



The Edit Filters dialog appears. This dialog lets you add and remove individual filters in the multi-filter search string:

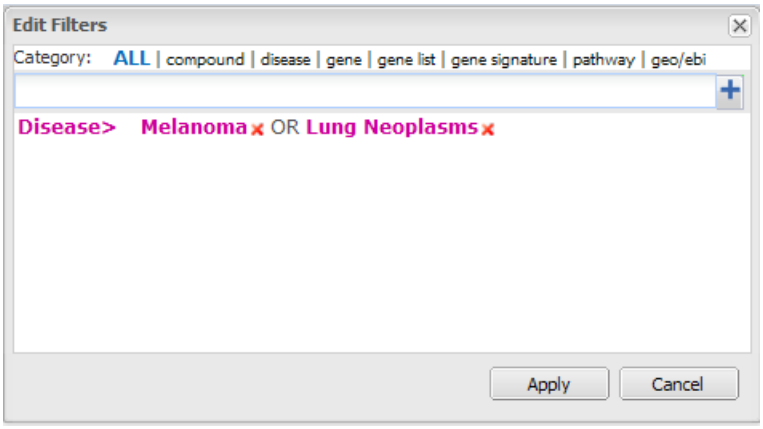


4. Type **lung** in the search field at the top of the Edit Filters dialog, then click **Disease> Lung Neoplasms** when the list of filters appears:



Note: The search does not start immediately when you click a filter in this dropdown list. The new filter is added to the Edit Dialog, where you can continue adding and removing filters.

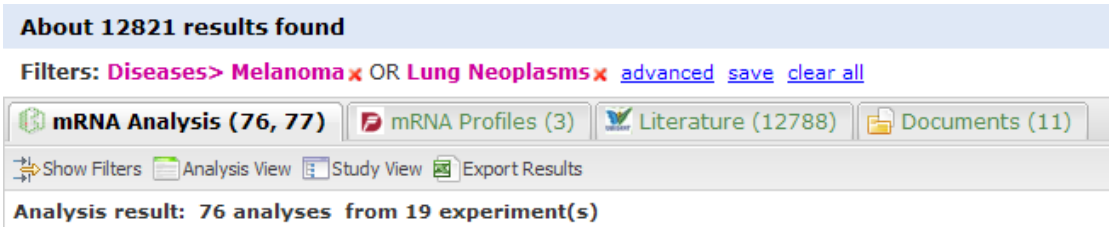
After you click **Disease> Lung Neoplasms** in the dropdown list, the Edit Filters dialog looks as follows:



Notice that two search filters in the same search category (in this case, Disease) are joined by the logical operator `OR`.

5. Click the **Apply** button to begin the search.

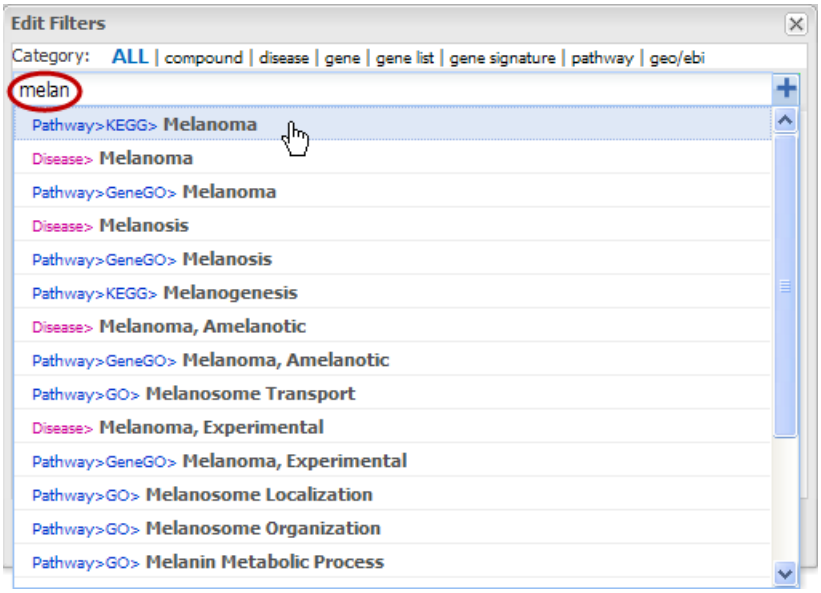
   With the search now based on two search filters, there are substantially more hits than when the search was based on the single search filter **Disease> Melanoma**:



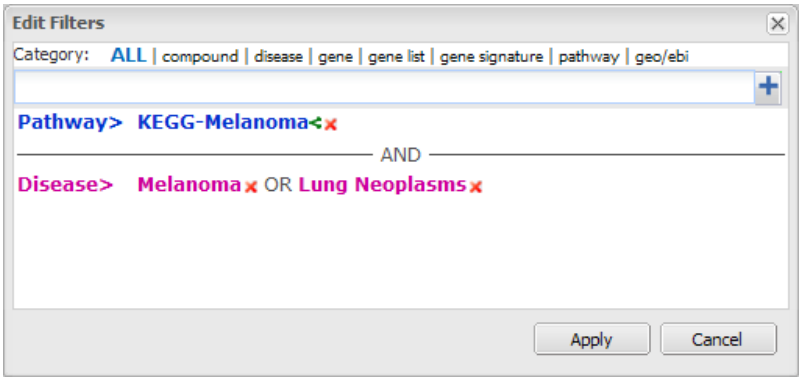
   But the search results are now too broad. You want to limit the results to records associated with melanoma or lung cancer, *and* the melanoma pathway.

6. Click the **advanced** button.

7. Type **melan** in the search field, then click **Pathway>KEGG> Melanoma** when the list of filters appears:



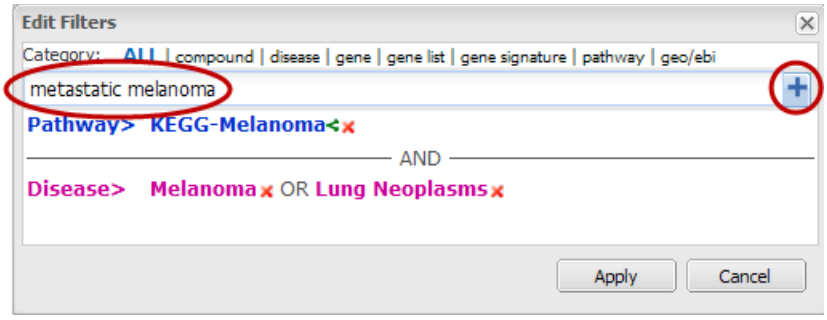
The Edit Filters dialog now looks as follows:



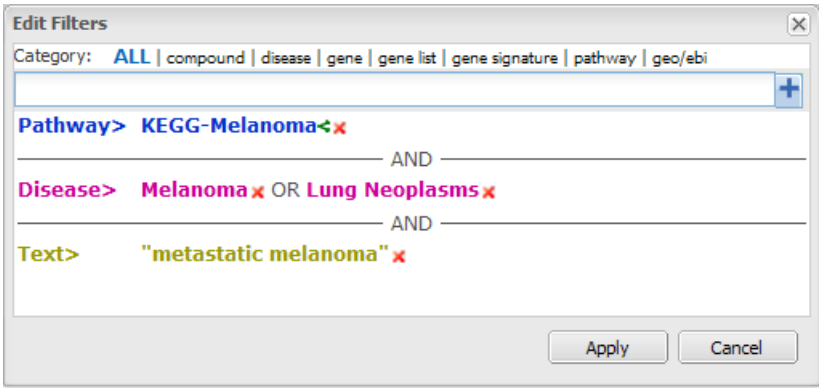
Notice that two search filters in different search categories (in this case, Disease and Pathway) are joined by the logical operator `AND`.

Finally, you want tranSMART to search proprietary document repositories for the search text **metastatic melanoma**.

8. Type **metastatic melanoma** in the search field, then click the **+** icon ( ) to the right of the field:



The multi-filter search string is now complete:

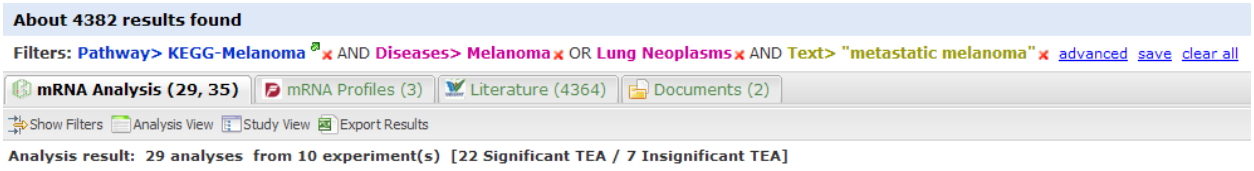


The search string can be expressed as follows:

```
Pathway>KEGG-Melanoma AND (Disease>Melanoma OR Disease>Lung Neoplasms) AND
Text>"metastatic melanoma"
```

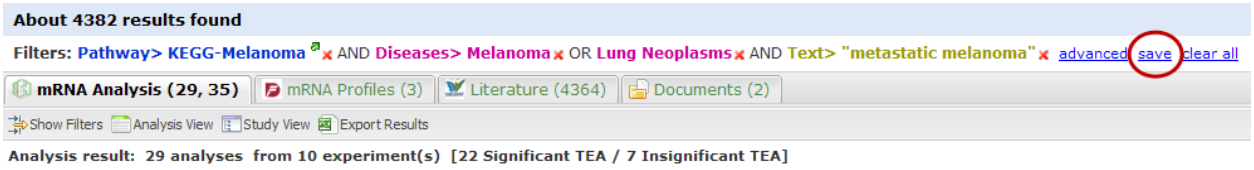
9. Click the **Apply** button to begin the search.

The following figure shows that 4382 records satisfied the search string:

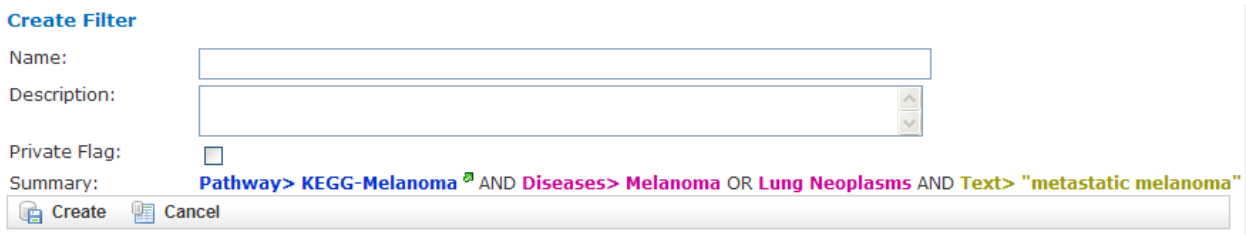


Now you decide to save the multi-filter search string.

10. Click **save**:



The Create Filter window appears:



11. Type **Melanoma** in the **Name** field.
12. Select the **Private Flag** checkbox.

    If **Private Flag** is cleared, other users can run the saved search if you email it to them.
13. Click **Create**.
14. The private search string now appears in your Saved Filters list.
15. Click **Return to Search**, located after the Saved Filters list.

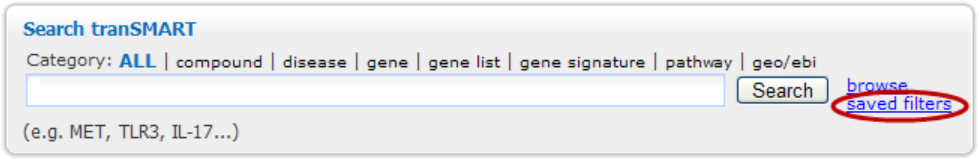
    The search window reappears, ready for a new search.

Note: To retrieve a saved filter, click **saved filters** on the tranSMART search window.

Deleting a Saved Search Filter

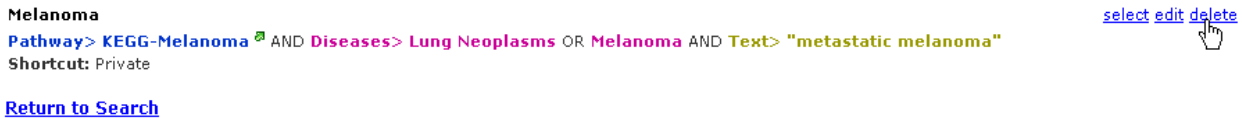
To delete a saved search filter:

1. In the tranSMART Search window, click **saved filters**:



   The Search > Saved Filters window appears.

2. Click **delete** for the saved filter to delete (**Melanoma** in this example):



3. Click **OK** to confirm the deletion.
4. Click **Return to Search** to prepare for the next lesson.

Lesson 3

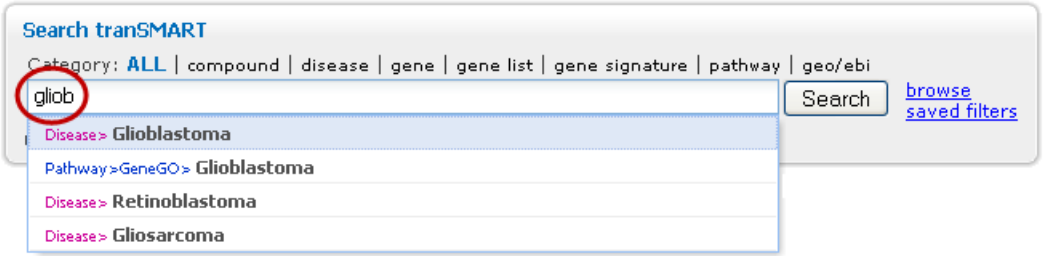
Searching Analyses of mRNA Experiments

Lesson Goal: Understand and use mRNA analysis search results.

Scenario: You want to review analyses of studies involving glioblastoma, and particularly those in which genes in the receptor signaling protein activity pathway were modulated.

1. In the tranSMART search window, type **gliob** in the search field.

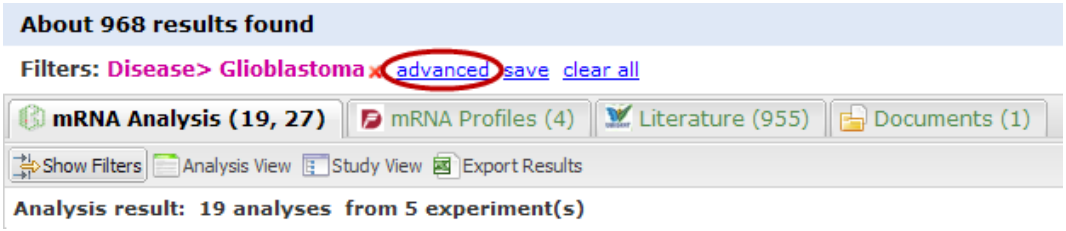
   The following dropdown list appears:



2. Click **Disease> Glioblastoma** in the dropdown list.

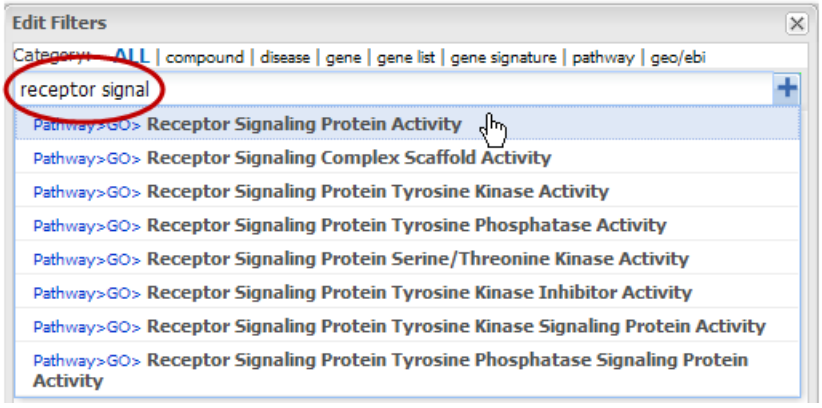
   The figure in the following step shows that 968 records of various types were returned – mRNA analyses, mRNA profiles, and references to public literature and internal documents – in which glioblastoma is a factor.

3. Click the **advanced** button:



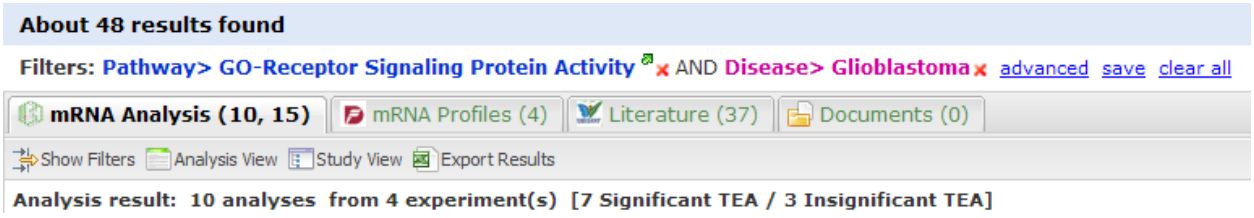
   The Edit Filters dialog appears.

4. Type **receptor signal** in the text box, then click **Pathway>GO>Receptor Signaling Protein Activity**:



5. Click **Apply**.

   In the figure below, the search returned just 48 records in which both glioblastoma and one or more genes in the pathway are factors:



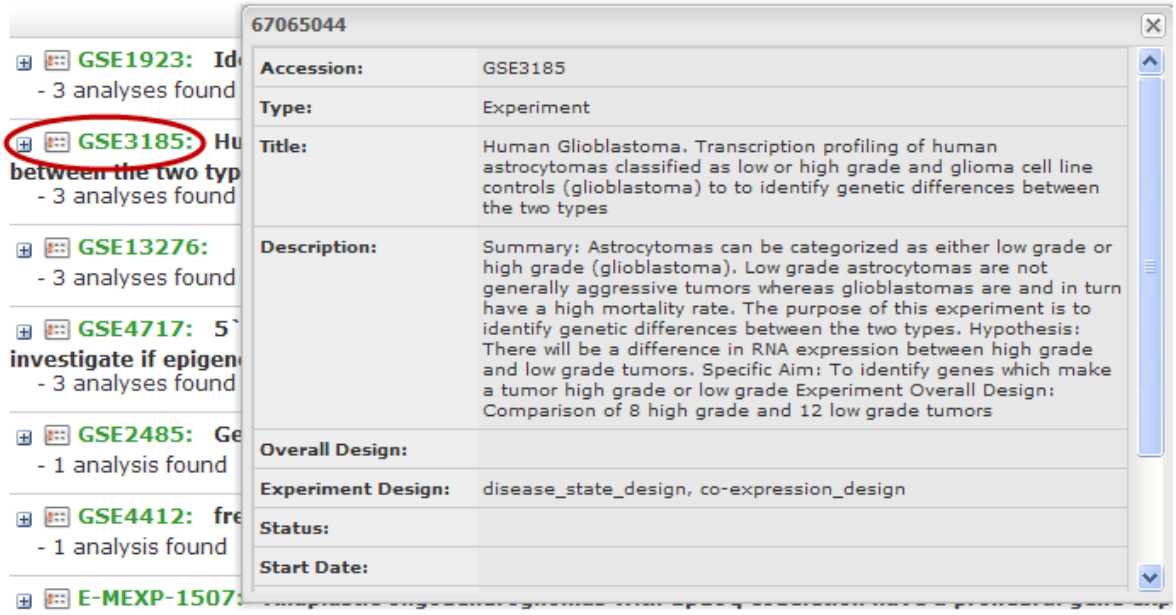
   The mRNA Analysis search result is selected by default. Notice the four buttons below the **mRNA Analysis** tab – Show Filters, Analysis View, Study View, and Export Results. All will be described in this lesson.

6. Click the **Study View** button.

   This view lists all studies in which glioblastoma and the pathway are factors.

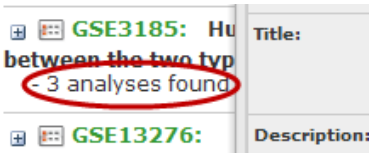
7. Click the name of study **GSE3185** (or hover the mouse pointer over the name).

A details box appears, summarizing the study:



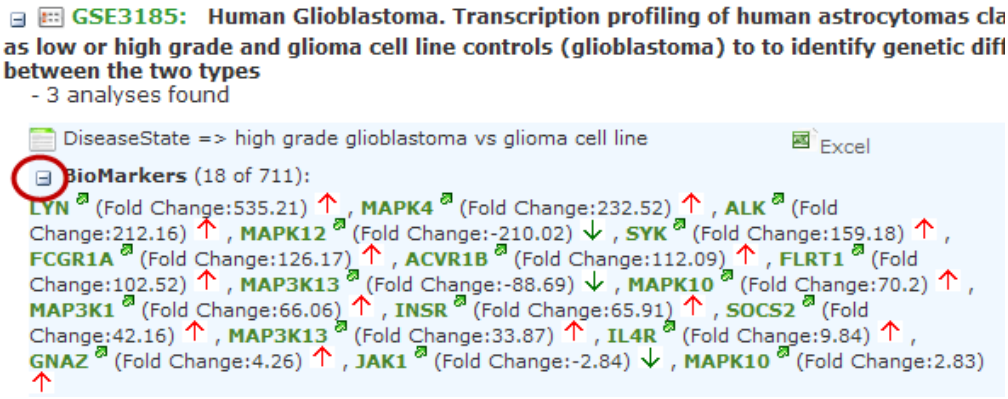
The study interests you. You decide to look at the analyses of the study.

Note: A study can have multiple analyses that satisfy the search criteria (glioblastoma and the pathway, in this case). This study has three:



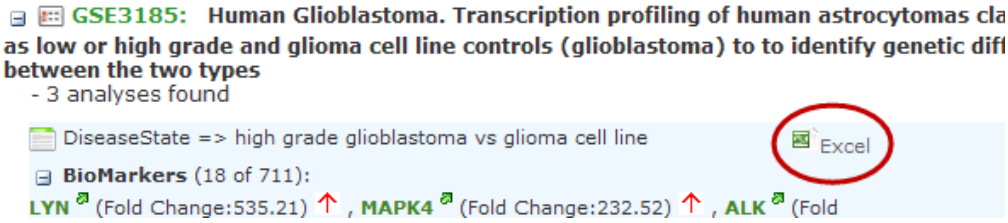
8. Click the **X** icon ( ) in the upper right corner of the details box to close it.
9. Click the **+** icon ( ) to the left of the study **GSE3185** to list the analyses.

10. Click the **+** icon (⊞) to the left of the first **BioMarkers** label to open a list of matching genes in the analysis (the + icon changes to a – icon after you click it):



In this figure, 18 genes referenced by the search criteria matched 18 of the 711 genes included in this analysis.

11. To see all 711 genes that were included in the analysis, click **Excel** to the right of the analysis title:



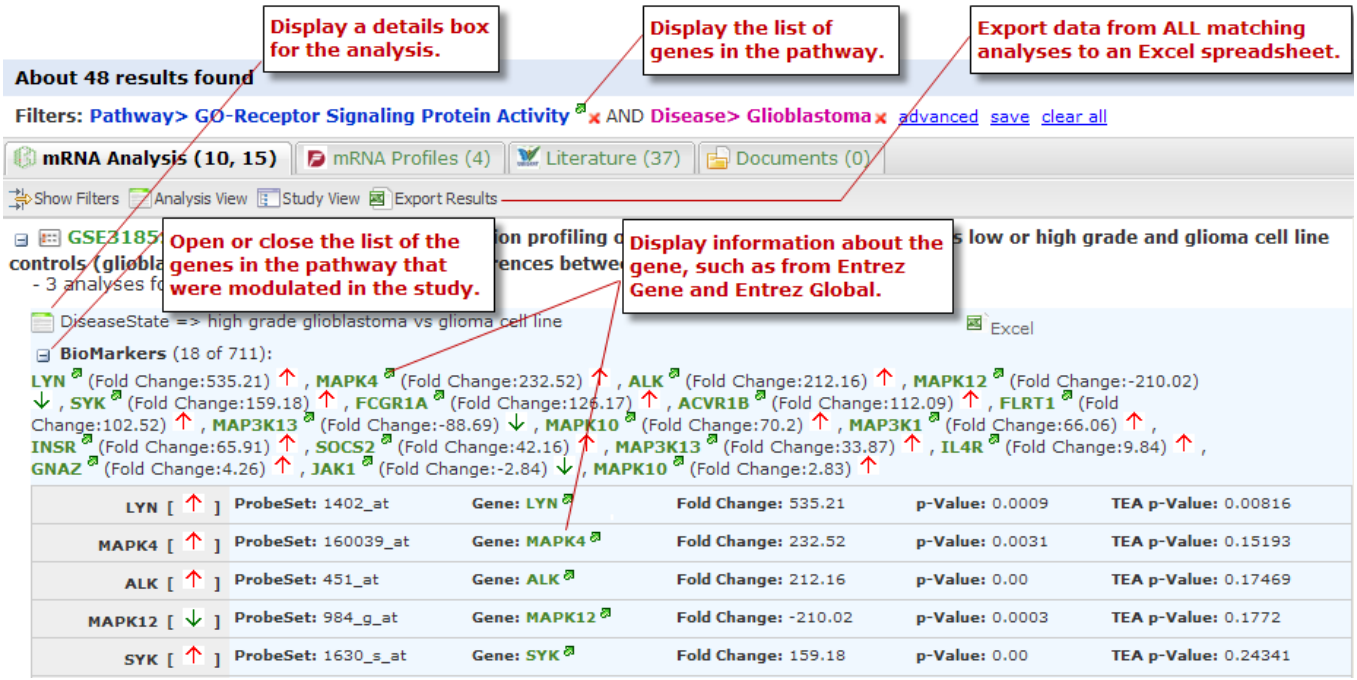
12. When prompted to either open or save the spreadsheet, click **Open**.

    You can now view fold change, p-value, and other data for all the genes included in the analysis, including the 18 matching genes.

    Note that the number of entries in the spreadsheet is higher than the number of genes in the analysis. That's because the Export function writes one row of data for each *probe set*, not each gene, and the same gene may be associated with multiple probe sets.

13. Close the spreadsheet without saving it.

14. Try some of the other actions that are available in Study View:



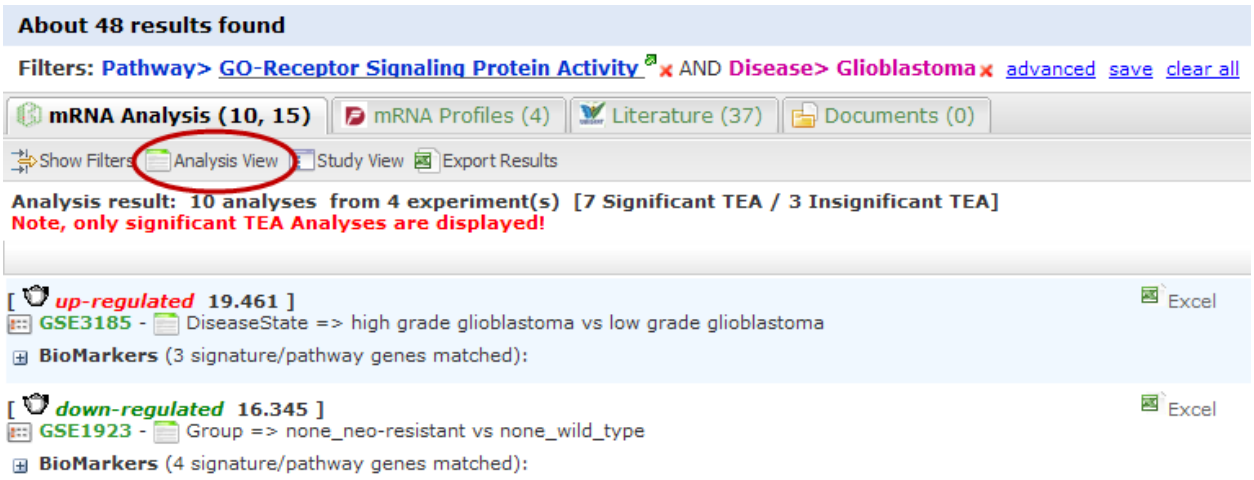
Study View displays all the studies that satisfy the search criteria, and all the analyses within each study that also match the criteria. There is no attempt to limit the view to just those analyses that are considered statistically significant.

Statistically significant analyses are displayed in the Analysis View.

Note: An analysis is determined to be "statistically significant" if one or more genes in the analysis has a normalized p-value of less than 0.05. In the figure above, the gene LYN is the only one that meets this threshold (TEA p-Value: 0.00816).

15. Click the **Analysis View** button.

    A partial display of the statistically significant analyses is shown below:



16. Notice the numbers in parentheses on the **mRNA Analysis** tab (10 and 15 in the figure above).

    The first number indicates the number of statistically significant analyses that were returned in the search (ten in this case). The second number indicates the total number of analyses returned.

17. Notice also that although there are ten statistically significant analyses, only seven are included in the Analysis View. That's because there are two levels of what is considered a "statistically significant" result, as indicated by the circled text below:

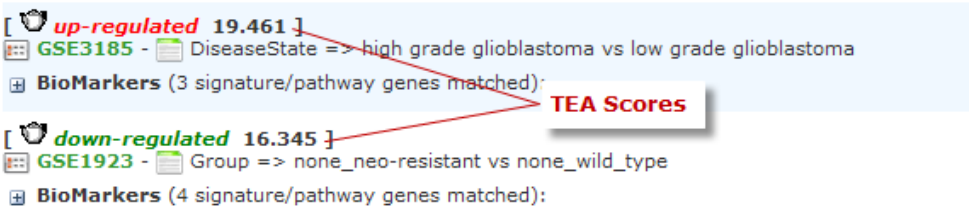
    These two levels of "statistically significant" analyses are based on the Target Enrichment Analysis (TEA) score of each analysis:

    - A TEA score of 2.9957 and above indicates a *significant* TEA analysis.
    - A TEA score of below 2.9957 indicates an *insignificant* TEA analysis.

    A **TEA score** is a binomial distribution of normalized p-values, calculated in the context of the following factors:

    - **With gene signatures and gene lists** – The level of co-regulation or anti-regulation of the genes within the gene signature or gene list, as compared with the experiment.
    - **With pathways** – The level of up-regulation or down-regulation of the genes within the pathway, as compared with the experiment.

TEA scores in the Analysis View appear as shown below:



You are interested in studies that focused on disease state, so you want to limit the analysis results to just those studies.

18. Click the **Show Filters** button.

    A form is displayed, allowing you to define search filters to narrow the search results.

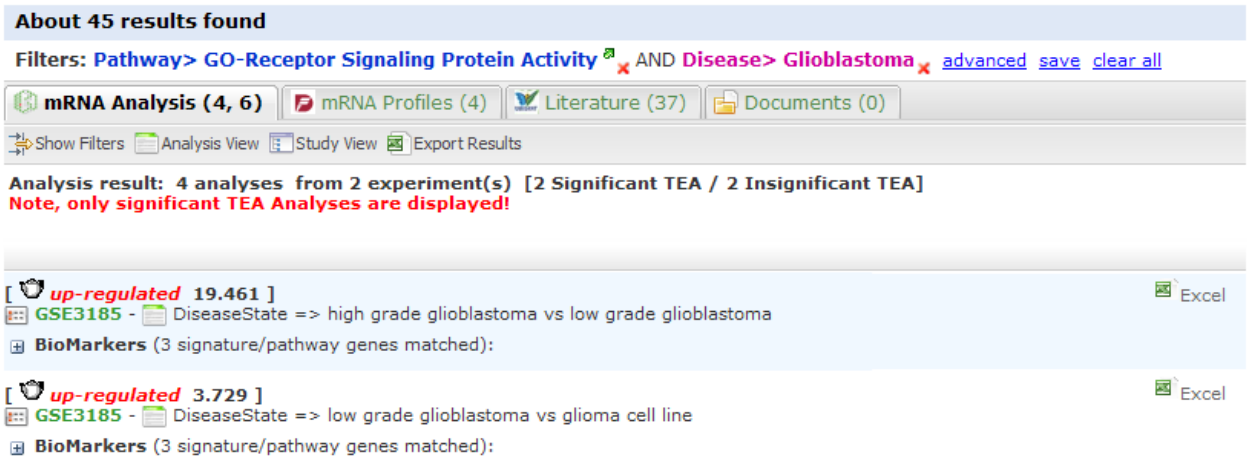
19. In the **Platform Species** field, select **Homo sapiens**.

20. In the **Disease** field, select **Glioblastoma**.

21. In the **Experimental Design** field, select **Disease State Design**.

22. Click **Filter Results**.

    The Analysis View now contains just two statistically significant analyses:



Note: To view all the studies that focused on disease state, whether or not the analyses in those studies are considered statistically significant, click **Study View**.

Lesson 4

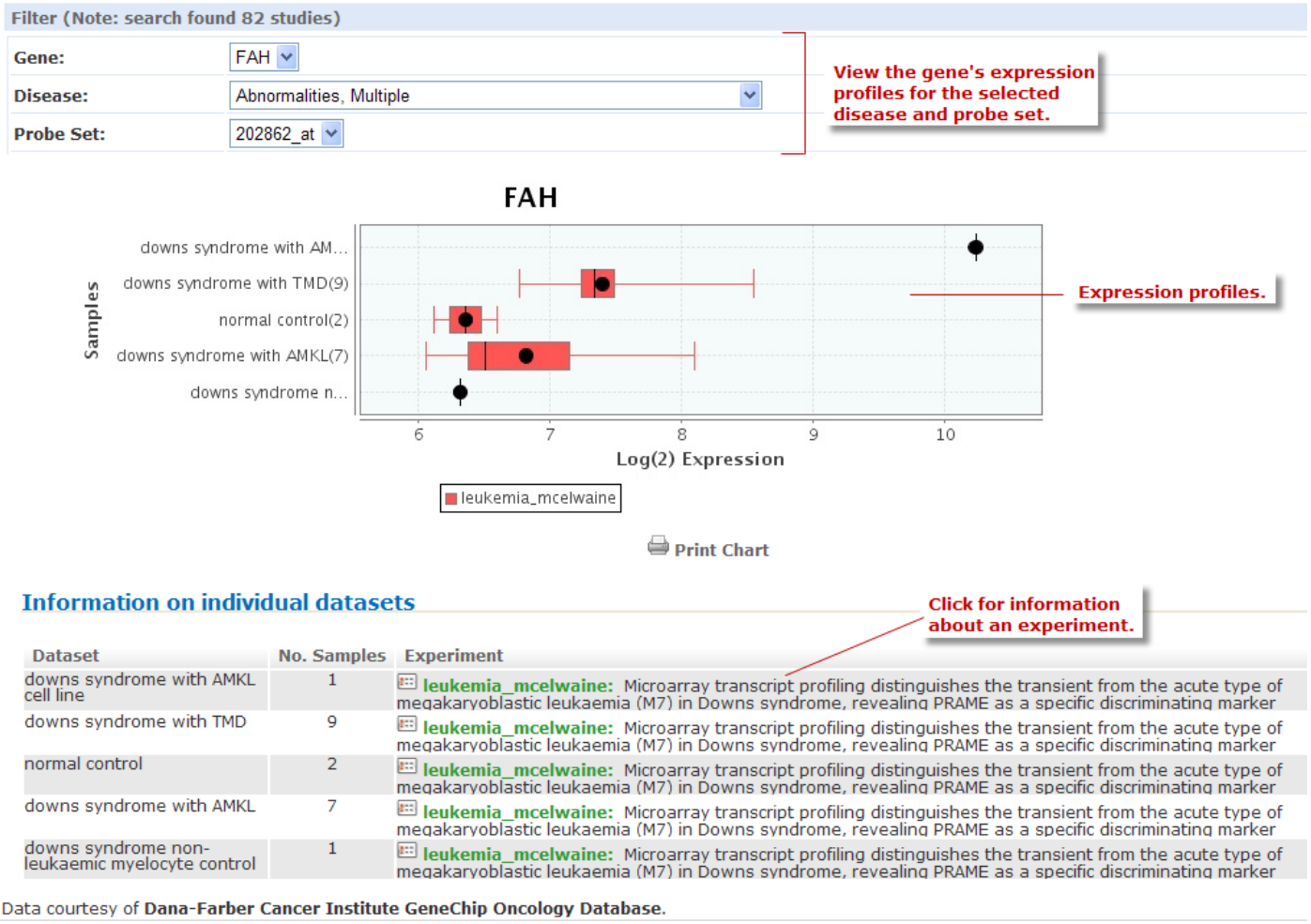
Searching Expression Profiles for Genes of Interest

Lesson Goal: Find gene expression profile data for a particular type of disease and probe set.

Scenario: You want to view different expression profiles for the FAH gene in relation to lung neoplasms.

1. Click the **tranSMART** logo to clear the results from the previous lesson.
2. Type **fah** in the search field, then click **Gene> FAH** in the dropdown list.
3. When the search results appear, click the **mRNA Profiles** tab.

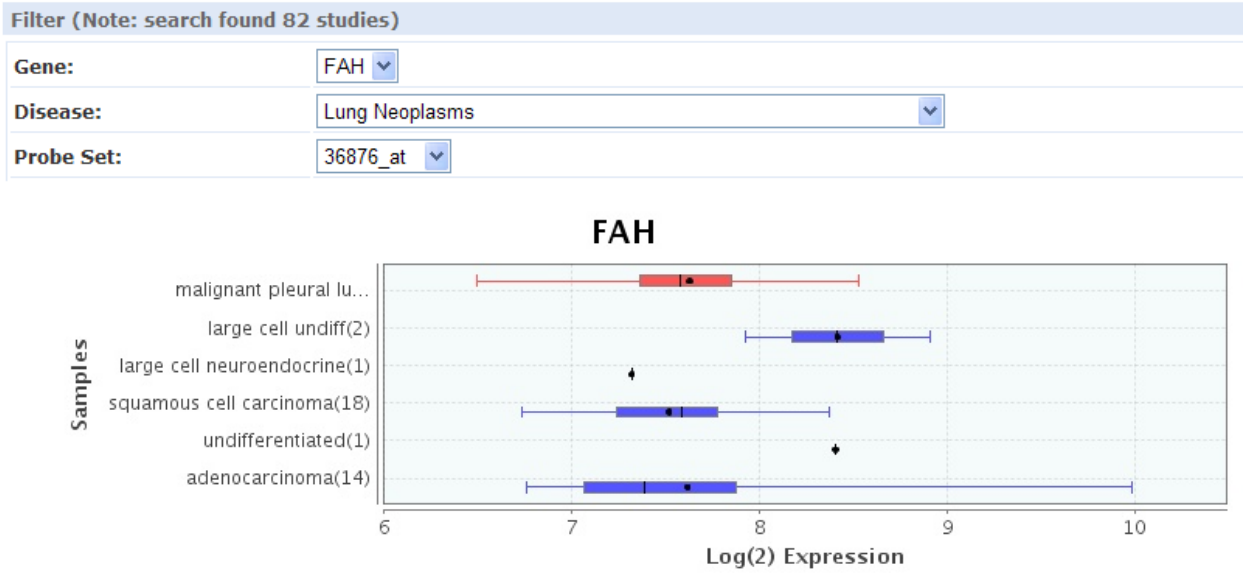
   The default mRNA Profiles search result for the FAH gene is displayed as follows:



4. Select **Lung Neoplasms** in the **Disease** dropdown.

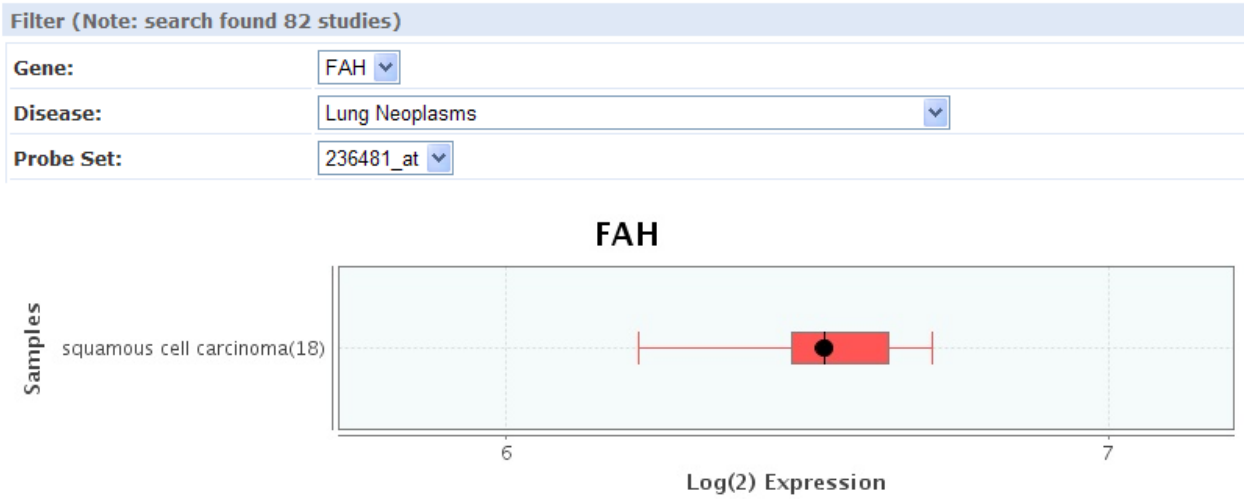
   After the display refreshes, notice that the probe set name changes to one that is appropriate to tests of lung neoplasms.

   Review the gene expressions that now appear in the box plot:



5. Select **236481_at** in the **Probe Set** dropdown.

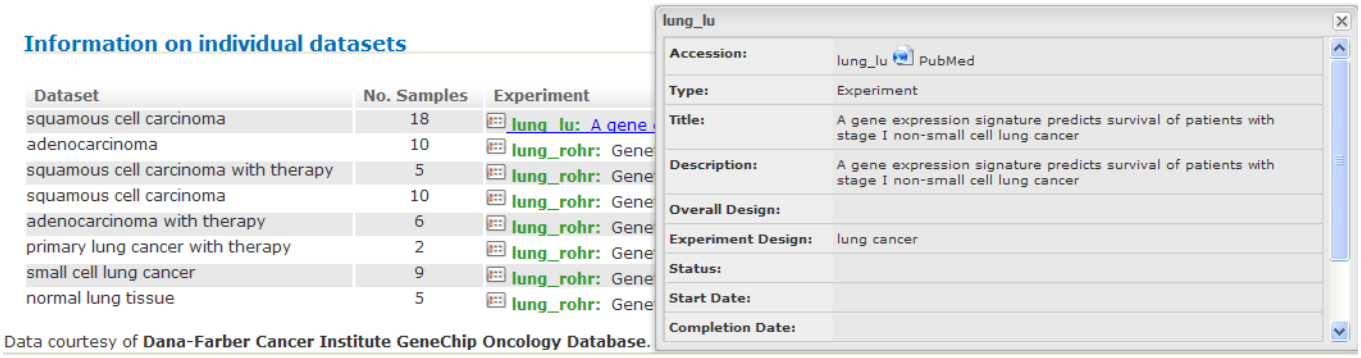
   Notice the change in the gene expressions associated with this probe set:



6. Repeat the previous step for the last probe set in the list, **202862_at**.

7. In the **Information on individual datasets** section, click the name of the first experiment in the list.

   A details box appears, describing the experiment:



8. Click the **PubMed** link in the details box to view relevant published information in a new browser window.

9. When finished, close the browser window, then close the details box.

Lesson 5

Searching Curated Literature

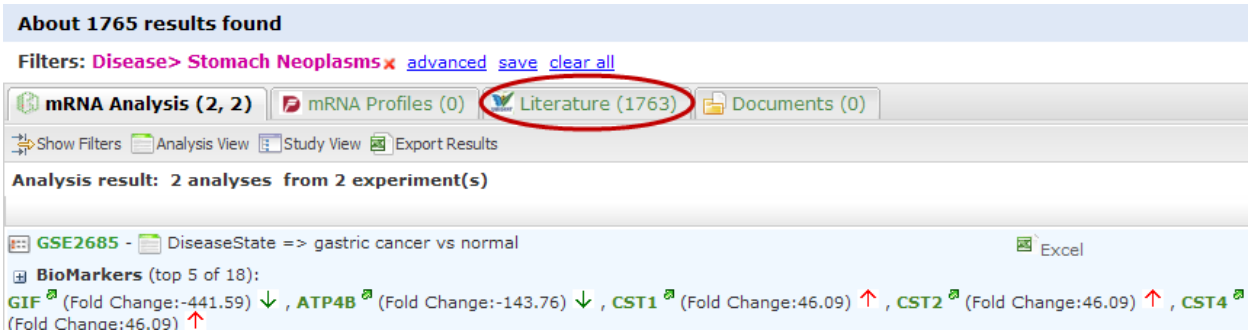
Lesson Goal: Explore literature curated by Jubilant.

Scenario: You want to find peer-reviewed literature about malignant stomach neoplasms in which the MET gene is active. You decide to take advantage of Jubilant's data collection and curation services.

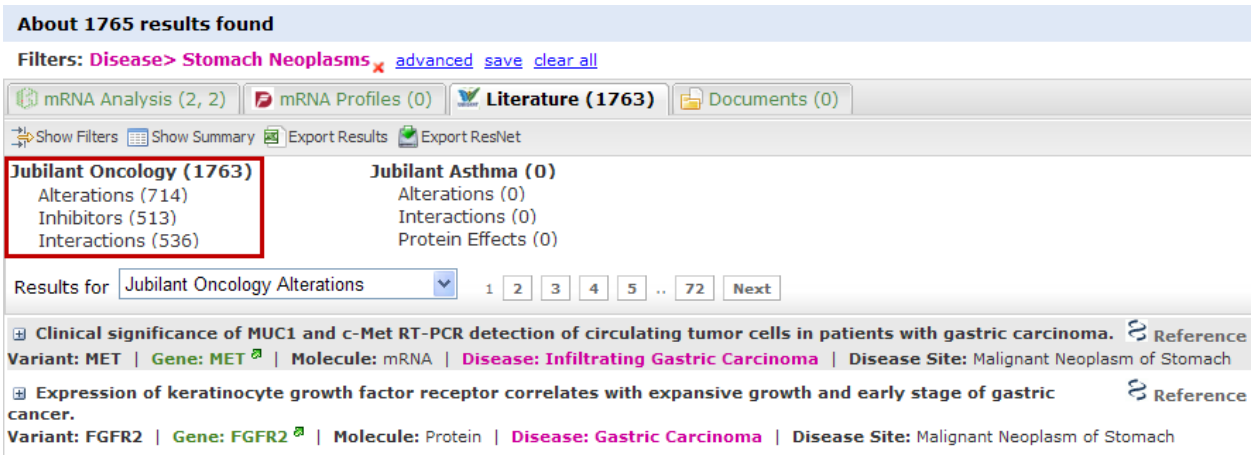
1. Click the **tranSMART** logo to clear the results from the previous lesson.
2. Click the filter category **disease** above the search field.
3. Type **stomach** in the search field.
4. Click **Disease> Stomach Neoplasms** in the dropdown list.

   The search begins immediately, and a result is returned in the **mRNA Analysis** tab by default.

5. Click the **Literature** tab:

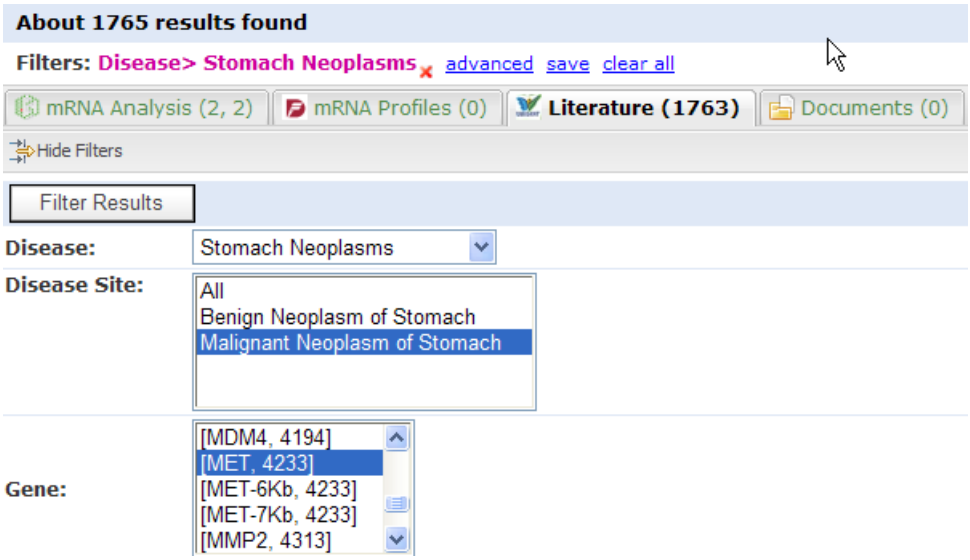


In the figure below, the results in the **Literature** tab contain 1763 records from a set of papers related to oncology, sub-categorized into the topics Alterations, Inhibitors, and Interactions:



You want to limit the search results within the scope of *malignant* stomach neoplasms, and then view the results in the Alterations category.

6. Click the **Show Filters** button.
7. Select the following choices from the following dropdown lists:
   - **Stomach Neoplasms** from **Disease**
   - **Malignant Neoplasm of Stomach** from **Disease Site**
   - **MET** from **Gene**.



In the **Gene** dropdown, note that the Entrez Gene ID (4233 in this case) is provided after the gene symbol.

8. Click the **Filter Results** button (you may need to scroll up first).

   There are still almost 350 results in the Alterations category – too many to scan. You decide to view a summary of the results.

9. Click the **Show Summary** button.

   Each summary record shows the total cases in the Alterations category that involved the specific factors in the alteration. For example, in the following figure, 883 of 1526 cases (a frequency of 57.8%) involved the data type Expression, alteration type Overall Expression (including mRNA and Protein), disease site Malignant Neoplasm of Stomach, and gene MET:

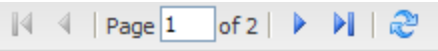
Jubilant Oncology Alteration Summary

Data Type	Alteration Type	Disease Site	Summary	Target	Variant	Frequency	Cases
Expression	Overall Expression (including mRNA and Protein)	Malignant Neoplasm of Stomach	Out of a total of 256 samples, HGF expression was seen in 75.4%. MET	MET		57.8%	883/1526

Note: Summaries are only provided for oncology extracts in the Alterations category.

10. Scan the summaries for groups of articles that interest you.

    Use the controls at the bottom of the window to advance to the next summary page (you may have to scroll down past white space to see the controls):



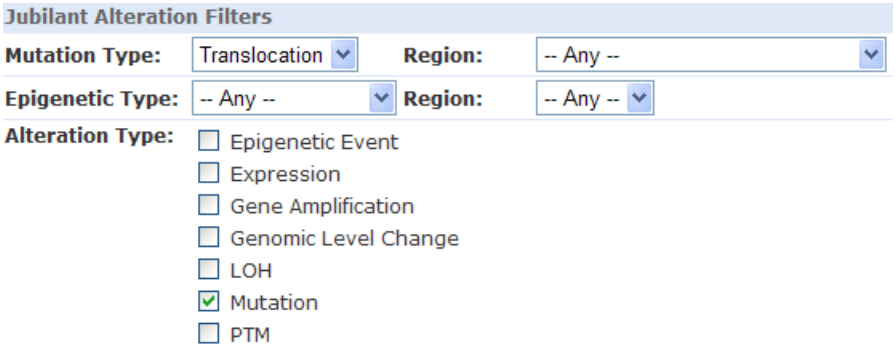
    Scanning the Data Type and Alteration Type columns on the second page, you notice several groups of articles involving translocation mutations.

11. Click the **Show Filters** button.

12. In the **Jubilant Alteration Filters** section, make the following changes:

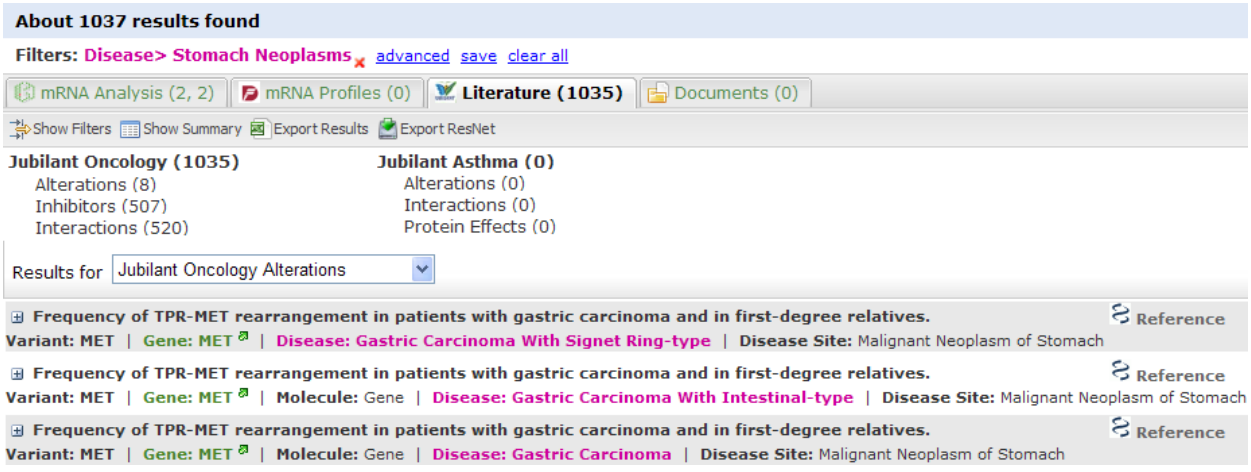
    - Select **Translocation** from the **Mutation Type** dropdown.
    - Clear all the **Alteration Type** check boxes except for **Mutation**.

    The section now looks as follows



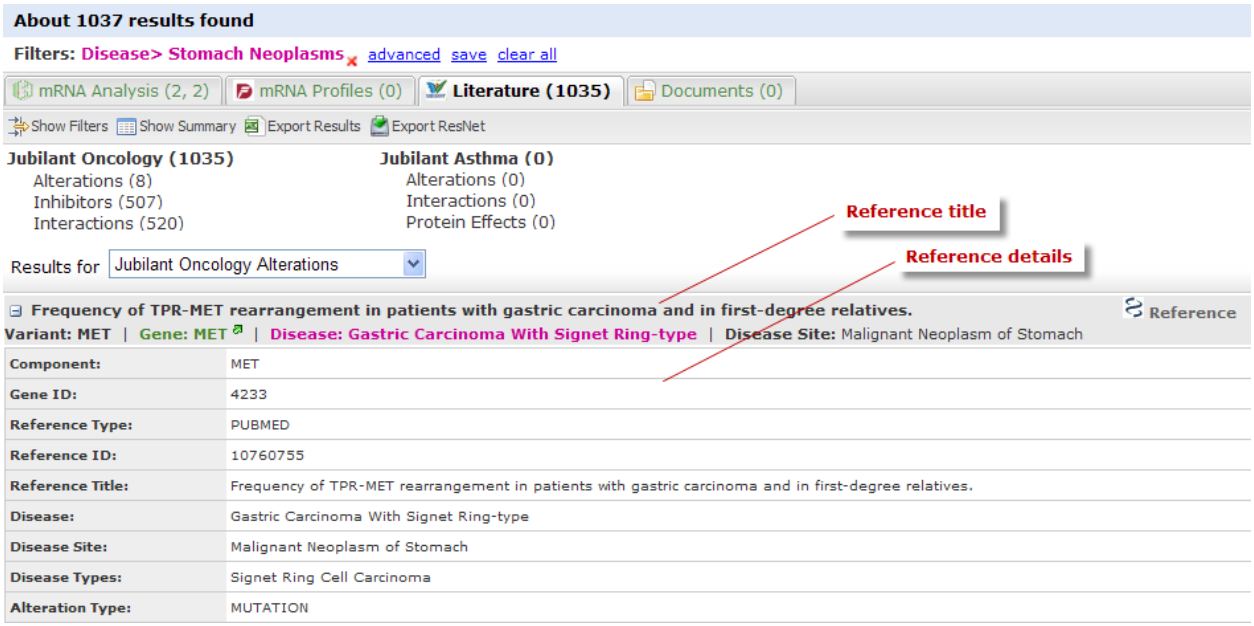
13. Click the **Filter Results** button.

    There are now just eight alterations results. A partial list is shown below:



    You now want to browse through the results to find any articles of interest to you.

14. In the list of reference titles, click the **+** icon (⊞) next to the first title in the list to display details about the referenced article:



15. Click the **Reference** button to the right of the title to display the referenced article in a new browser window:

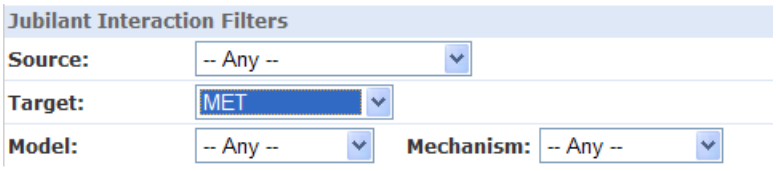
16. After reading through the article, close the browser window containing the article.

    You don't have time to read through any more articles now. You decide to export the reference information to a Microsoft Excel spreadsheet, so you can quickly look up the articles later.

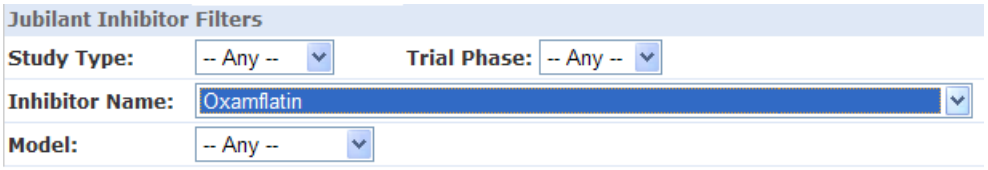
    The export function exports the Inhibitor and Interactions categories as well as Alterations category that you filtered. You decide to filter those categories as well before doing the export.

17. Click the **Show Filters** button.

18. In the **Jubilant Interaction Filters** section, select **MET** from the **Target** dropdown:

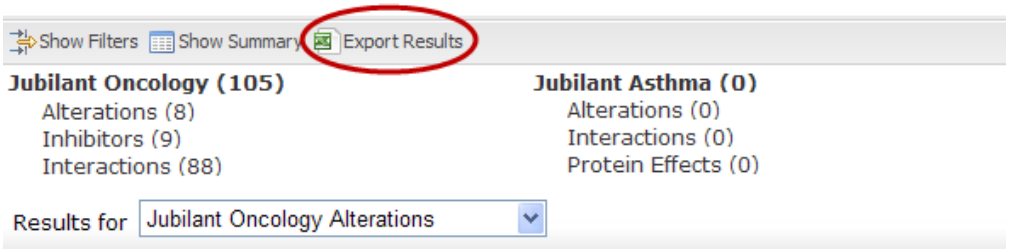


19. In the **Jubilant Inhibitor Filters** section, select **Oxamflatin** from the **Inhibitor Name** dropdown:



20. Click the **Filter Results** button.

21. Click the **Export Results** button:



22. When prompted to open or save the spreadsheet, click **Open**.

    The results are exported to the spreadsheet, with the Alterations, Inhibitors, and Interactions information written to separate worksheets.

    Note that the publication's reference ID is one of the items of information exported to the file:

A	B	C	D	E	F	G	H
Componer	Gene ID	Molecule	Variant	Reference Type	Reference ID	Reference Title	Disease
MET	4233			PUBMED	10760755	Frequency of TPR-MET rearrangement i	Gastric Carcinoma With Signet Ring-type
MET	4233	Gene		PUBMED	10760755	Frequency of TPR-MET rearrangement i	Gastric Carcinoma With Intestinal-type
MET	4233	Gene		PUBMED	10760755	Frequency of TPR-MET rearrangement i	Gastric Carcinoma
MET	4233			PUBMED	10760755	Frequency of TPR-MET rearrangement i	Gastric Carcinoma With Intestinal-type
TPR-MET	4233	Gene	MET	PUBMED	2052572	The TPR-MET oncogenic rearrangement is present and	Gastric Carcinoma

    Use that ID when you want to reference a particular article. For example, to open the PubMed article with ID 2052572, enter the following address in your browser:

    http://www.ncbi.nlm.nih.gov/pubmed/2052572

23. Close the spreadsheet without saving it.

    Normally you would save the spreadsheet for future reference.

Lesson 6

Searching Documents in Proprietary Repositories

Lesson Goal: Use the free-text search feature to search internal document repositories for specific words or phrases used in the documents.

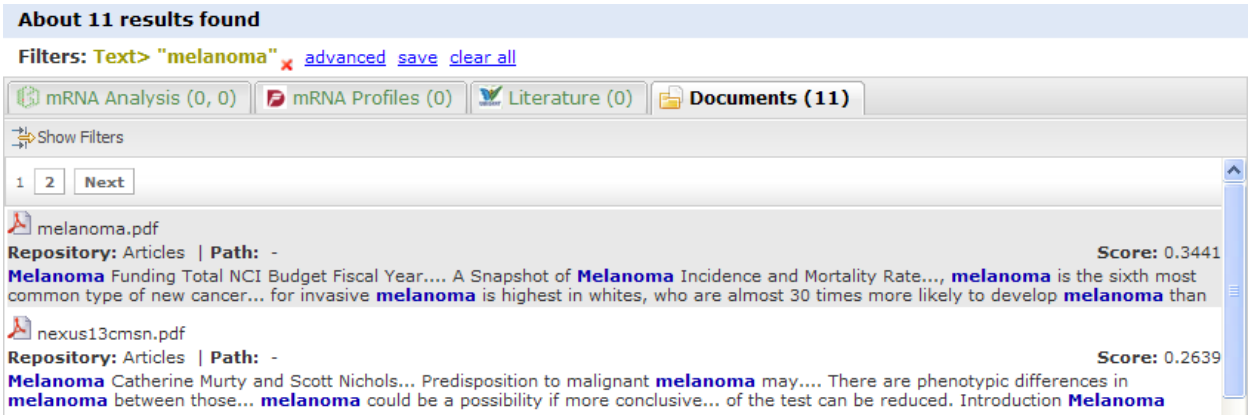
Scenario: You want to find documents about advanced melanoma.

> **Note:** This tranSMART feature searches proprietary document repositories. But for training purposes only, the repository consists of a small set of publicly available documents related to melanoma and asthma.

1. Click the **tranSMART** logo to clear the results from the previous lesson.
2. Type **melanoma** in the search field, then click the **Search** button.

   Alternatively, press the **Enter** key to initiate a free-text search. However, if you press **Enter** after the dropdown list of search filters appears, the search will be based on the filter that happens to be selected in the list, not the search term you typed in the search field.

   Eleven results are returned:



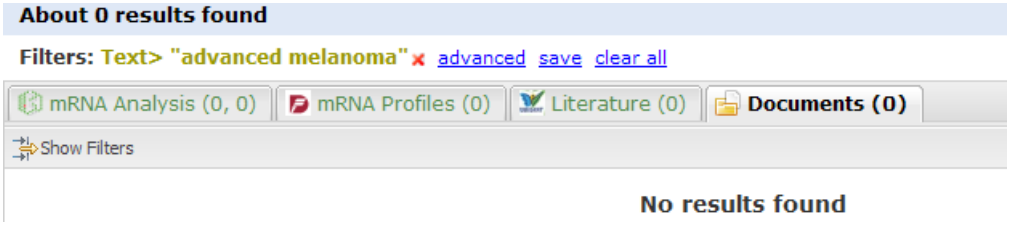
   Notice that no results are returned for any of the other tabs (mRNA Analysis, mRNA Profiles, Literature). That's because a search initiated by clicking the **Search** button or by pressing the **Enter** key searches *only* in the proprietary repositories.

   You are interested in documents on advanced melanoma, so you decide to narrow the results by making the search term more specific.

3. Type the word **advanced** and a space before the word **melanoma** in the search field.

4. Click the **Search** button.

   No results are returned:



   The search looks for the exact word or phrase you typed. If the words "advanced" and "melanoma" are in a document, but not used in the phrase "advanced melanoma," the search will not return a reference to the document.

5. Replace the word **advanced** with the word **metastatic**, then click the **Search** button.

   Two results are returned.

   Notice that each result is assigned a score, and that the results are listed from highest score to lowest. The higher the score, the more instances of the search term were found in the document, relative to the other documents in the search results:

   Notice also that the results are similar to a Google search result – each result has a file name and some brief text extracts that contain part or all of the search term, highlighted in blue.

6. Click the file name of the document **pathways.pdf** to open it in a separate window.

7. When finished reading through the document, close the document's window.

8. Click the **tranSMART** logo to clear the search results.

9. Type **melanoma** in the search field, as you did in Step 2, *but this time don't click the Search button or press Enter*.

10. Click **Disease>Melanoma** in the dropdown list of search filters.

    Notice that the **Documents** tab shows 11 documents returned, as it did in Step 2. But this time, the search was also conducted for the records represented by the other tabs, with results returned for **mRNA Analysis** and **Literature**.

Lesson 7

Using Dataset Explorer

Lesson Goal: Become acquainted with the Dataset Explorer interface.

Dataset Explorer lets you view metrics, observations, and other data from both public studies and from private studies that you are authorized to access.

You can view the entire set of data from a study or just a subset of the data, based on criteria that you specify. You can also include all the test subjects in the study, just a subset of the subjects, or define two groups of cohorts to compare.

Quick Tour

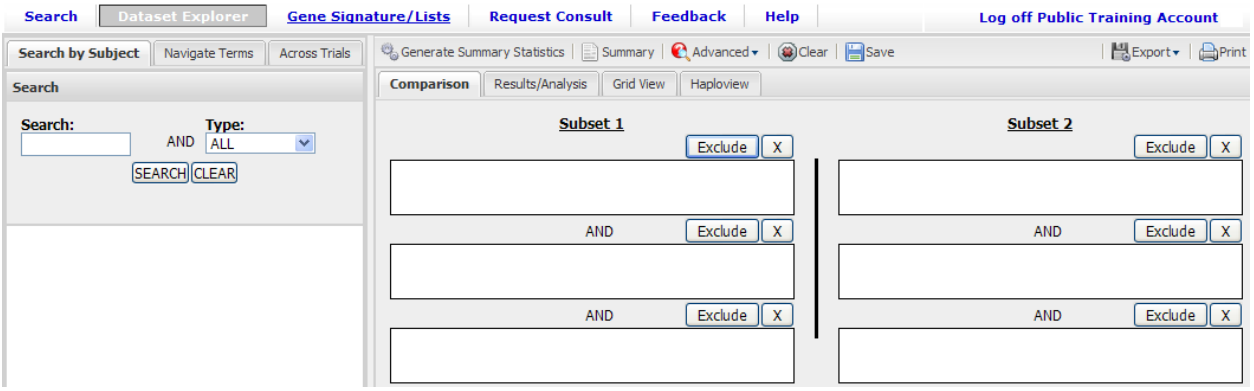
The figure below shows the Dataset Explorer interface. It is divided into two panes:

Left pane

- Lets you select the study of interest.
- Provides a Microsoft Windows Explorer-like navigation tree where you select the concepts that are most relevant to your area of interest.

Right pane

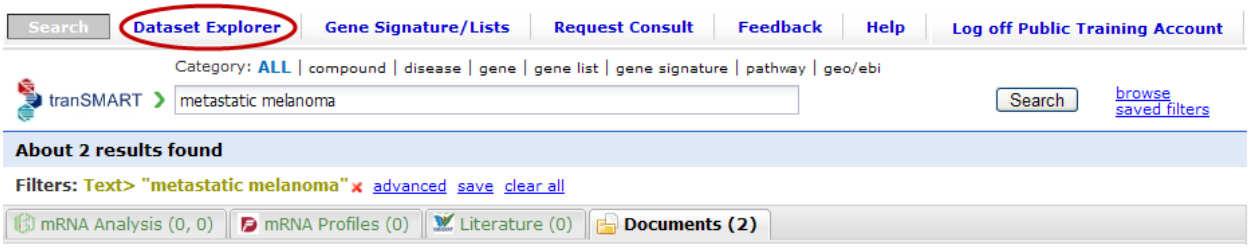
- Lets you define the criteria that determine the subset of subjects and data you want to focus on. You can define a single subset, or you can define two subsets for purposes of comparison. Subjects and data that do not satisfy the criteria you define are excluded from the subset(s).
- Provides summary data about the selected subjects, and several different kinds of views of the data (such as charts, tables, and grid).



Viewing the List of Studies and Opening a Study

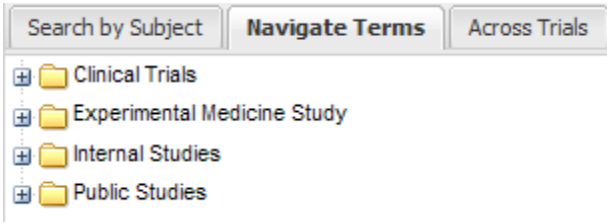
Scenario: You want to see a list of all the studies in the data warehouse and then open one of the studies.

1. Click the tranSMART **Dataset Explorer** tab to display the Dataset Explorer window:



2. In the left pane of the Dataset Explorer window, click the **Navigate Terms** tab.

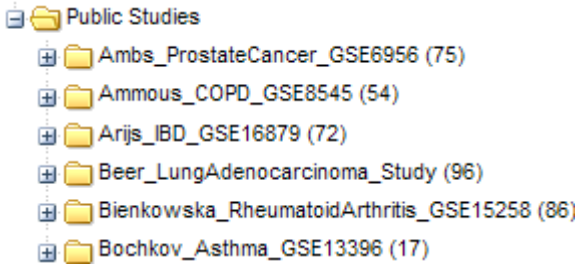
   The navigation tree appears, showing the categories of available studies:



   During training, the only category that will be used is **Public Studies**. The other categories are empty during training.

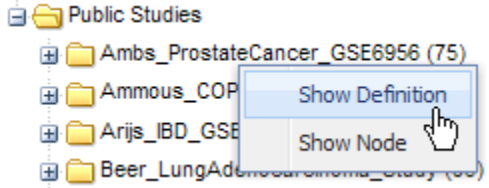
3. Open the **Public Studies** category by clicking the **+** icon (⊞) to the left of the category name.

   A list of all public studies appears. A partial list is shown below:

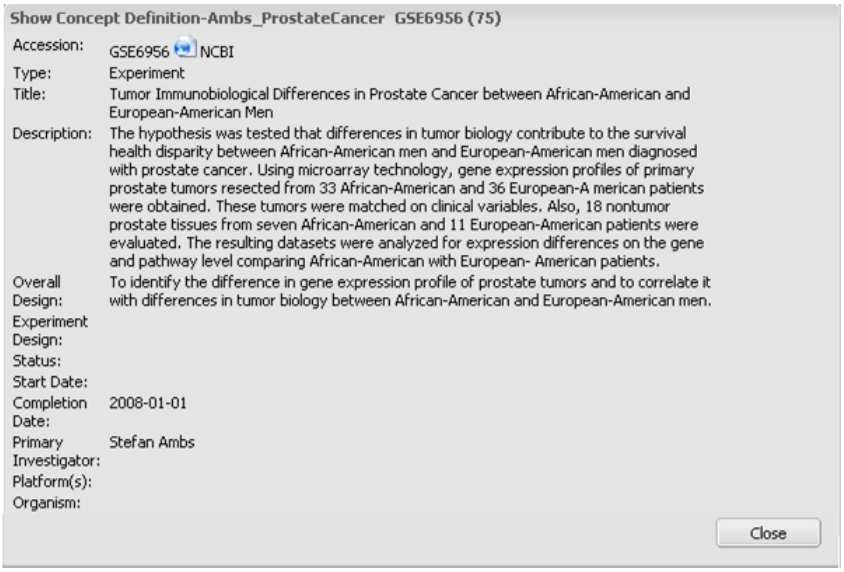


   The number in parentheses after each study name indicates the number of subjects in the study.

4. Right-click the study **Ambs_ProstateCancer_GSE6956**, then click **Show Definition**.

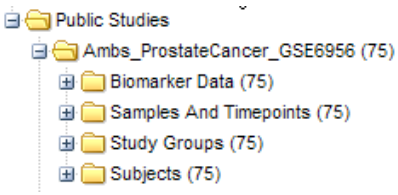


A details box appears that describes the study:



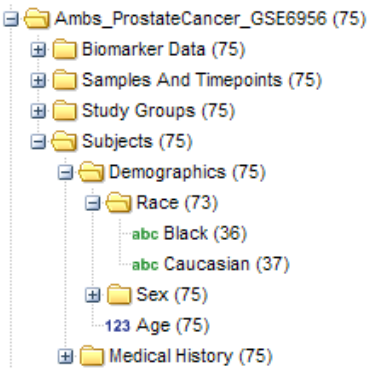
5. Click **Close** to close the study description.
6. Click the **+** icon (⊞) to the left of study **Ambs_ProstateCancer_GSE6956** to open it.

   The nodes in the navigation tree that appear represent various aspects of the study:



   To see more detail about a particular aspect of the study, click the **+** icon (⊞) next to the node to open it. For example, to see a breakdown of the subjects in the study by race, open the following nested nodes in the following order:

   a. Subjects
   b. Demographics
   c. Race



   In the figure above, the numbers in parentheses after **Black** and **Caucasian** tell you that 36 African Americans and 37 Caucasians participated in the study.

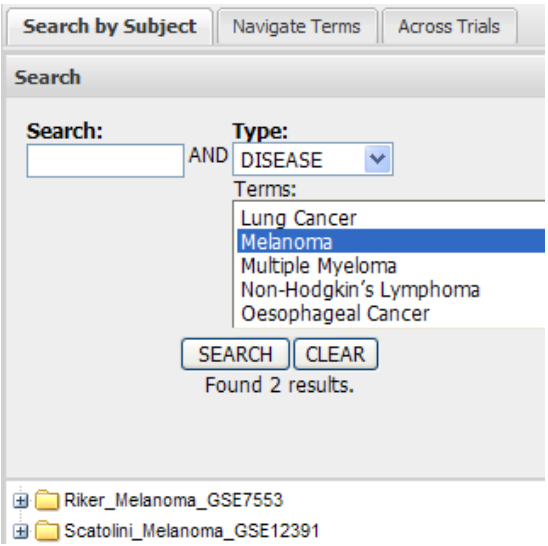
Finding a Study

Scenario: You want to find a study involving melanoma.

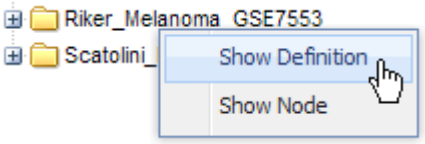
1. In the left pane, click the **Search by Subject** tab.
2. In the **Type** dropdown, select **DISEASE**.
3. Select **Melanoma** from the **Terms** dropdown.

   You may need to widen the left pane to see the scroll bar for the **Terms** dropdown.
4. Click **Search**.

   tranSMART displays a list of the studies involving melanoma:



   To see more information about a listed study, right-click the study name, then click **Show Definition**.



5. To open a study, click the **+** icon (⊞) next to the study name.

Alternatively, to browse through the list of studies, click the **Navigate Terms** tab.

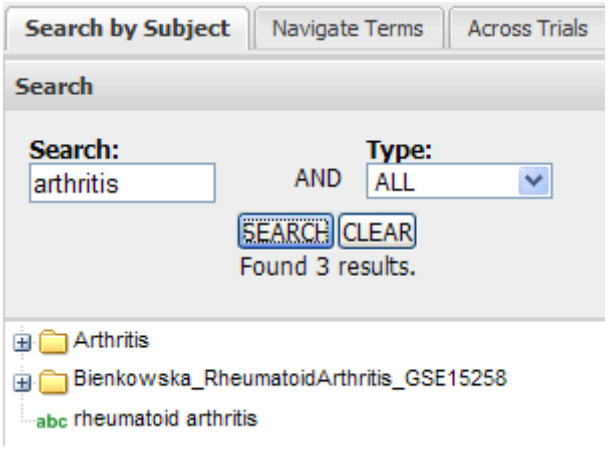
Finding an Area of Interest in a Study

Scenario: You want to find studies in which arthritis is an area of interest.

1. In the left pane, click the **Search by Subject** tab.
2. Click **Clear** to remove the search criteria from the previous lesson.

3. In the **Search** box, type **arthritis**, then click **Search**.
4. The search results contain all nodes relating to arthritis.

   The returned nodes might represent an entire study or a specific node within a study, as shown below:

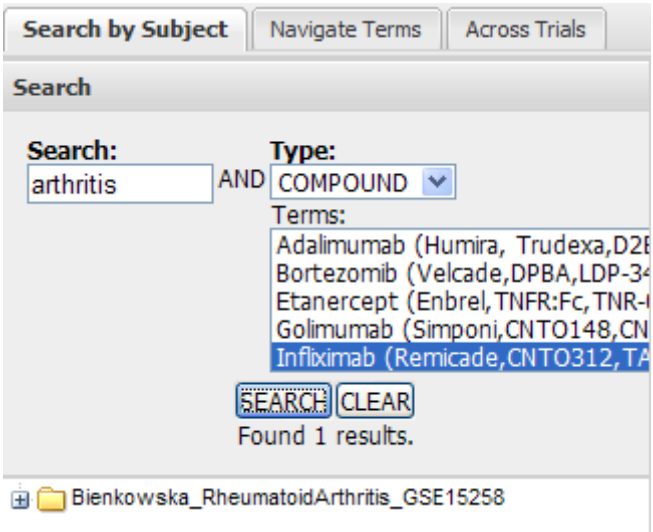


5. To find out more about an item in the list, do either or both of the following:

   - Hover the mouse pointer over a node to see the name of the study that contains the node.
   - Right click a node, then click **Show Definition** to see a description of the associated study.

6. Optionally, you can further filter the search results by using the **Type** field in combination with the **Search** field. For example, to list only those asthma studies that involved the compound Infliximab, select **COMPOUND** in the **Type** field, then select **Infliximab** in the **Terms** list.

   When you click **Search**, the result now contains just one item:

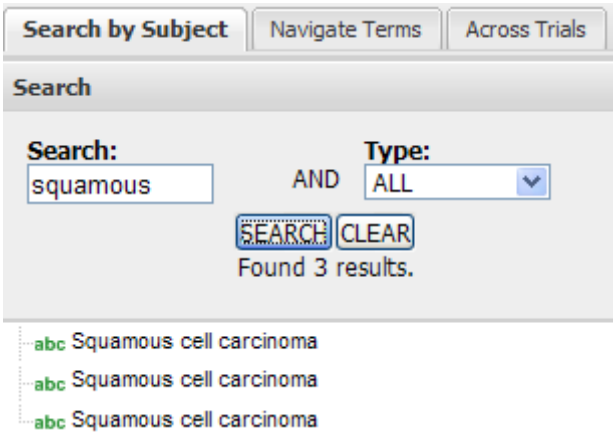


Opening a Study from the Search Result

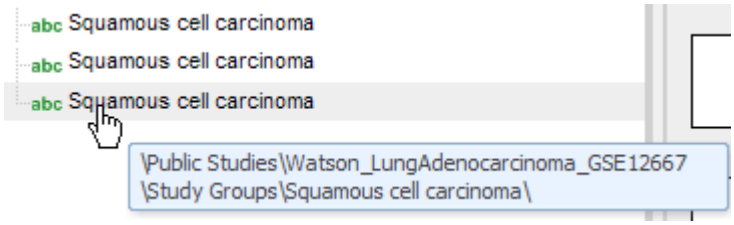
Scenario: The item of interest in a search result is a study subnode rather than the top-level node of the study. You want to open the complete study.

1. Click **Clear** to remove the search criteria from the previous lesson.
2. Type **squamous** in the **Search** field, then click the **Search** Button.

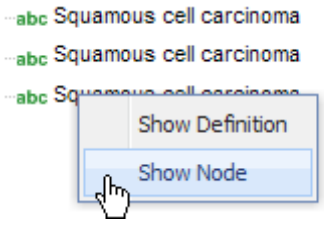
   The search result includes study subnodes:



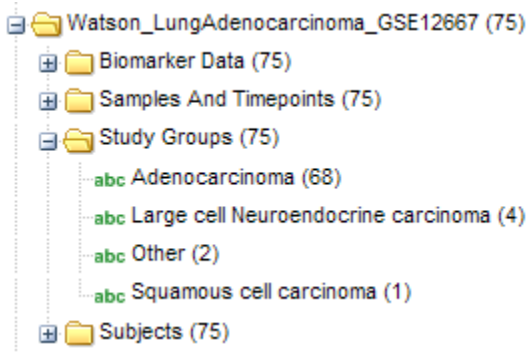
3. Hover the mouse pointer over each item in the result to see the names of the studies – for example:



4. When you find an item that interests you, right-click the item, then click **Show Node**:



The **Navigate Terms** tab opens showing the list of public studies, with the study you selected open. You may need to scroll down in the list to see it:



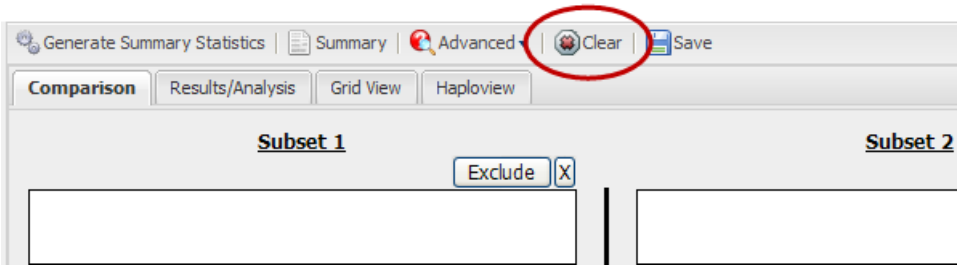
Specifying Subsets of Data and Subjects in a Study

Once you have selected a study, you specify the data and subjects from the study that are most relevant to your area of interest. You do so by selecting concepts from the study's tree structure and dragging them into the subset definition boxes in the right pane of Dataset Explorer.

You can specify a single subset of data and subjects, or two subsets that can be compared.

Scenario: You want to compare survival statistics for breast cancer patients under 60 years old, divided into two groups: those with negative estrogen receptor tumors, and those with positive estrogen receptor tumors.

1. Make sure the subset definition boxes in the right pane are empty. If they are not, click the **Clear** button above the boxes:



2. In the left pane, the **Navigate Terms** tab should already be open. If it is not, click the tab to open it.

   The studies are listed in the Navigate Terms tab in alphabetical order.

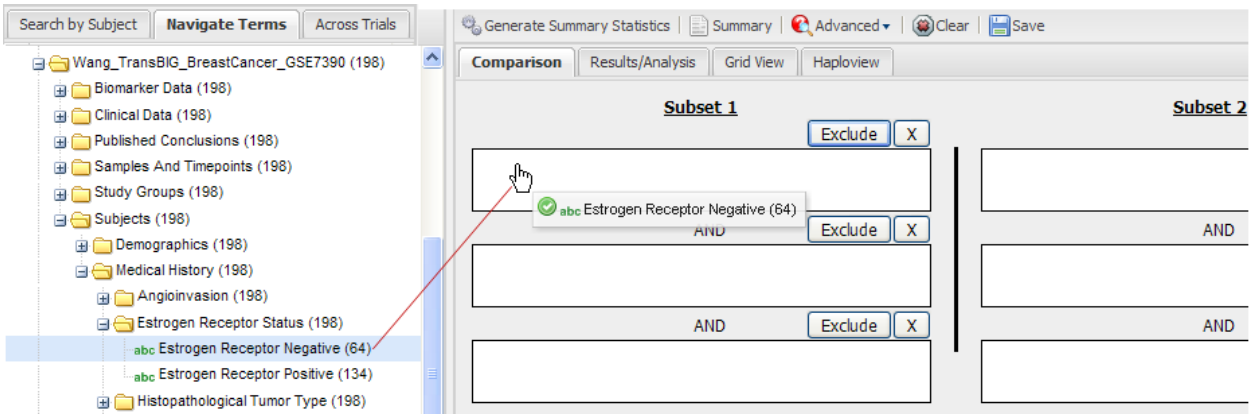
3. Open the study **Wang_TransBIG_BreastCancer_GSE7390**.

4. Open the following nested nodes in the following order:

   a. Subjects
   b. Medical History
   c. Estrogen Receptor Status

Notice that the nodes **Estrogen Receptor Negative** and **Estrogen Receptor Positive** have the green icon **abc** before them. This icon indicates that the node is not associated with a numeric value that you need to assign.

5. Drag **Estrogen Receptor Negative** into a subset definition box in Subset 1:



When you release the mouse button, **Estrogen Receptor Negative** becomes part of the subset definition for Subset 1.

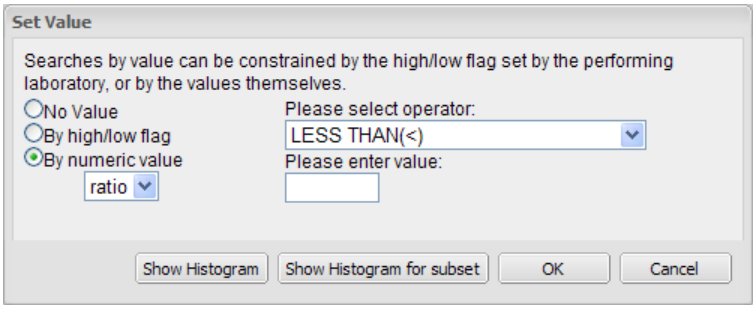
6. Drag **Estrogen Receptor Positive** into a box in Subset 2.

7. Return to the **Subjects** node and open the node **Demographics**.

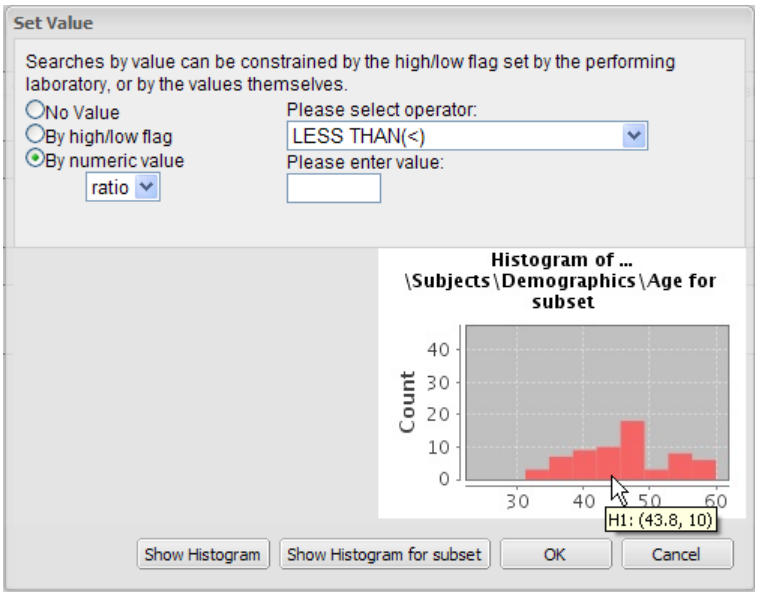
Notice that the node **Age** has the blue icon **123** before it. This icon indicates that the node is associated with a numeric value that you need to assign. You are prompted to assign the value immediately after you drop the node into a subset definition box.

8. Drag **Age** into an empty box in Subset 1.

The Set Value dialog appears:



Note: The **Show Histogram** buttons on this dialog are useful if you need help deciding on the value to enter. Clicking **Show Histogram for subset** (or **Show Histogram** to display a histogram for both subsets) displays a histogram like the following, showing how the values (in this case, ages) are distributed across the cohorts in the subset(s). Also, to see the average value in any given bar of the histogram, and the number of cohorts to whom the value applies, hover the mouse pointer over the bar.



In this case, however, we know the value we want to use – 60.

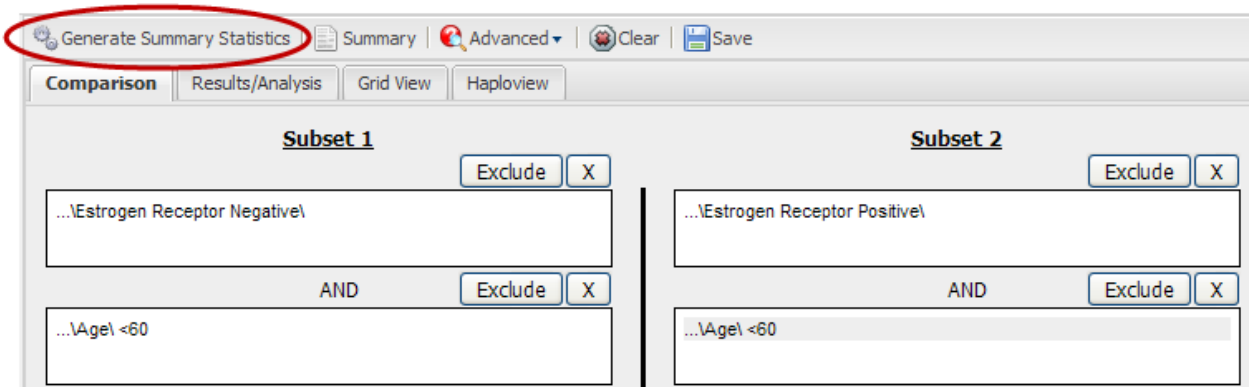
9. In **Please enter value**, type **60**, then click **OK**.

   Age < 60 becomes part of the subset definition for Subset 1.

10. Repeat Step 8 and Step 9 for Subset 2.

    Now that the subsets are defined, you are ready to generate data from the study that applies to the subsets.

11. Click the **Generate Summary Statistics** button:



In the Results/Analysis view shown in step 13 below, you see that 64 subjects meet the criteria defined for Subset 1, and 131 meet the criteria for Subset 2.

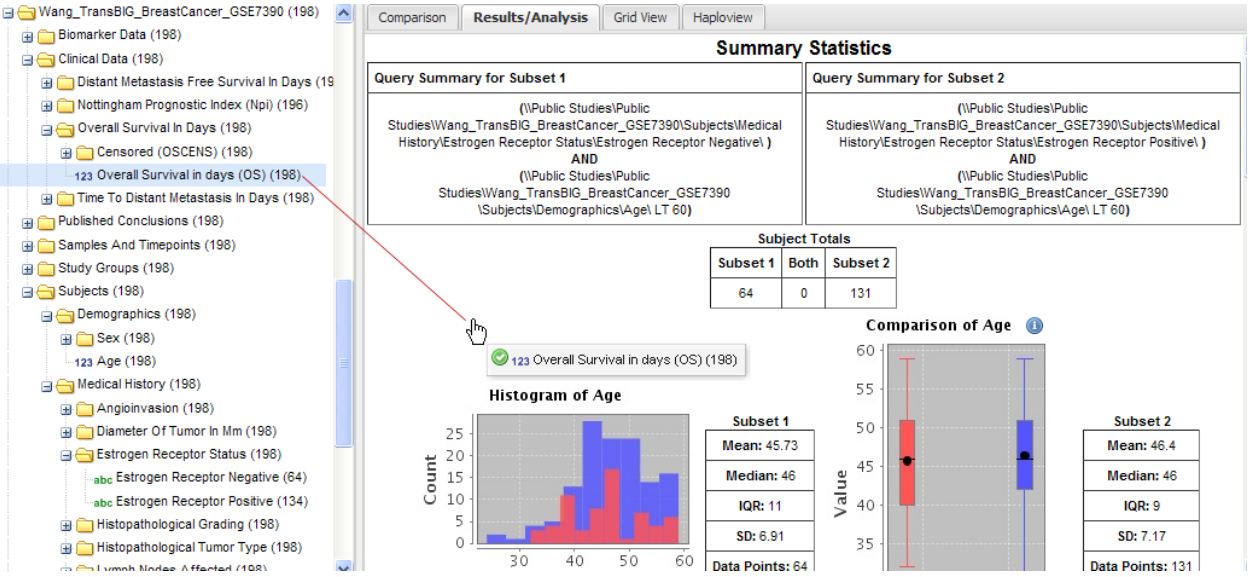
You now want to apply points of comparison to the subsets, based on clinical observations of the subjects in each subset.

12. Open the following nodes in the following order:

    a. Clinical Data
    b. Overall Survival in Days

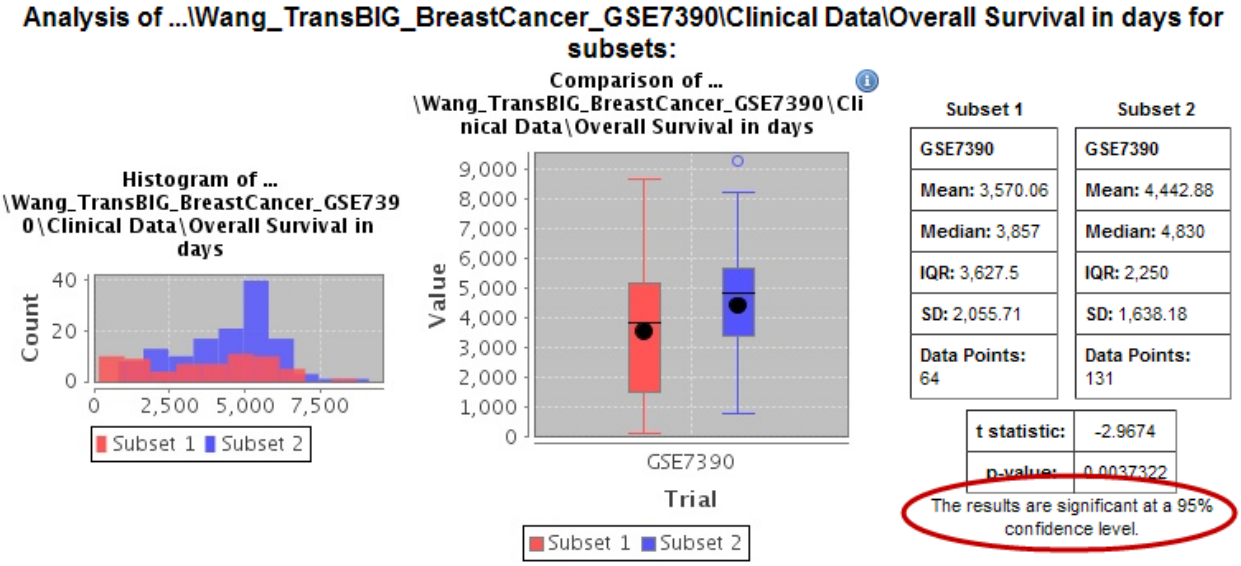
    Notice that inside the **Overall Survival in Days** node is a subnode of the same name.

13. Drag the subnode **Overall Survival in days** anywhere into the Results/Analysis view:



When you release the mouse button, the following charts and tables appear at the top of the Results/Analysis view, presenting comparison data based on overall survival statistics in days for the subjects in each subset.

Notice that the results of the analysis are significant at a 95% confidence level:

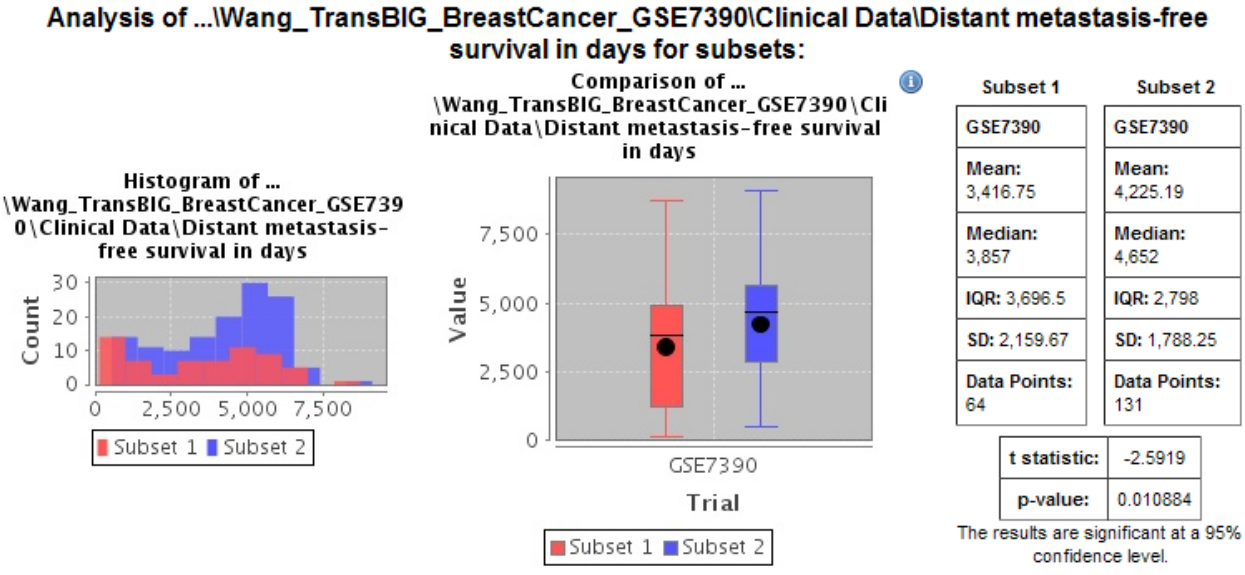


Also notice the three colors in the histogram chart on the left. Blue indicates data observed in Subset 2 subjects. The lighter shade of red indicates data observed in Subset 1 subjects. The darker shade of red represents overlapping data – in this case, the same number of survival days observed in subjects in each subset.

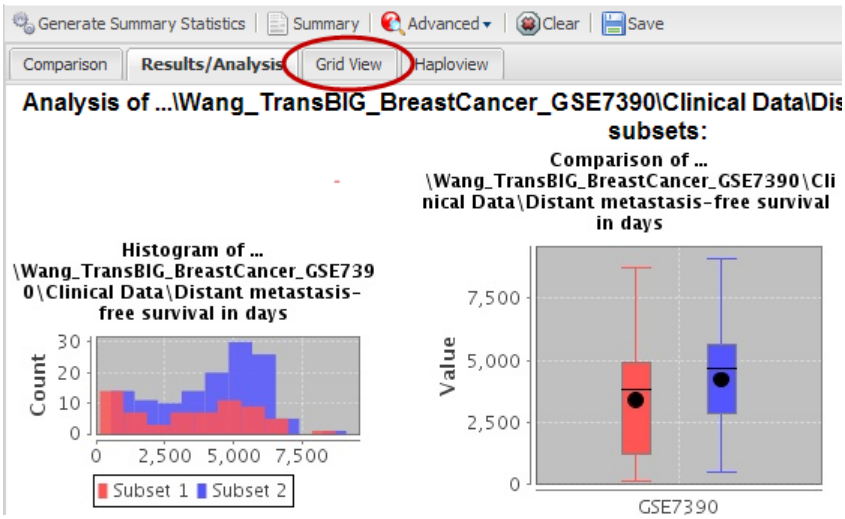
You now want to introduce another point of comparison.

14. Open the **Distant Metastasis Free Survival in Days** node, then drag the subnode **Distant metastasis-free survival in days** anywhere into the Results/Analysis view.

The following charts and tables appear at the top of the Results/Analysis view, presenting the comparison data:



15. Click the **Grid View** tab.



Clicking this tab displays a table of data for each of the subjects in the subsets.

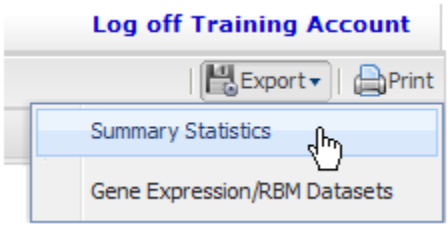
16. Click the **Results/Analysis** tab to return to the graphical view.

17. Optionally, continue dragging other points of comparison into the Results/Analysis view.

You decide to save the summary statistics results to a file.

> **Note:** This is a public study. All users have full access to public studies, However, with private studies, you must have permission from the study owner to save study data to a file.

18. Click **Export**, then click **Summary Statistics**:



19. When prompted to open or save the spreadsheet, click **Open**.

    Summary statistics for several of the subjects in Subset 1 appear below:

	A	B	C	D	E	F	G	H	I	J	K
1	subject	subset	TRIAL	SEX_CD	AGE_IN_Y	RACE_CD	\Medical_	\Medical_	\Subjects\Demographics\Age		
2	84351	subset1	GSE7390	F	42	NULL	Negative	N	42		
3	84354	subset1	GSE7390	F	37	NULL	Negative	N	37		
4	84357	subset1	GSE7390	F	38	NULL	Negative	N	38		
5	84360	subset1	GSE7390	F	53	NULL	Negative	N	53		
6	84362	subset1	GSE7390	F	37	NULL	Negative	N	37		
7	84366	subset1	GSE7390	F	46	NULL	Negative	N	46		
8	84370	subset1	GSE7390	F	59	NULL	Negative	N	59		
9	84371	subset1	GSE7390	F	41	NULL	Negative	N	41		
10	84374	subset1	GSE7390	F	53	NULL	Negative	N	53		
11	84389	subset1	GSE7390	F	45	NULL	Negative	N	45		

20. Close the spreadsheet without saving it.

    Normally you would save the spreadsheet for future reference.