JnJ Training: Advanced

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: **jnjuser***xx*

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

  For example, **jnjuser5** or **jnjuser21**.

- 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 references to specific data in this tutorial may be different than what you see on your screen.

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

Lesson 1

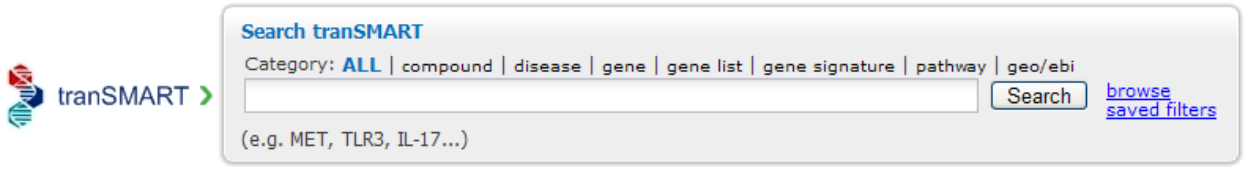
Remove Parts of a Search String

Lesson Goal: Remove parts of a search string, including individual genes in a pathway.

Scenario: You are interested in the relationship between melanoma and the gene EDNRB, and secondarily, in relationships between melanoma and the other genes in the melanogenesis pathway.

1. Start up tranSMART as described in Logging In to tranSMART Training on page 1.

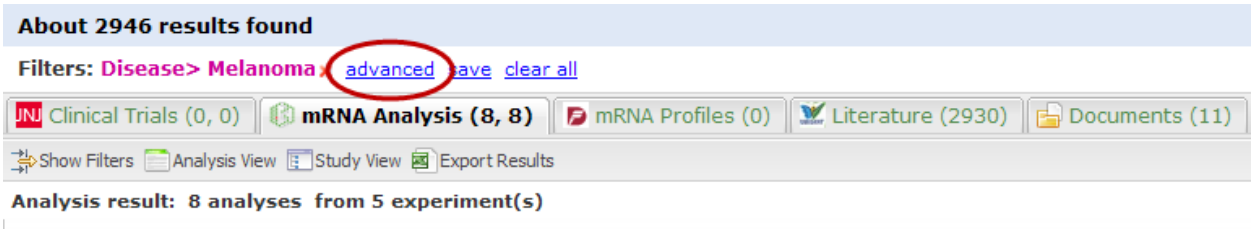
   The tranSMART Search window appears:



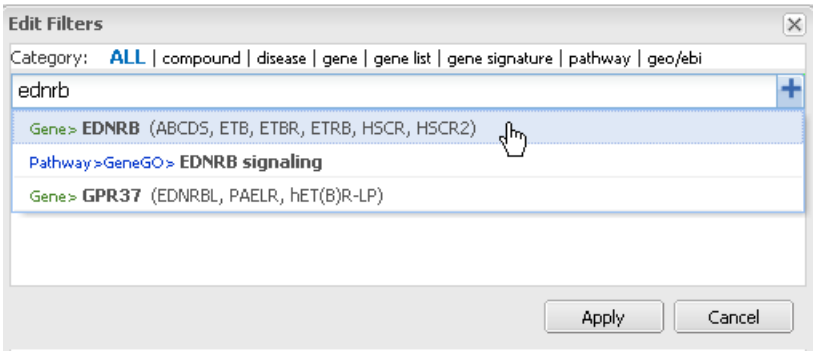
2. Type **melan** in the search field.
3. Click **Disease> Melanoma** in the dropdown list of search filters.

   After the search result is returned, you want to add the gene EDNRB to the search string.

4. Click **advanced**:

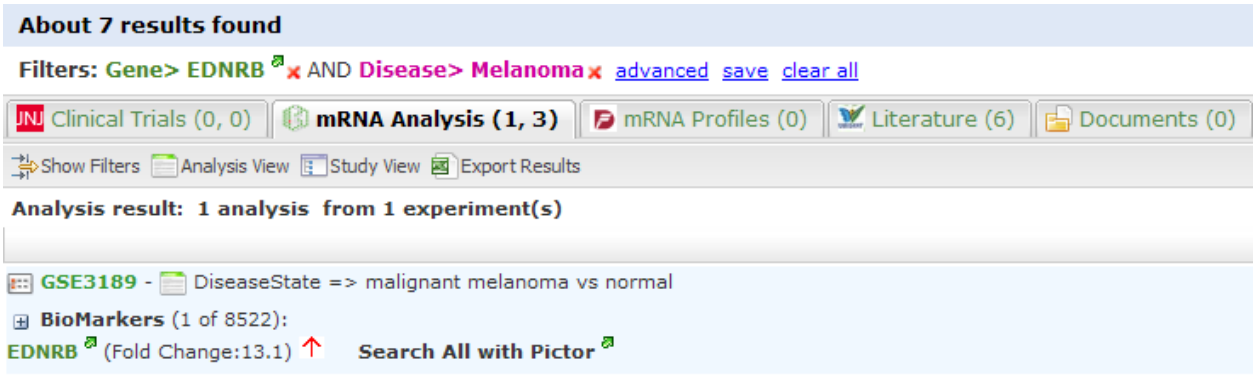


5. Type **EDNRB** in the search field of the Edit Filters dialog.
6. Click **Gene> EDNRB** in the dropdown list:



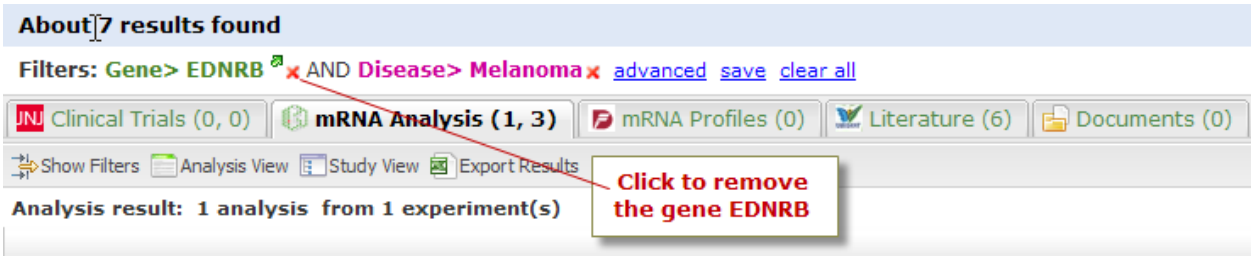
7. Click **Apply**.

   As shown in the figure below, the search returns a single study, **GSE3189**, that matches the search criteria. The study has three matching analyses, but only one, **malignant melanoma vs normal**, that is considered statistically significant.



   You now want to look for relationships between melanoma and the genes in the melanogenesis pathway. First, you want to remove EDNRB from the search string.

8. Click the red **X** after the name **EDNRB**:



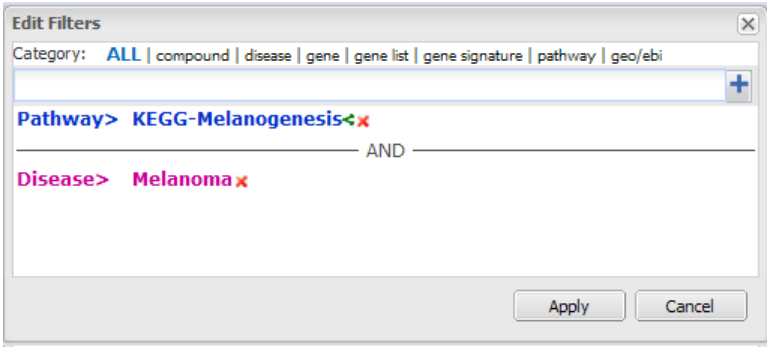
   A search begins immediately based on the modified search filter, and a result is returned.

9. Click **advanced**.

10. Type **melan** in the search field of the Edit Filters dialog.

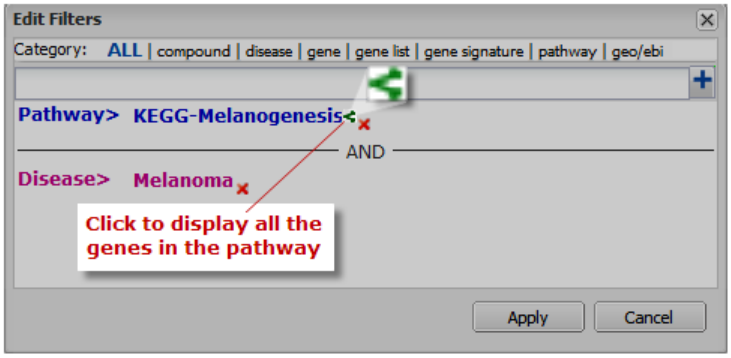
11. Click **Pathway>KEGG>Melanogenesis** in the dropdown list.

    The Edit Filters dialog now looks as follows:



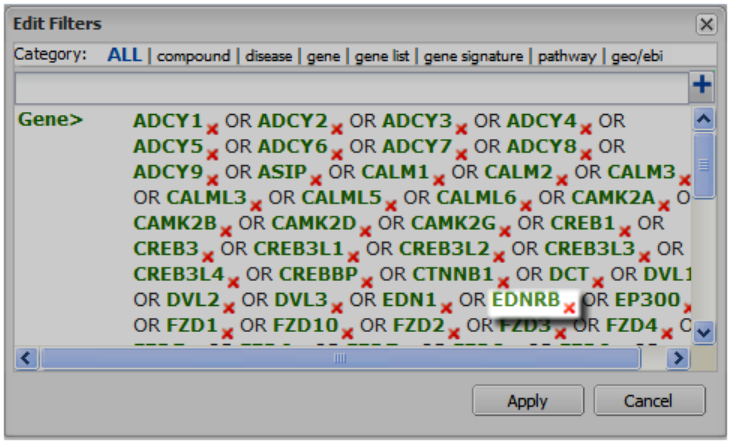
Since you've already searched for relationships between melanoma and EDNRB, you would like to conduct this new search without EDNRB as a factor. To do so, you must remove EDNRB from the melanogenesis pathway.

12. Click the green angle bracket immediately after the name of the melanogenesis pathway:



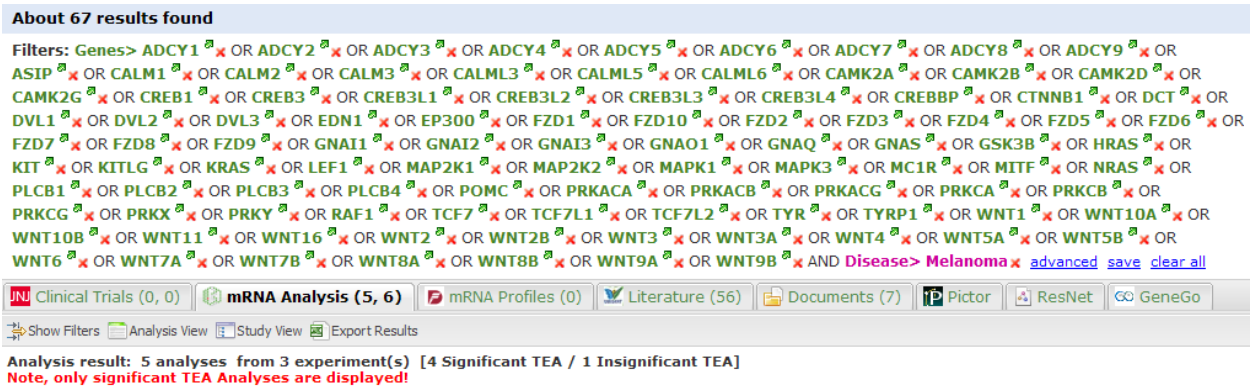
The name of the melanogenesis pathway is replaced by a list of the individual genes in the pathway.

13. Locate **EDNRB** in the alphabetized list, then click the red **X** after the gene name:



14. Click **Apply**.

tranSMART returns the new search result:



Lesson 2

Create a Gene Signature

Lesson Goal: Use the tranSMART gene signature wizard to create a gene signature.

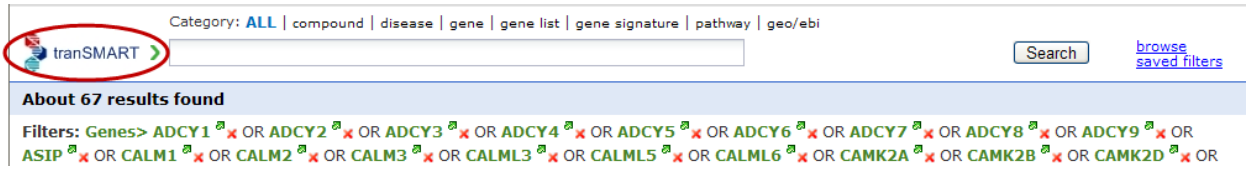
Scenario: You are interested in lung adenocarcinoma, and want to create a gene signature consisting of genes that were strongly up-regulated in an experiment involving lung adenocarcinoma patients.

This lesson involves two basic tasks:

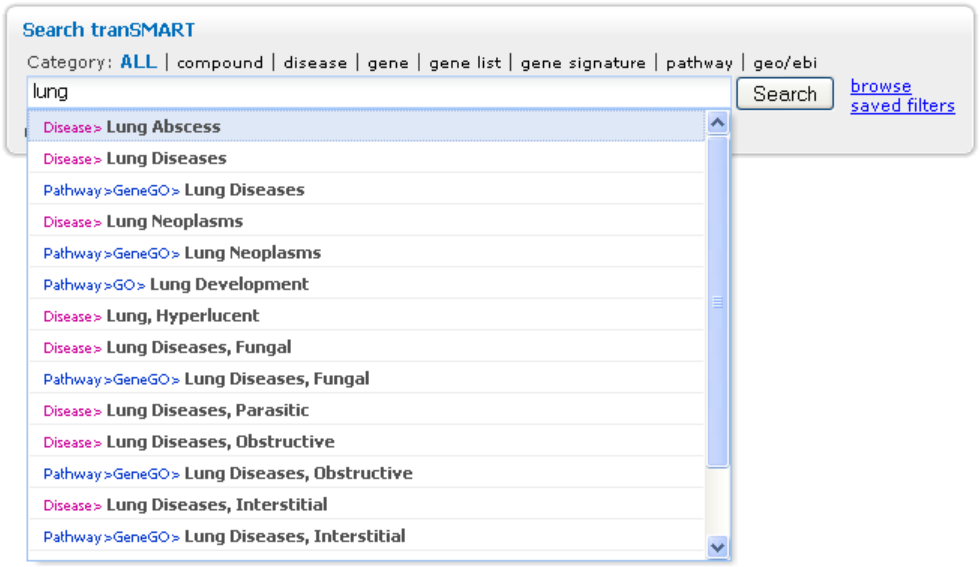
1. Find the genes to include in the gene signature, then write the genes to a tab-delimited text file that can be imported into the gene signature.
2. Define the gene signature and import the file containing the genes.

Task 1: Find the Genes for the Gene Signature and Write them to a File

1. Click the tranSMART logo to clear the results from the previous lesson:



2. Type **lung** in the Search field:

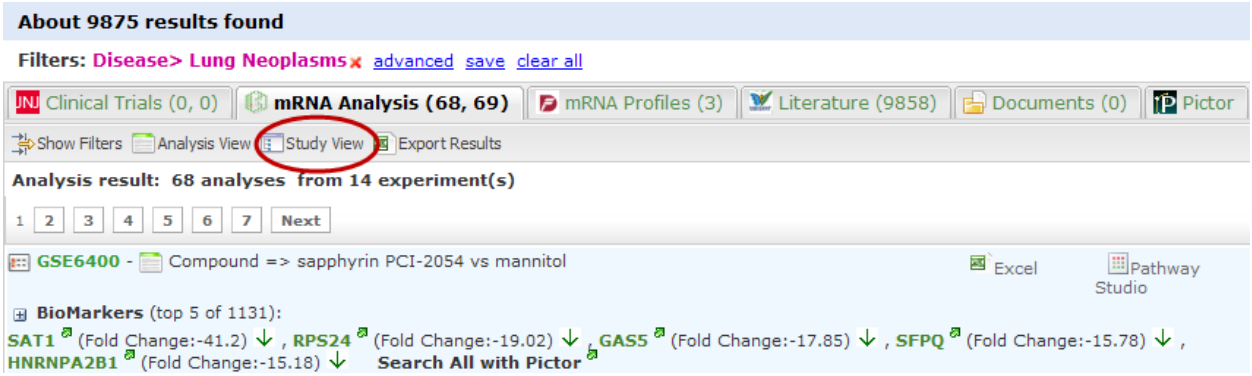


Lung adenocarcinoma is not listed in the dropdown list of search filters, but lung neoplasms is listed.

3. Click **Disease> Lung Neoplasms**.

   In a few seconds, the search result appears.

4. Click the **Study View** button in the **mRNA Analysis** tab:

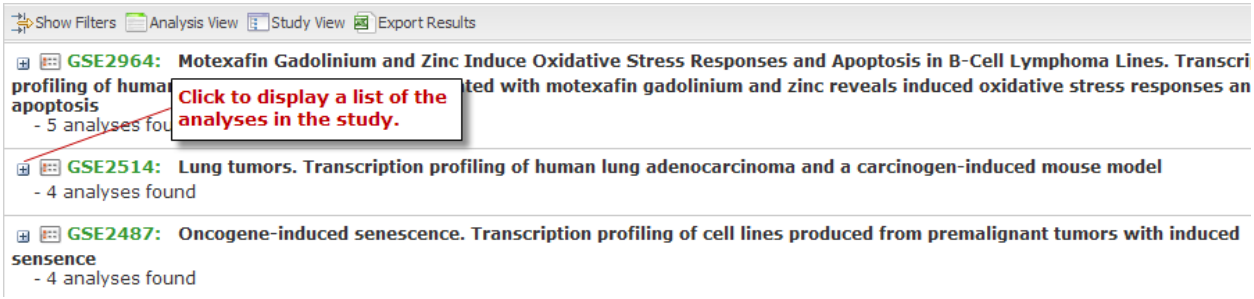


   tranSMART displays a list of all the experiments related to lung neoplasms.

5. Scroll through the list of experiments until the experiment **GSE2514** appears.

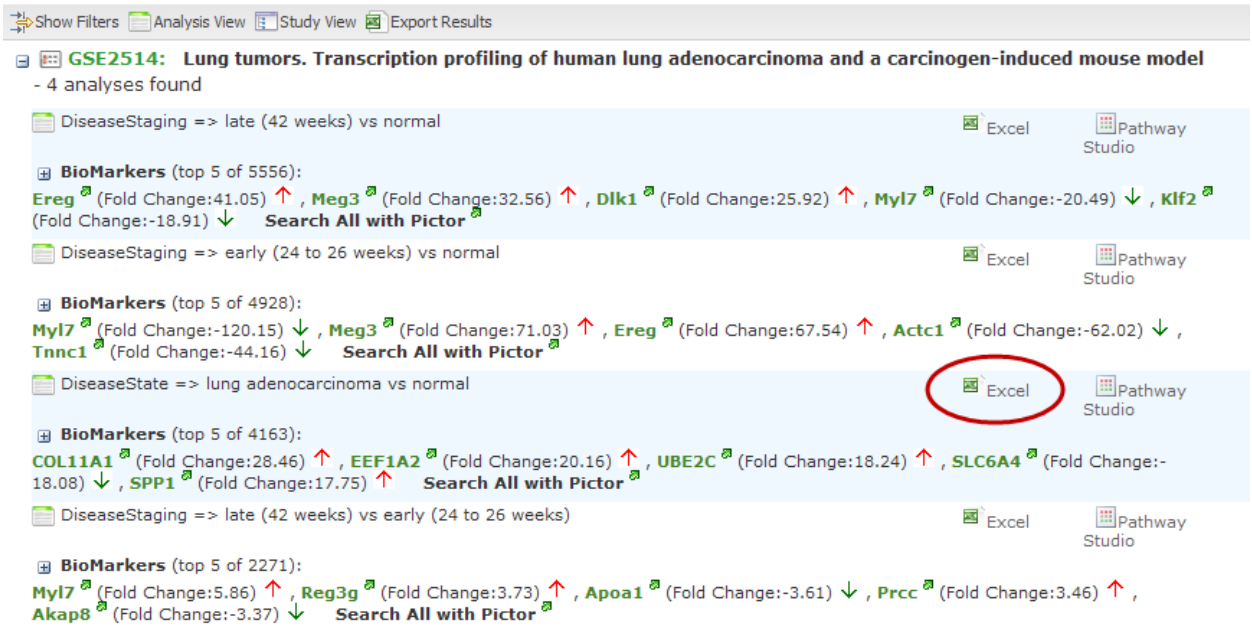
   You notice that this experiment focused specifically on lung adenocarcinoma.

6. Click the **+** icon (⊞) to the left of the experiment name:



   A list of the analyses based on this experiment appears. The analysis **Lung adenocarcinoma vs normal** interests you.

7. Click the **Excel** button for the analysis **Lung adenocarcinoma vs normal** to export the analysis data to a Microsoft Excel file:



8. When the File Download dialog appears, click **Open**.

   Excel starts up and displays the analysis data. The following figure shows the first few rows of data:

	A	B	C	D	E	F	G
1	Analysis	Probe Set	Fold Change Ratio	p-Value	adjusted p-value	TEA p-Value	Gene
2	DiseaseState => lung adenocarcinoma vs norma	479_at	-1.45	0.0009	0.0051	0.2136	DAB2
3	DiseaseState => lung adenocarcinoma vs norma	36275_at	-2.84	0	0.00004	0.05363	SEMA6A
4	DiseaseState => lung adenocarcinoma vs norma	34830_at	1.18	0.0174	0.0558	0.26259	ECOP
5	DiseaseState => lung adenocarcinoma vs norma	1970_s_at	-5.31	0	0.00003	0.0011	FGFR2
6	DiseaseState => lung adenocarcinoma vs norma	33908_at	1.17	0.0313	0.0887	0.26451	CAPN1
7	DiseaseState => lung adenocarcinoma vs norma	32146_s_at	-2	0	0	0.13197	ADD1
8	DiseaseState => lung adenocarcinoma vs norma	33062_at	-1.55	0.0359	0.0987	0.1969	GSTA1

   You want to select the most strongly up-regulated and down-regulated genes for your gene signature. To find them, you will sort the rows according to the **Fold Change Ratio** column.

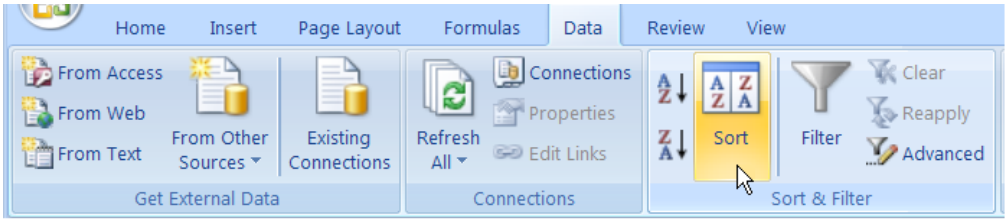
> **Note:** The instructions in this lesson for performing Excel operations are based on Microsoft Excel 2007. If you have a different version, some of the steps and graphics may be different for you.

9. Select all the data in all the columns, as follows:

   a. Click the letter **A** above the leftmost column.

   b. Press and hold down the **Shift** key, then click the letter above the rightmost column (letter G in the figure above).

10. Click the **Data** menu, then click **Sort**:

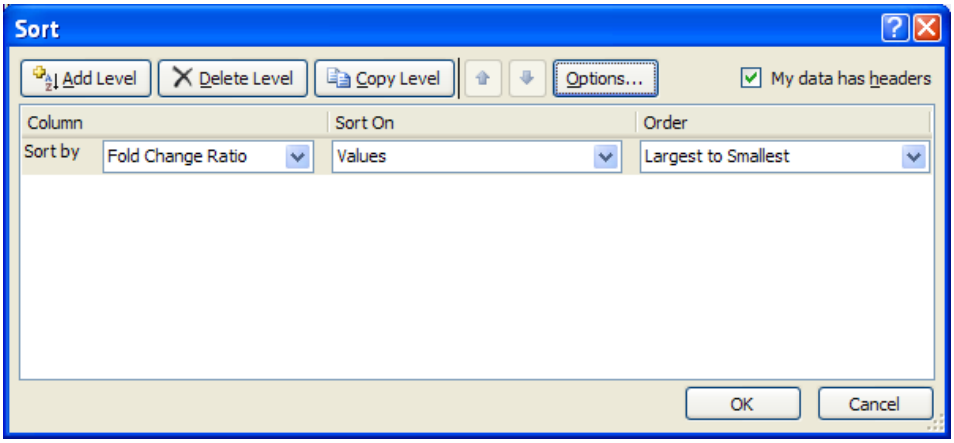


The Sort dialog appears.

11. Select **Fold Change Ratio** in the dropdown list under **Column**.
12. Select **Largest to Smallest** in the dropdown list under **Order**.

Note: If the choices in the dropdown are **A to Z** and **Z to A**, choose **Z to A**.

The Sort dialog now appears as follows:



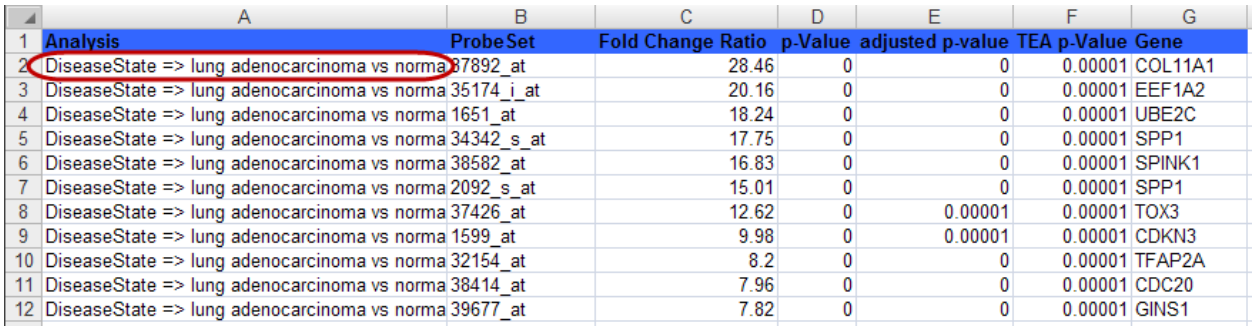
13. Click **OK**.

Note: If the Sort Warning dialog appears, select **Sort anything that looks like a number, as a number**, then click **OK**.

The rows are now sorted from highest fold change value to lowest. You decide to create a gene signature based on the genes with a fold change value above absolute 10.

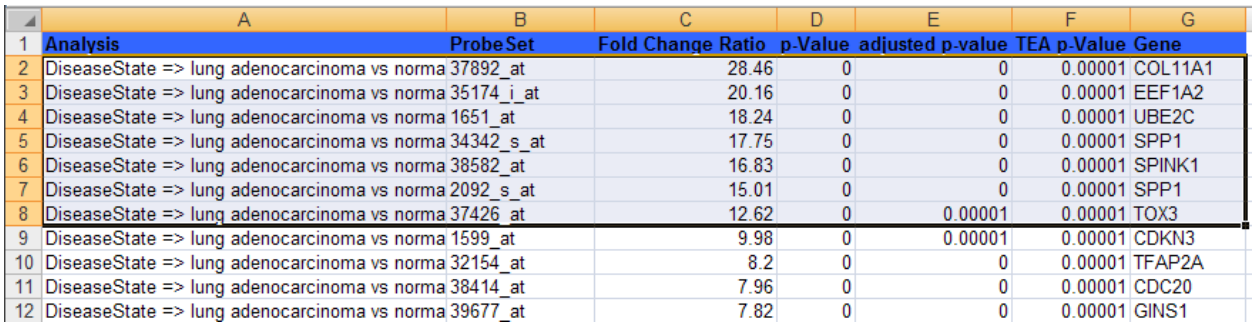
14. Select and copy all rows with a fold change value above 10, as follows:

    a. In column A (the **Analysis** column), click the first cell under the column heading:



	A	B	C	D	E	F	G
1	Analysis	Probe Set	Fold Change Ratio	p-Value	adjusted p-value	TEA p-Value	Gene
2	DiseaseState => lung adenocarcinoma vs norma	37892_at	28.46	0	0	0.00001	COL11A1
3	DiseaseState => lung adenocarcinoma vs norma	35174_i_at	20.16	0	0	0.00001	EEF1A2
4	DiseaseState => lung adenocarcinoma vs norma	1651_at	18.24	0	0	0.00001	UBE2C
5	DiseaseState => lung adenocarcinoma vs norma	34342_s_at	17.75	0	0	0.00001	SPP1
6	DiseaseState => lung adenocarcinoma vs norma	38582_at	16.83	0	0	0.00001	SPINK1
7	DiseaseState => lung adenocarcinoma vs norma	2092_s_at	15.01	0	0	0.00001	SPP1
8	DiseaseState => lung adenocarcinoma vs norma	37426_at	12.62	0	0.00001	0.00001	TOX3
9	DiseaseState => lung adenocarcinoma vs norma	1599_at	9.98	0	0.00001	0.00001	CDKN3
10	DiseaseState => lung adenocarcinoma vs norma	32154_at	8.2	0	0	0.00001	TFAP2A
11	DiseaseState => lung adenocarcinoma vs norma	38414_at	7.96	0	0	0.00001	CDC20
12	DiseaseState => lung adenocarcinoma vs norma	39677_at	7.82	0	0	0.00001	GINS1

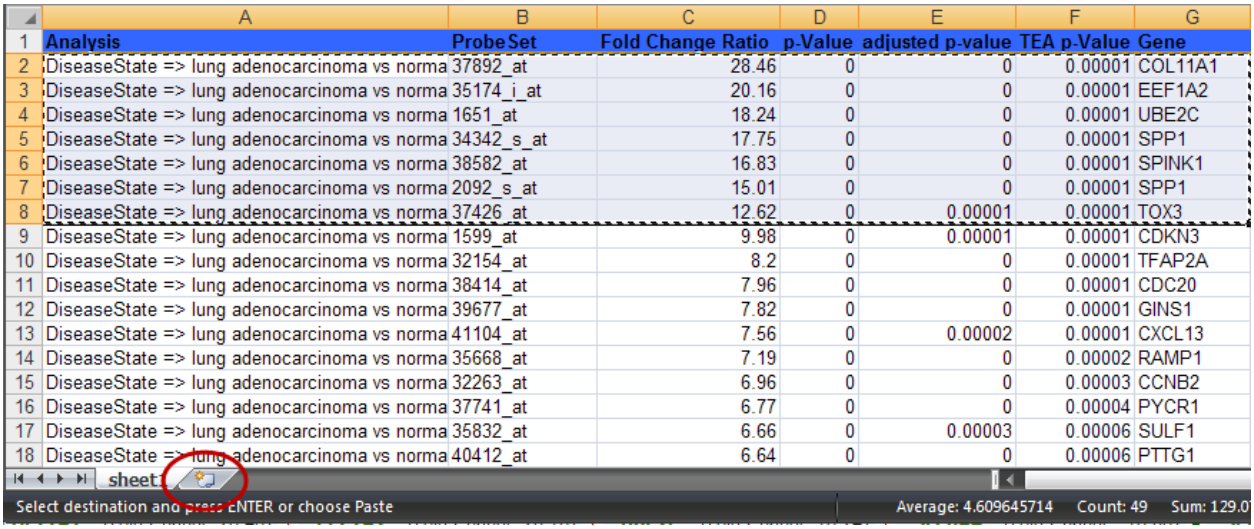
    b. Press and hold down the **Shift** key, then click the cell in column G (the **Gene** column) in the last row with a fold change value above 10:



	A	B	C	D	E	F	G
1	Analysis	Probe Set	Fold Change Ratio	p-Value	adjusted p-value	TEA p-Value	Gene
2	DiseaseState => lung adenocarcinoma vs norma	37892_at	28.46	0	0	0.00001	COL11A1
3	DiseaseState => lung adenocarcinoma vs norma	35174_i_at	20.16	0	0	0.00001	EEF1A2
4	DiseaseState => lung adenocarcinoma vs norma	1651_at	18.24	0	0	0.00001	UBE2C
5	DiseaseState => lung adenocarcinoma vs norma	34342_s_at	17.75	0	0	0.00001	SPP1
6	DiseaseState => lung adenocarcinoma vs norma	38582_at	16.83	0	0	0.00001	SPINK1
7	DiseaseState => lung adenocarcinoma vs norma	2092_s_at	15.01	0	0	0.00001	SPP1
8	DiseaseState => lung adenocarcinoma vs norma	37426_at	12.62	0	0.00001	0.00001	TOX3
9	DiseaseState => lung adenocarcinoma vs norma	1599_at	9.98	0	0.00001	0.00001	CDKN3
10	DiseaseState => lung adenocarcinoma vs norma	32154_at	8.2	0	0	0.00001	TFAP2A
11	DiseaseState => lung adenocarcinoma vs norma	38414_at	7.96	0	0	0.00001	CDC20
12	DiseaseState => lung adenocarcinoma vs norma	39677_at	7.82	0	0	0.00001	GINS1

    c. Press the **Ctrl + C** keys to copy the selected rows.

15. At the bottom of the Excel window, click the **Insert Worksheet** icon ( ) to open a new worksheet:



	A	B	C	D	E	F	G
1	Analysis	Probe Set	Fold Change Ratio	p-Value	adjusted p-value	TEA p-Value	Gene
2	DiseaseState => lung adenocarcinoma vs norma	37892_at	28.46	0	0	0.00001	COL11A1
3	DiseaseState => lung adenocarcinoma vs norma	35174_i_at	20.16	0	0	0.00001	EEF1A2
4	DiseaseState => lung adenocarcinoma vs norma	1651_at	18.24	0	0	0.00001	UBE2C
5	DiseaseState => lung adenocarcinoma vs norma	34342_s_at	17.75	0	0	0.00001	SPP1
6	DiseaseState => lung adenocarcinoma vs norma	38582_at	16.83	0	0	0.00001	SPINK1
7	DiseaseState => lung adenocarcinoma vs norma	2092_s_at	15.01	0	0	0.00001	SPP1
8	DiseaseState => lung adenocarcinoma vs norma	37426_at	12.62	0	0.00001	0.00001	TOX3
9	DiseaseState => lung adenocarcinoma vs norma	1599_at	9.98	0	0.00001	0.00001	CDKN3
10	DiseaseState => lung adenocarcinoma vs norma	32154_at	8.2	0	0	0.00001	TFAP2A
11	DiseaseState => lung adenocarcinoma vs norma	38414_at	7.96	0	0	0.00001	CDC20
12	DiseaseState => lung adenocarcinoma vs norma	39677_at	7.82	0	0	0.00001	GINS1
13	DiseaseState => lung adenocarcinoma vs norma	41104_at	7.56	0	0.00002	0.00001	CXCL13
14	DiseaseState => lung adenocarcinoma vs norma	35668_at	7.19	0	0	0.00002	RAMP1
15	DiseaseState => lung adenocarcinoma vs norma	32263_at	6.96	0	0	0.00003	CCNB2
16	DiseaseState => lung adenocarcinoma vs norma	37741_at	6.77	0	0	0.00004	PYCR1
17	DiseaseState => lung adenocarcinoma vs norma	35832_at	6.66	0	0.00003	0.00006	SULF1
18	DiseaseState => lung adenocarcinoma vs norma	40412_at	6.64	0	0	0.00006	PTTG1

sheet1

Select destination and press ENTER or choose Paste Average: 4.609645714 Count: 49 Sum: 129.0

16. Press the **Ctrl + V** keys to paste the selected rows at the top of the new worksheet.

17. Return to the original worksheet.

18. Repeat Step 9 through Step 14, but this time sort from smallest fold change value to largest.

19. After copying the selected rows with **Ctrl + C**, return to the new worksheet.

20. Paste the selected rows (**Ctrl + V**) in the first empty row below the data you previously pasted.

    The new worksheet now looks as follows:

	A	B	C	D	E	F	G
1	DiseaseSt	37892_at	28.46	0	0	0.00001	COL11A1
2	DiseaseSt	35174_i_at	20.16	0	0	0.00001	EEF1A2
3	DiseaseSt	1651_at	18.24	0	0	0.00001	UBE2C
4	DiseaseSt	34342_s_a	17.75	0	0	0.00001	SPP1
5	DiseaseSt	38582_at	16.83	0	0	0.00001	SPINK1
6	DiseaseSt	2092_s_at	15.01	0	0	0.00001	SPP1
7	DiseaseSt	37426_at	12.62	0	0.00001	0.00001	TOX3
8	DiseaseSt	34604_at	-18.08	0	0	0.00001	SLC6A4
9	DiseaseSt	773_at	-16.65	0	0	0.00001	MYH11
10	DiseaseSt	35868_at	-16.63	0	0	0.00001	AGER
11	DiseaseSt	38430_at	-13.92	0	0	0.00001	FABP4
12	DiseaseSt	34174_s_a	-13.22	0	0	0.00001	LPHN2
13	DiseaseSt	32527_at	-12.41	0	0	0.00001	C10orf116
14	DiseaseSt	39066_at	-11.81	0	0	0.00001	MFAP4
15	DiseaseSt	39577_at	-11.74	0	0	0.00001	SOSTDC1
16	DiseaseSt	37777_at	-10.87	0	0	0.00001	PTPRB
17	DiseaseSt	34637_f_at	-10.73	0	0	0.00001	ADH1A
18	DiseaseSt	34708_at	-10.61	0	0	0.00001	FCN3

    In the next steps, you will organize the rows of data in the new spreadsheet into a format that can be imported into the gene signature, and then write the formatted data to a text file.

21. Select the column containing the fold change values (column C in the above figure – click the column letter to select the column).

22. Right-click the column of fold change values, then select **Cut**.

23. Select the column immediately to the right of the column containing the gene names.

24. Right-click in the column you just selected, then select **Paste**.

The columns of data should now appear as shown below:

	A	B	C	D	E	F	G	H
1	DiseaseSt	37892_at		0	0	0.00001	COL11A1	28.46
2	DiseaseSt	35174_i_at		0	0	0.00001	EEF1A2	20.16
3	DiseaseSt	1651_at		0	0	0.00001	UBE2C	18.24
4	DiseaseSt	34342_s_at		0	0	0.00001	SPP1	17.75
5	DiseaseSt	38582_at		0	0	0.00001	SPINK1	16.83
6	DiseaseSt	2092_s_at		0	0	0.00001	SPP1	15.01
7	DiseaseSt	37426_at		0	0.00001	0.00001	TOX3	12.62
8	DiseaseSt	34604_at		0	0	0.00001	SLC6A4	-18.08
9	DiseaseSt	773_at		0	0	0.00001	MYH11	-16.65
10	DiseaseSt	35868_at		0	0	0.00001	AGER	-16.63
11	DiseaseSt	38430_at		0	0	0.00001	FABP4	-13.92
12	DiseaseSt	34174_s_at		0	0	0.00001	LPHN2	-13.22
13	DiseaseSt	32527_at		0	0	0.00001	C10orf116	-12.41
14	DiseaseSt	39066_at		0	0	0.00001	MFAP4	-11.81
15	DiseaseSt	39577_at		0	0	0.00001	SOSTDC1	-11.74
16	DiseaseSt	37777_at		0	0	0.00001	PTPRB	-10.87
17	DiseaseSt	34637_f_at		0	0	0.00001	ADH1A	-10.73
18	DiseaseSt	34708_at		0	0	0.00001	FCN3	-10.61

You will now delete all columns except for the **Gene** and **Fold Change Ratio** columns (in the figure above, columns G and H, respectively).

25. Click column A to select it.

    Ensure that only column A is selected (highlighted).

26. Press and hold down the **Shift** key, then click the column (F in the above figure) immediately to the left of the **Gene** column.

27. Right-click anywhere in the selected area, then select **Delete**.

    The remaining data appears as follows:

	A	B
1	COL11A1	28.46
2	EEF1A2	20.16
3	UBE2C	18.24
4	SPP1	17.75
5	SPINK1	16.83
6	SPP1	15.01
7	TOX3	12.62
8	SLC6A4	-18.08
9	MYH11	-16.65
10	AGER	-16.63
11	FABP4	-13.92
12	LPHN2	-13.22
13	C10orf116	-12.41
14	MFAP4	-11.81
15	SOSTDC1	-11.74
16	PTPRB	-10.87
17	ADH1A	-10.73
18	FCN3	-10.61

28. Click the Office button (), then click **Save As > Other Formats**.

    The Save As dialog appears.

29. In the **Save in** field at the top of the dialog, select the root directory on the C:\ drive – for example:

    `Local Disk (C:)`

30. In the **File name** field, type **lung adenocarcinoma vs normal**.

31. In the **Save as type** field, select **Text (Tab delimited) (*.txt)**.

32. Click **Save**.

    A warning dialog appears, advising you that you can only save the active worksheet.

33. Click **OK** to acknowledge the warning.

    Another warning dialog appears, informing you that some features will be lost when saving to a text file.

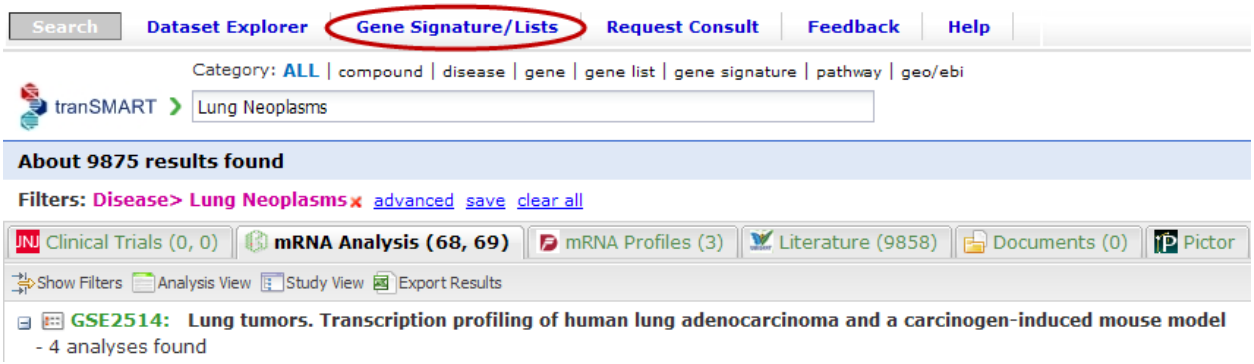
34. Click **Yes** to acknowledge the warning.

35. Close Excel.

36. Click **No** if prompted to save the file. It has already been saved.

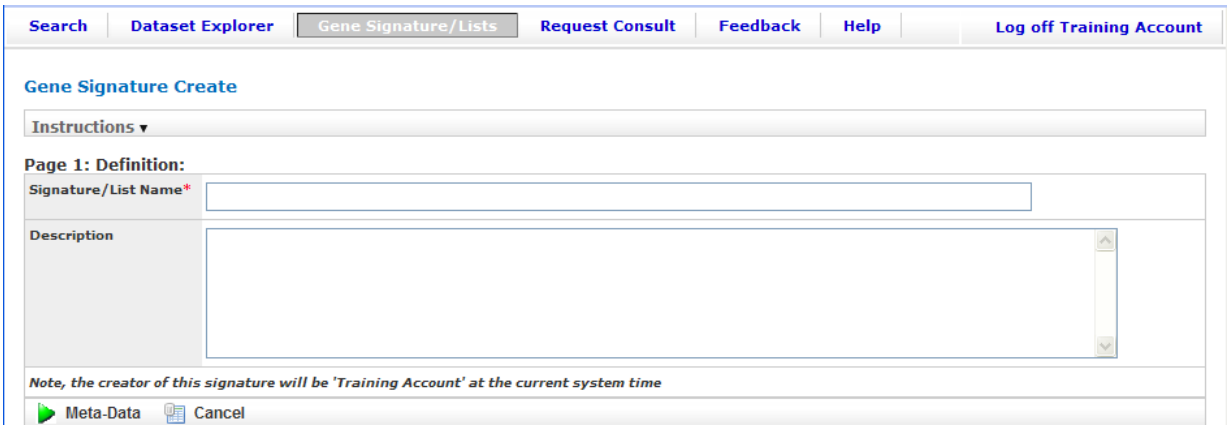
Task 2: Define the Gene Signature and Import the Gene File

1. Click the tranSMART **Gene Signature/Lists** tab to open the Gene Signature tool:

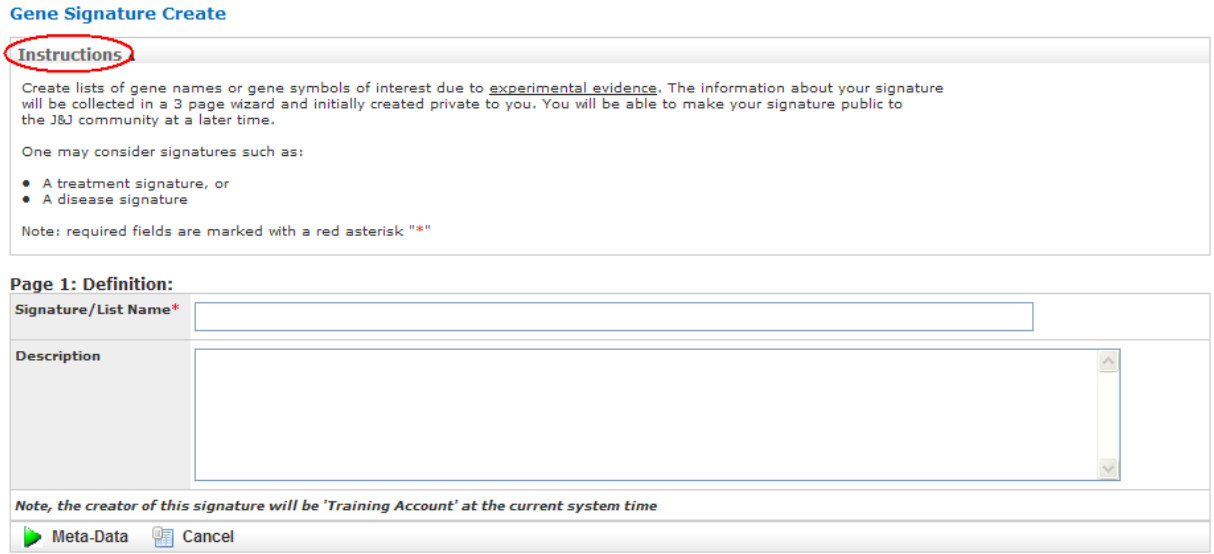


2. Click the **New Signature** button.

The first page of the gene signature wizard appears:



3. Click **Instructions** to read the instructions for creating a gene signature:



4. In **Signature/List Name**, type **<Your Training ID> Gene Signature**.

   For example, if your training ID is **jnjuser9**, type:

   JnJUser9 Gene Signature

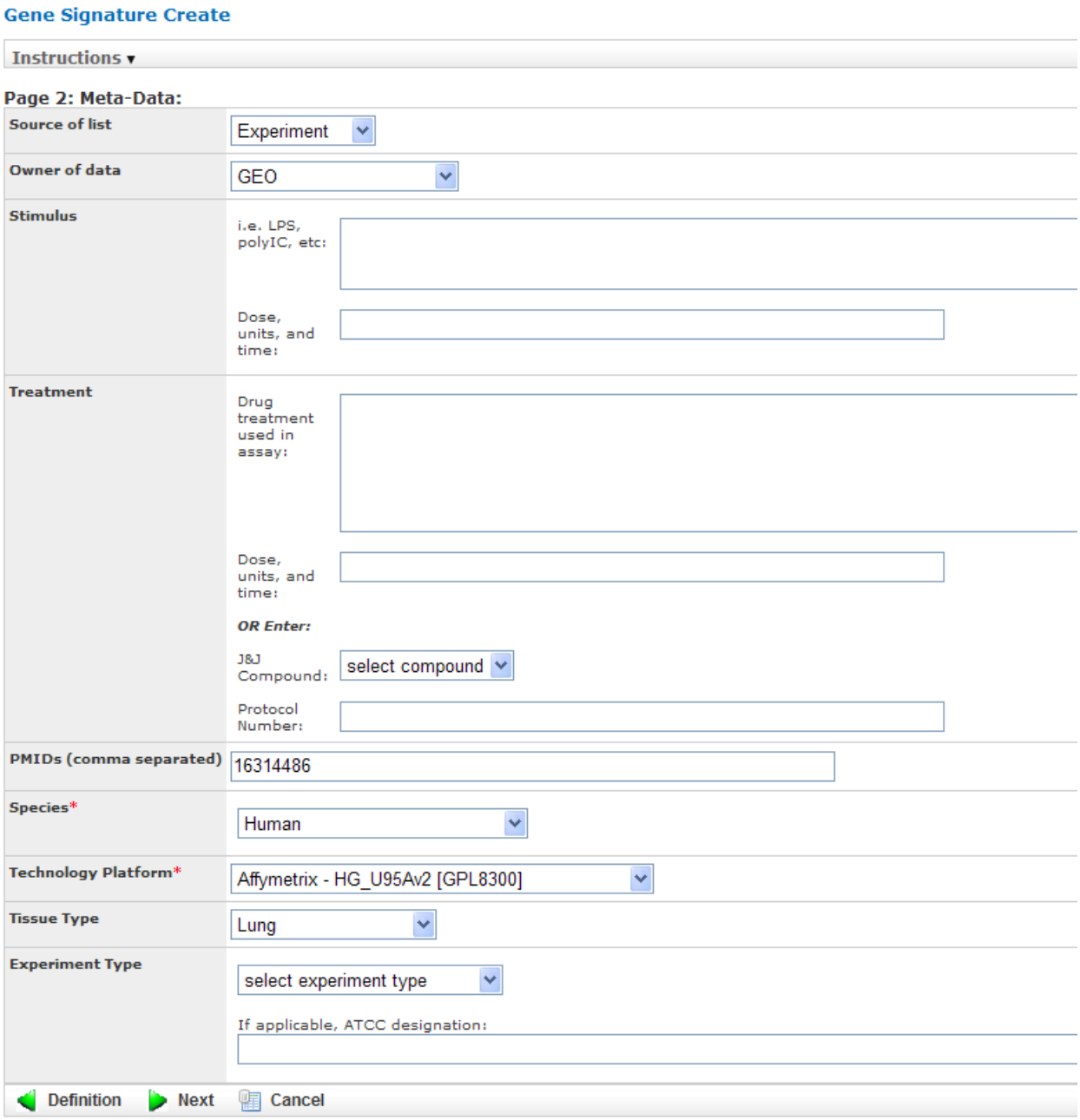
5. In **Description**, type the following text:

   `Genes from` **`lung adenocarcinoma experiment with fold change value above absolute 10.`**

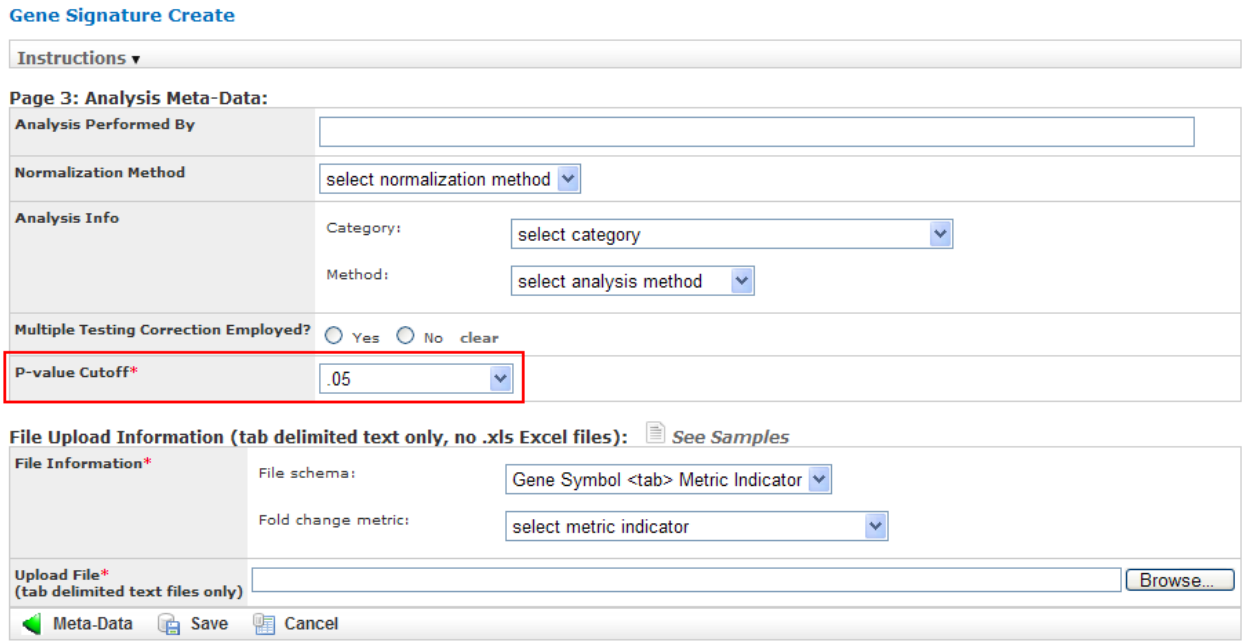
6. Click **Meta-Data** to proceed to the next wizard page.
7. In **Source of list**, select **Experiment**.
8. In **Owner of data**, select **GEO**.
9. Scrolling down to **PMIDs**, type **16314486**.
10. In **Species**, select **Human**.

11. In **Technology Platform**, select **Affymetrix - HG_U95Av2 [GPL8300]**.
12. In **Tissue Type**, select **Lung**.

The wizard page now appears as follows:



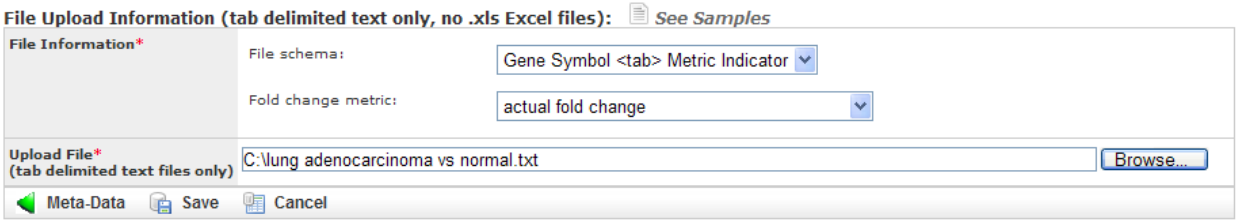
13. Click **Next** to proceed to the final wizard page.
14. In **P-value Cutoff**, select .05:



You are now ready to specify the format of the gene file you created in Task 1, and then upload the file into the gene signature.

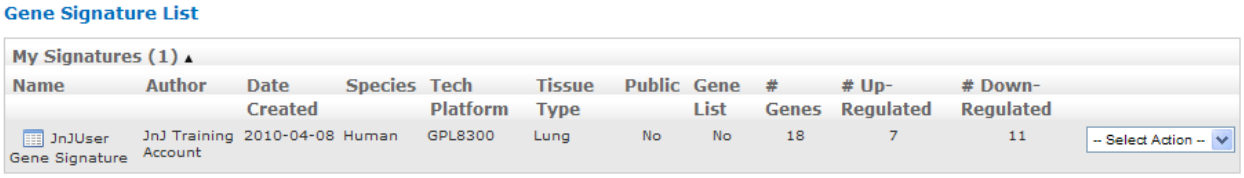
15. In **File Information**:
    - Select **Gene Symbol <tab> Metric Indicator** in **File schema**.
    - Select **actual fold change** in **Fold change metric**.
16. Click the **Browse** button to the right of the **Upload File** field.
17. In the Choose File dialog, navigate to the **C:** directory and select the file **lung adenocarcinoma vs normal.txt**.
18. Click **Open**.

The path and file name of the **lung adenocarcinoma vs normal.txt** file now appears in the **Upload File** field:



19. Click **Save** to save the new gene signature.

    The new signature now appears in the **My Signatures** section of the gene signature list:



20. Click the **Select Action** dropdown at the right of the gene signature entry to see a list of the actions you can perform on the gene signature:

    Notice that one of the actions is **Make Public**. If a gene signature is private, only you or an administrator can view it and use it as a filter in a tranSMART Search operation (as you will do in the next lesson). If a gene signature is public, anyone can view it and use it as a search filter.

> **Note:** You will use your new gene signature in the next lesson.

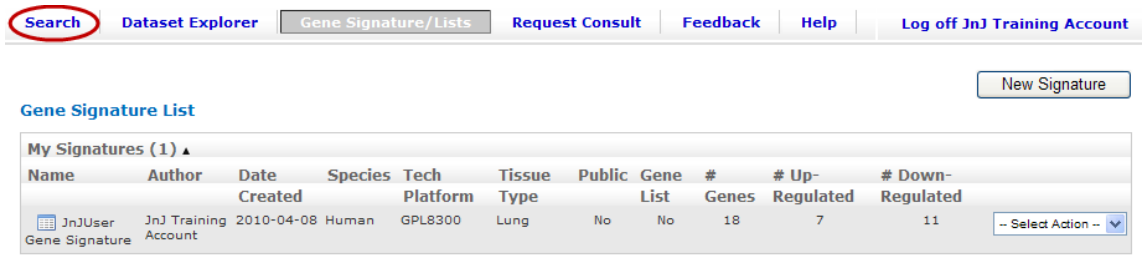
Lesson 3

Search for Studies Using a New Gene Signature as a Filter

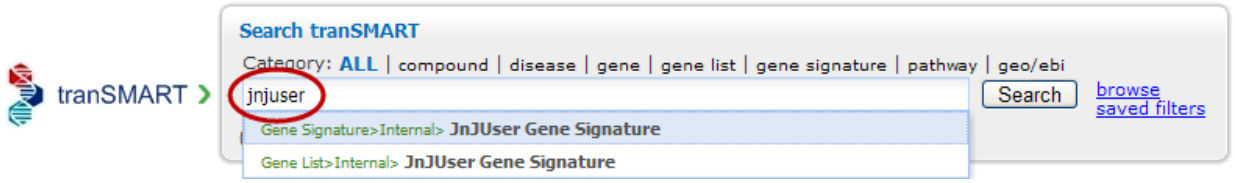
Lesson Goal: Use a newly created gene signature to generate hypotheses in tranSMART.

Scenario: You want to find studies where the differentially regulated genes overlap with the genes contained in your new gene signature. This will generate a set of hypotheses about diseases or treatments that may have similar genes disregulated, and that can help you develop a further set of experiments.

1. Click the tranSMART **Search** tab to display the Search window.



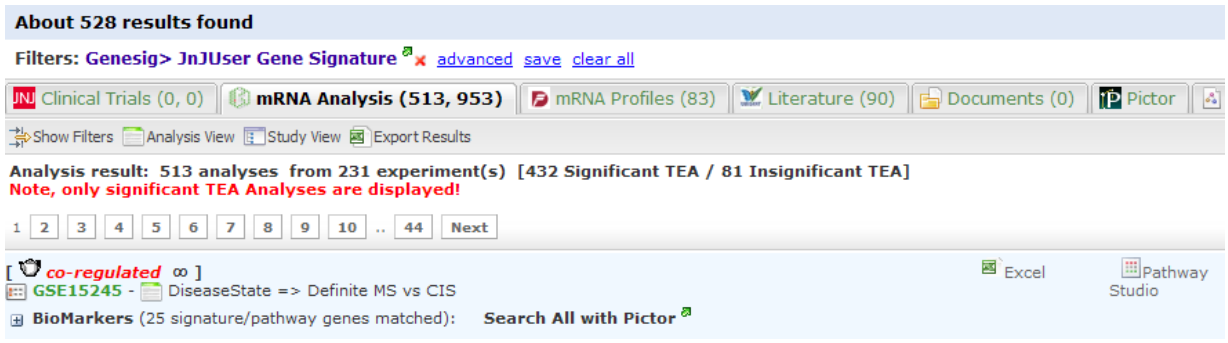
2. Type **jnjuser** in the Search field:



Note: The following steps use **jnjuser** as the login ID that is used as part of the gene signature name. Substitute your own login ID for **jnjuser**.

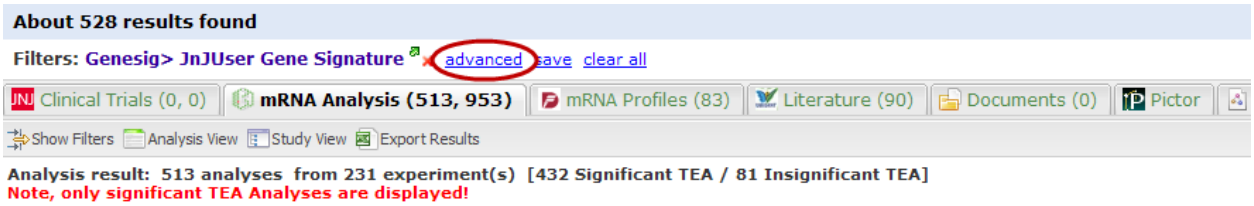
3. Click **Gene Signature>Internal> JnJUser Gene Signature** in the dropdown list.

   The search results include studies involving genes that matched the ones in your gene signature:

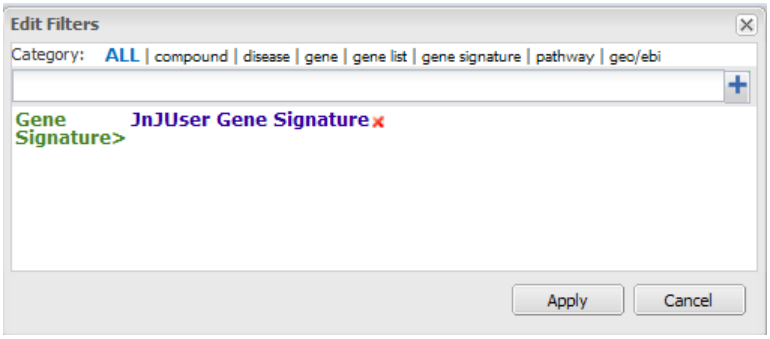


You are only interested in studies related to lung neoplasms, so you want to filter the results further.

4. Click **advanced**:

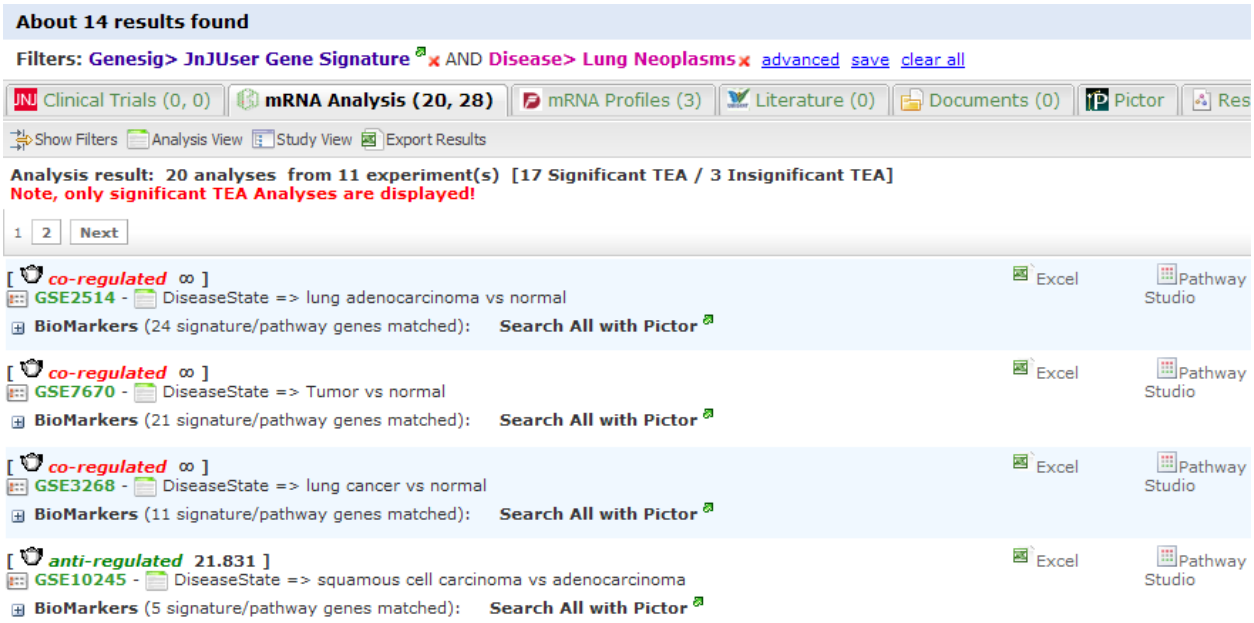


The Edit Filters dialog appears:



5. Type **lung neo** in the search field.
6. Click **Disease> Lung Neoplasms** in the dropdown list.
7. Click **Apply**.

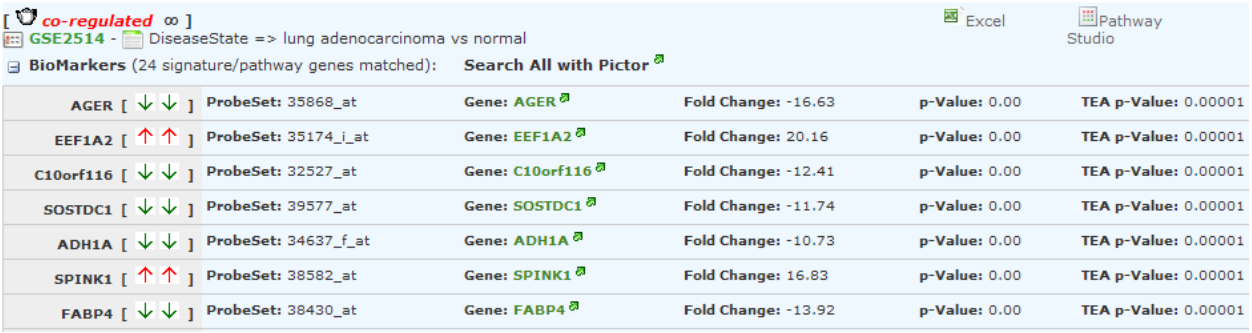
The following figure shows a portion of the mRNA Analysis results. The analyses that are returned involve both lung neoplasms and one or more genes in your gene signature:



Now you want to browse through the analyses to see how the genes that match those in your gene signature behaved during the experiments.

In the figure above, the analysis **lung adenocarcinoma vs normal** in experiment **GSE2514** is the first in the list. This is the analysis from which you derived the genes for your gene signature during the previous exercise.

8. Click the **+** icon (⊞) to the left of the label **BioMarkers** for the analysis **lung adenocarcinoma vs normal**. A partial list of the matching genes is shown below:


[co-regulated ∞]
GSE2514 - DiseaseState => lung adenocarcinoma vs normal
Excel Pathway Studio
BioMarkers (24 signature/pathway genes matched): Search All with Pictor

AGER [↓ ↓]	ProbeSet: 35868_at	Gene: AGER	Fold Change: -16.63	p-Value: 0.00	TEA p-Value: 0.00001
EEF1A2 [↑ ↑]	ProbeSet: 35174_i_at	Gene: EEF1A2	Fold Change: 20.16	p-Value: 0.00	TEA p-Value: 0.00001
C10orf116 [↓ ↓]	ProbeSet: 32527_at	Gene: C10orf116	Fold Change: -12.41	p-Value: 0.00	TEA p-Value: 0.00001
SOSTDC1 [↓ ↓]	ProbeSet: 39577_at	Gene: SOSTDC1	Fold Change: -11.74	p-Value: 0.00	TEA p-Value: 0.00001
ADH1A [↓ ↓]	ProbeSet: 34637_f_at	Gene: ADH1A	Fold Change: -10.73	p-Value: 0.00	TEA p-Value: 0.00001
SPINK1 [↑ ↑]	ProbeSet: 38582_at	Gene: SPINK1	Fold Change: 16.83	p-Value: 0.00	TEA p-Value: 0.00001
FABP4 [↓ ↓]	ProbeSet: 38430_at	Gene: FABP4	Fold Change: -13.92	p-Value: 0.00	TEA p-Value: 0.00001

Notice that the list of genes in the analysis contains the same genes as in your gene signature, the same gene expression values (for values above absolute 10), and the same probe sets that produced the expression results in your gene signature.

You now want to browse through the list of genes in the other analyses. You are interested in finding those genes whose expressions were produced by the same probe set as in your gene signature.

If you compared the genes and probe sets for each of the analyses to the ones in your gene signature, you would find that the most strongly up-regulated gene in your gene signature, COL11A1, is the only gene that is associated with the same probe set (37892_at) in both your gene signature and in any of the analyses.

You decide to view a profile of COL11A1 to learn more about it.

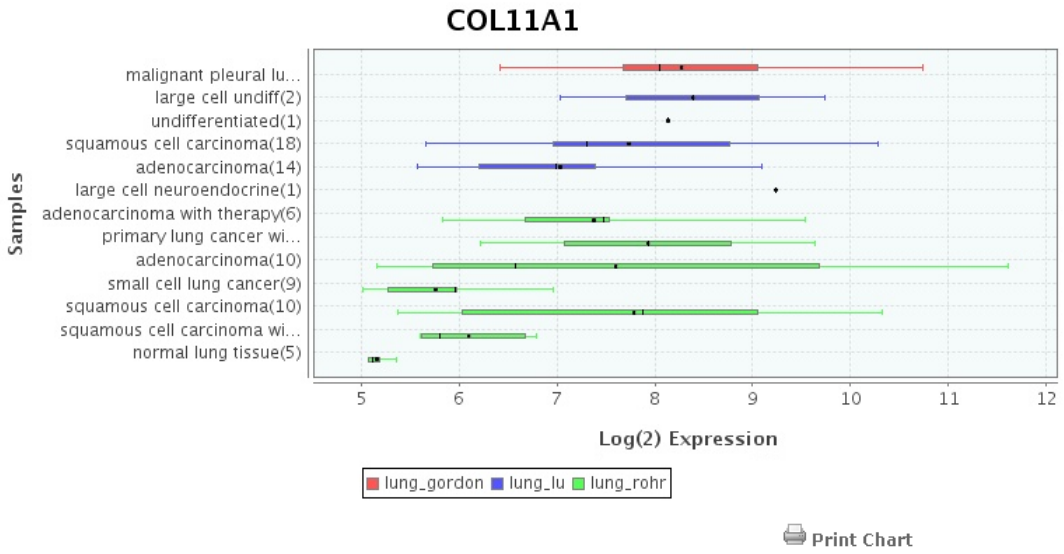
9. Click the **mRNA Profiles** tab.
10. Select **COL11A1** from the **Gene** dropdown.
11. Select **Lung Neoplasms** from the **Disease** dropdown.
12. Select **37892_at** from the **Probe Set** dropdown.

This is the probe set that produced the expression results for COL11A1 in your pathway and in some of the analyses.

The following chart and the datasets related to the COL11A1 gene are displayed for your further study:

Filter (Note: search found 3 studies)

Gene:	COL11A1
Disease:	Lung Neoplasms
Probe Set:	37892_at



Print Chart

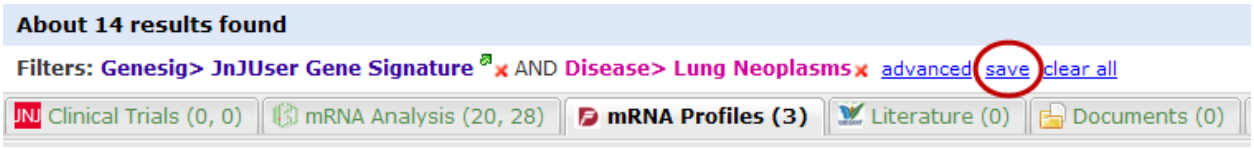
Information on individual datasets

Dataset	No. Samples	Experiment
malignant pleural lung mesothelioma	33	**lung_gordon:** Translation of microarray data into clinically relevant cancer diagnostic tests mesothelioma
large cell undiff	2	**lung_lu:** A gene expression signature predicts survival of patients with stage I non-small
undifferentiated	1	**lung_lu:** A gene expression signature predicts survival of patients with stage I non-small
squamous cell carcinoma	18	**lung_lu:** A gene expression signature predicts survival of patients with stage I non-small
adenocarcinoma	14	**lung_lu:** A gene expression signature predicts survival of patients with stage I non-small
squamous cell carcinoma	18	**lung_lu:** A gene expression signature predicts survival of patients with stage I non-small
large cell neuroendocrine	1	**lung_lu:** A gene expression signature predicts survival of patients with stage I non-small
adenocarcinoma with therapy	6	**lung_rohr:** Genetic programming and gene expression profling for molecular discriminatior
primary lung cancer with therapy	2	**lung_rohr:** Genetic programming and gene expression profling for molecular discriminatior
adenocarcinoma	10	**lung_rohr:** Genetic programming and gene expression profling for molecular discriminatior
small cell lung cancer	9	**lung_rohr:** Genetic programming and gene expression profling for molecular discriminatior
squamous cell carcinoma	10	**lung_rohr:** Genetic programming and gene expression profling for molecular discriminatior
squamous cell carcinoma with therapy	5	**lung_rohr:** Genetic programming and gene expression profling for molecular discriminatior
normal lung tissue	5	**lung_rohr:** Genetic programming and gene expression profling for molecular discriminatior

Data courtesy of **Dana-Farber Cancer Institute GeneChip Oncology Database**.

You decide to save the search string.

13. Click **save**:



The Create Filter window appears.

14. In **Name**, type **Gene Signature Search**.
15. In Description, type **Lung neoplasm experiments involving genes in my gene signature**.
16. Check the **Private** check box to keep the saved query private.
17. Click **Create**.
18. Click **Return to Search**.

You will use the saved query in an upcoming lesson.

Lesson 4

Use a Heat Map to Compare Treatment Results

Lesson Goal: Use the tranSMART Dataset Explorer to create a heat map, and save the heat map data to a file.

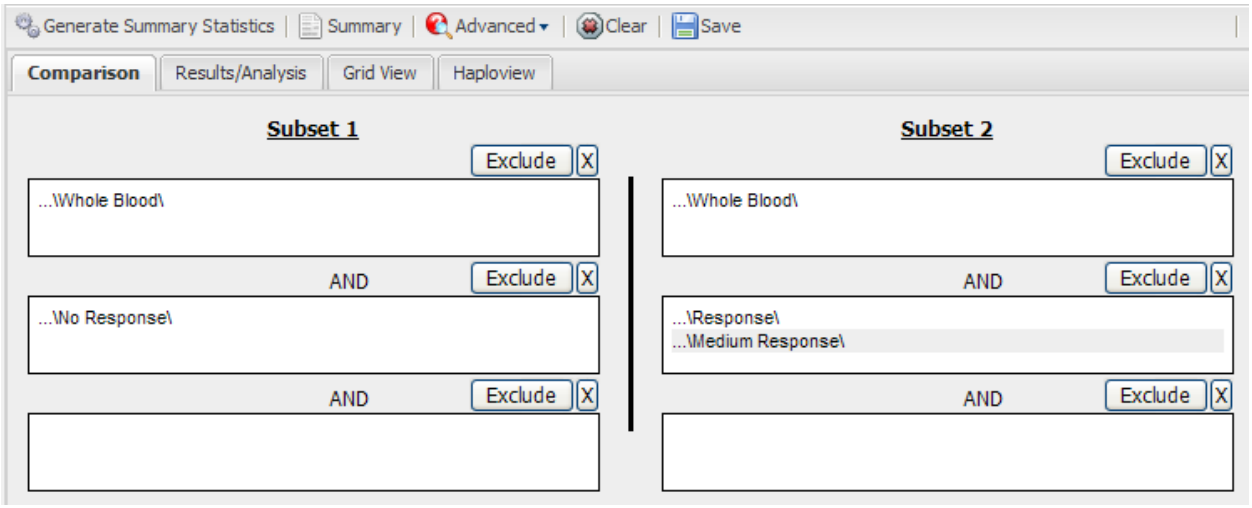
Scenario: You are analyzing the results of a study on rheumatoid arthritis. You want to see a visualization of gene expression data for the gene REL in two cohorts: those who responded to anti-Tnf therapy and those who did not.

> **Note:** For a heat map to be generated in Dataset Explorer, at least one of the subsets must contain the following elements:
>
> - One or more test subjects.
> - A platform of biomarkers (for example, a gene pathway or RBM antigens).

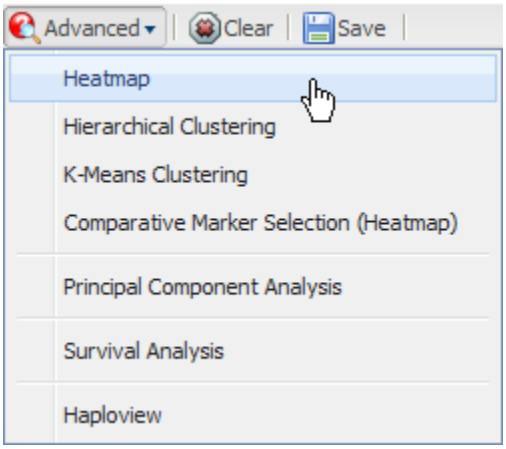
> **Note:** The heat map illustrated in this section was generated with Internet Explorer version 8. If you are using a different browser or a different version of Internet Explorer, you might see slight differences between the illustration and the heat map displayed on your screen.

1. Click the tranSMART **Dataset Explorer** tab to display the Dataset Explorer window.
2. In the left pane of Dataset Explorer, click the **Navigate Terms** tab.
3. In **Public Studies**, open the study **Bienkowska_RheumatoidArthritis_GSE15258**.
4. Open the following nested nodes in the following order:
   a. Biomarker Data
   b. Affymetrix GeneChip Human Genome U133 Plus 20 Array
5. Drag **Whole Blood** into subset definition boxes in Subset 1 and Subset 2.
6. Open the following nested nodes in the following order:
   a. Clinical Data
   b. Response To Anti Tnf Therapy
7. Drag **No Response** into an empty box in Subset 1.
8. Drag **Response** into an empty box in Subset 2.
9. Drag **Medium Response** into the same box where you placed **Response**.

The subset definition boxes look as follows:



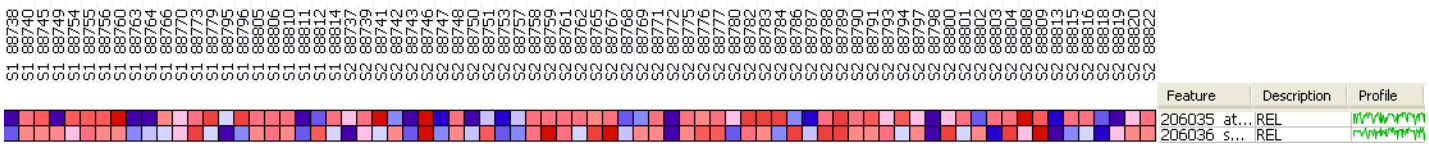
10. Click the **Advanced** tab, then click **Heatmap**:



The Compare Subsets-Pathway Selection dialog appears.

11. Type **rel** in the **Select a Gene/Pathway** field.
12. Click **Gene> REL** in the dropdown list.
13. Click **Run Workflow**.

In a few seconds, the heat map appears in a new browser window:



Notice that:

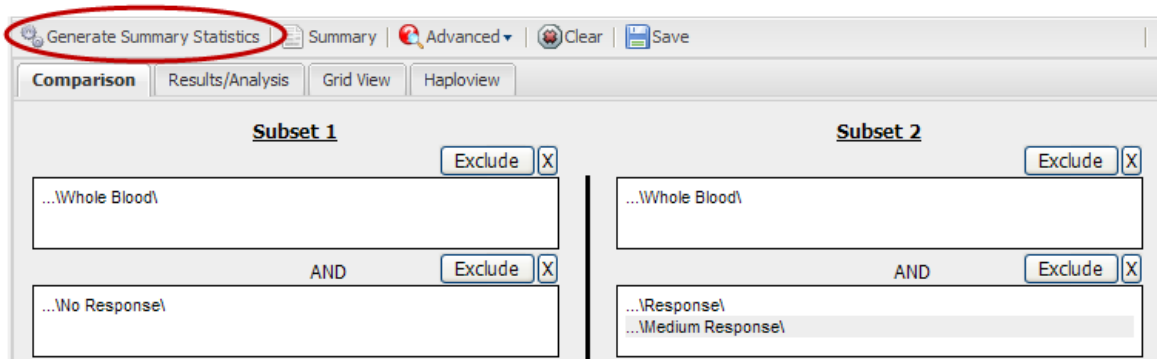
- The column headings represent the subjects in the study.
- The prefix S1_ represents Subset 1 subjects. Prefix S2_ represents Subset 2 subjects. The numbers following the prefixes are the IDs of the subjects.
- REL expression data is represented by the colored cells – up-regulation is expressed in shades of red. Down-regulation is expressed in shades of blue.
- In this example, two probe sets were used, yielding two sets of REL data.

14. When finished comparing the biomarker metrics, close the browser window containing the heat map.

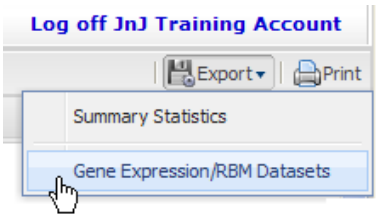
    You decide to save the REL expression data to a Microsoft Excel spreadsheet.

    Note: With private studies, you must have permission from the study owner to save study data to a file.

15. Click **Generate Summary Statistics**:



16. Click **Export**, then click **Gene Expression/RBM Datasets**:



17. When the File Download dialog appears, click **Open**.

    REL data observed in several of the subjects in Subset 1 appear below:

	A	B	C	D	E	F	G	H	I	J	K	L
1	NAME	Descriptic	S1_88738	S1_88740	S1_88745	S1_88749	S1_88754	S1_88755	S1_88756	S1_88760	S1_88763	S1_88764
2	206035_at	REL	-2.47133	0.13106	0.27076	-2.35473	0.24986	0.27869	0.23062	0.80504	-2.3824	-2.47734
3	206036_s_	REL	-1.54764	0.04338	0.10132	0.04961	-0.25269	0.34217	0.1117	0.02272	-1.38417	-0.66737

18. Close the Excel file without saving it.

    Normally you would save the file for future reference.

Lesson 5

Analyze Gene Expression Data from Different Perspectives

Lesson Goal: Create different heat map visualizations of study data.

Scenario: You want to analyze gene expression data for the gene IL6R, collected in a study of patients diagnosed with multiple myeloma. You are particularly interested in the data collected from the study's proliferation group.

In the previous lesson, you created a standard heat map. A standard heat map organizes its data points according to the numeric order of the IDs of the subjects in the subset(s).

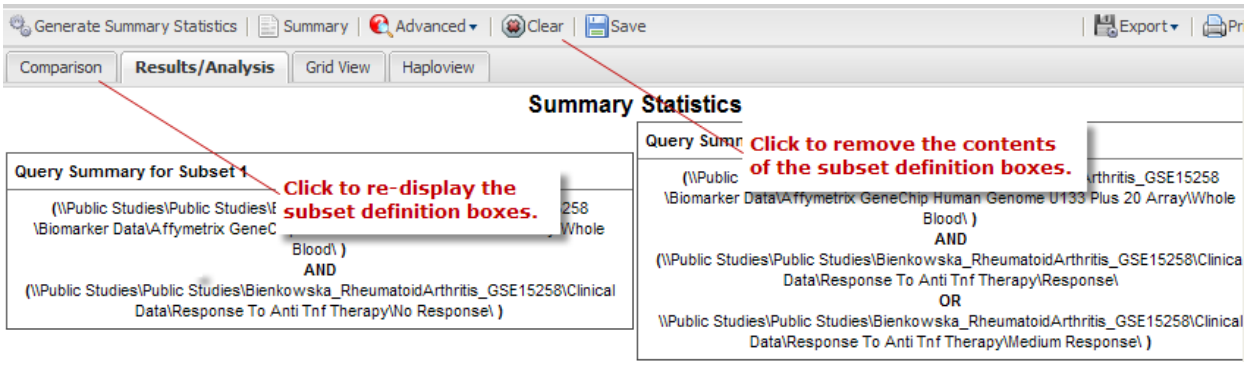
You can also organize a heat map's data points by their gene expression values. The following types of heat maps are organized by gene expression values:

- Class discovery (hierarchical clustering) heat map – A visualization of patterns of related data points in gene expression and RBM data.
- Class discovery (k-means clustering) heat map – A visualization of groupings of the most closely related data points, based on the number of groupings you specify.
- Differential Analysis/Marker Selection heat map – A visualization of differentially expressed genes in distinct phenotypes.

You will generate all these types of heat maps in this lesson.

> **Note:** The heat maps illustrated in this section were generated with Internet Explorer version 8. If you are using a different browser or a different version of Internet Explorer, you might see slight differences between the illustrations and the heat maps displayed on your screen.

1. Click the Dataset Explorer **Comparison** tab, then click the **Clear** button. These actions clear the subset definitions and results from the previous lesson.



2. In **Public Studies**, open the study **Shaughnessy_MultipleMyeloma_GSE2658**.

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

   a. Biomarker Data
   b. Affymetrix GeneChip Human Genome U133 Plus 20 Array

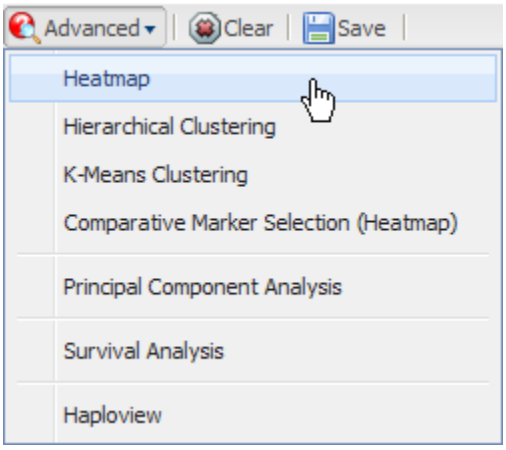
4. Drag **Bone Marrow** into a subset definition box in Subset 1.

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

   a. Published Conclusions
   b. Disease Subtype Classification
   c. RNA

6. Drag **Proliferation group** into an empty box in Subset 1.

7. Click the **Advanced** tab, then click **Heatmap** to generate a standard heat map:



   The Compare Subsets-Pathway Selection dialog appears.

8. Type **il6r** in the **Select a Gene/Pathway** field.

9. Click **Gene> IL6R** in the dropdown list.

10. Click **Run Workflow**.

    In a few seconds, the standard heat map appears in a new browser window:

    Leave the heat map window open, so you can compare it with the other heat maps you will generate. Shades of color are more easily distinguished on your screen than on the black-and-white printed page.

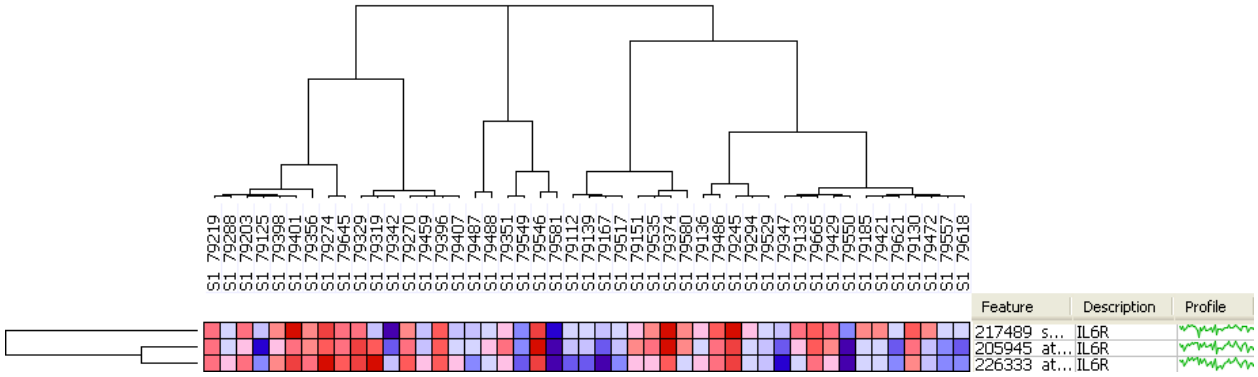
    Now you want to organize the same data in hierarchical patterns of related data points.

11. Click the **Advanced** tab, then click **Hierarchical Clustering**.

    The Compare Subsets-Pathway Selection dialog appears, with IL6R already in the **Select a Gene/Pathway** field.

12. Click **Run Workflow**.

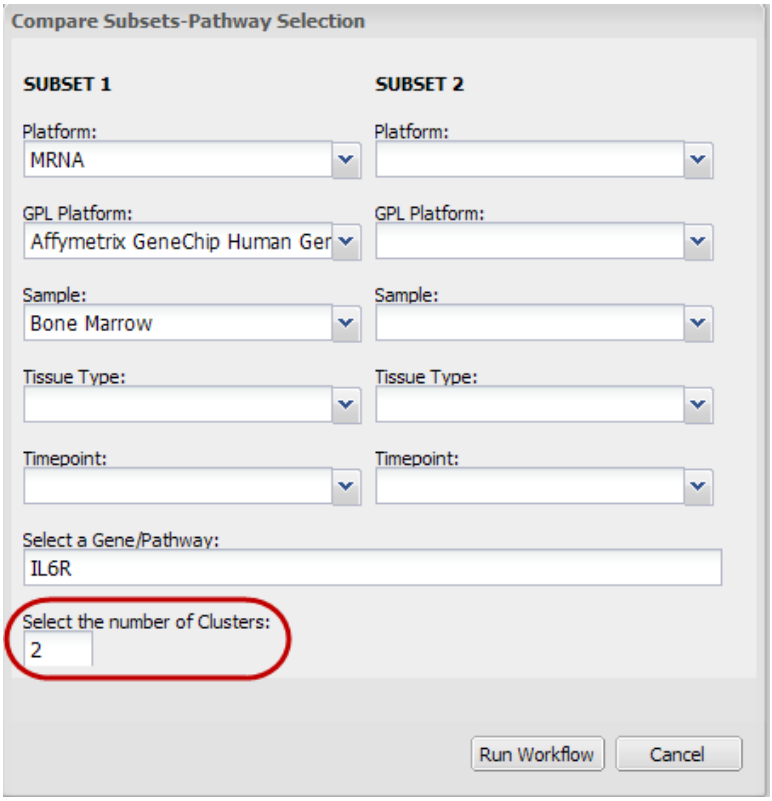
    The hierarchical clustering heat map appears in a new browser window:



    Now you want to organize the same data in three clusters of the most related data points:

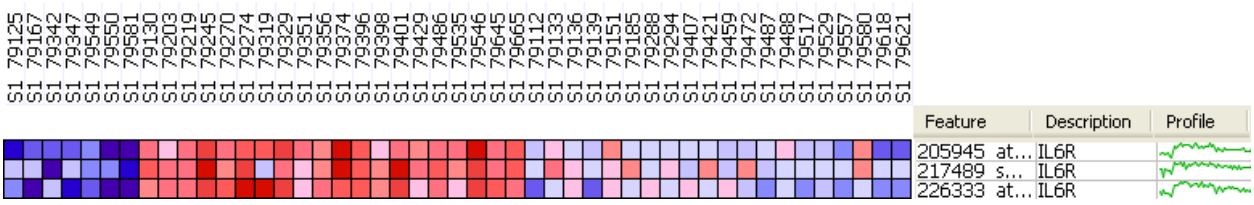
13. Click the **Advanced** tab, then click **K-Means Clustering**.

    The Compare Subsets-Pathway Selection dialog again appears with IL6R in the **Select a Gene/Pathway** field, but it now contains a new field, **Select the number of Clusters**:



14. Type **3** in the **Select the number of Clusters** field, overwriting the default value of **2**.
15. Click **Run Workflow**.

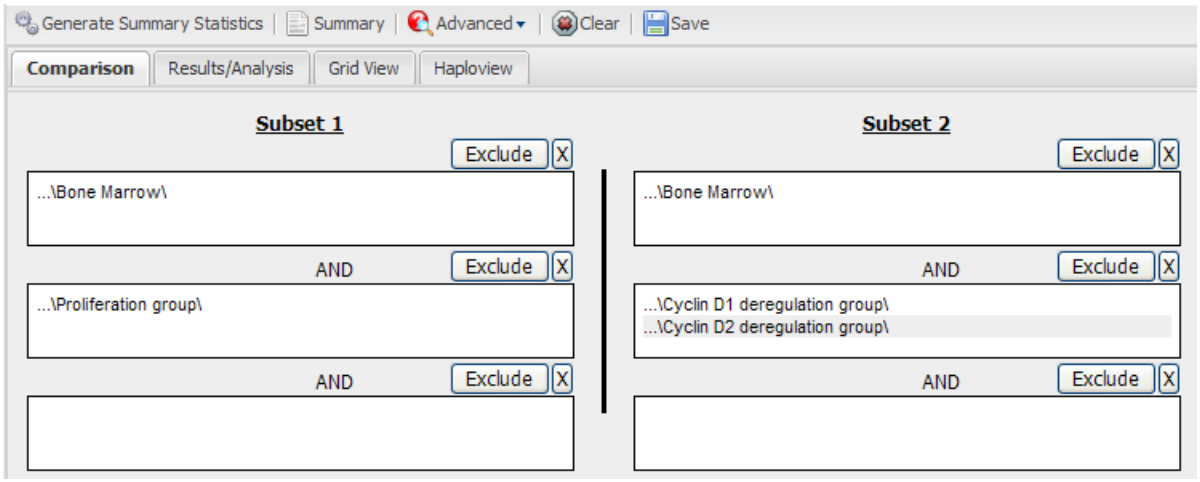
    The k-means clustering heat map appears in a new browser window:



    Finally, you want to generate a heat map of differentially expressed genes in the proliferation group as compared with another set of phenotypes that you define.

16. Return to the node Biomarker Data > Affymetrix GeneChip Human Genome U133 Plus 20 Array, and drag **Bone Marrow** into a subset definition box in Subset 2.
17. Return to the node Published Conclusions > Disease Subtype Classification > RNA, and drag **Cyclin D1 deregulation group** into an empty box in Subset 2.
18. Drag **Cyclin D2 deregulation group** into the same box.

    The subset definition boxes now look as follows:



19. Click the **Advanced** tab, then click **Comparative Marker Selection (Heatmap)**.

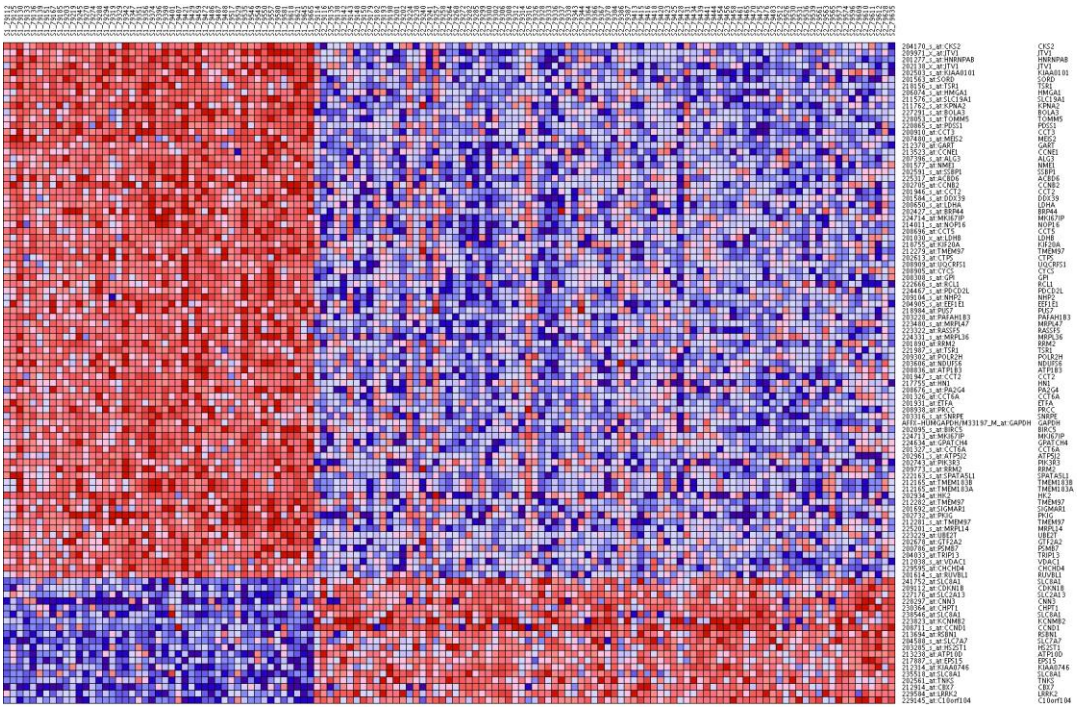
    This time, when the Compare Subsets-Pathway Selection appears, the **Select a Gene/Pathway** field is not on it. With this type of heat map, tranSMART searches for *all differentially expressed* genes between the two subsets.

20. Click **Run Workflow**.

    Note: Due to the large amount of data being searched, the heat map may take several minutes to appear. Meanwhile, explore the interactive features of heat maps. See the section Interactive Heat Maps on page 33.

    Note: If you are using Internet Explorer 8 and a window labeled **Upregulated features** appears, click **View > Heatmap** to display the heat map.

A miniaturized version of the heat map appears below:



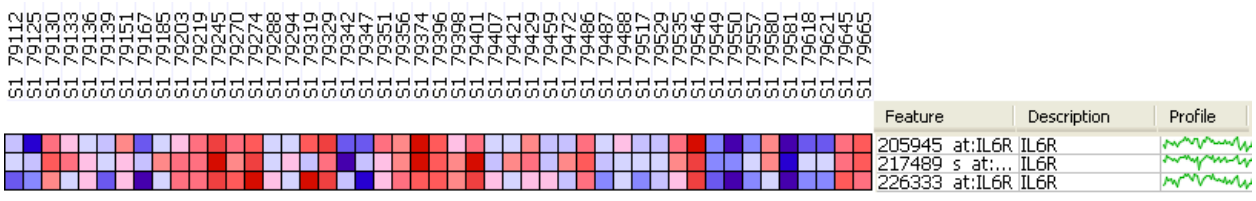
21. Keep the standard heat map open (the one you generated in Step 10), but close the other heat map windows. You will use the standard heat map in the next section.

Interactive Heat Maps

Dataset Explorer heat maps generated with Internet Explorer versions below version 8 are static. With Internet Explorer 8, you can generate interactive heat maps.

Continue with this section if you are using Internet Explorer 8.

Return to the standard heat map you created in Step 2 through Step 10 of the previous section. The heat map looks as follows:



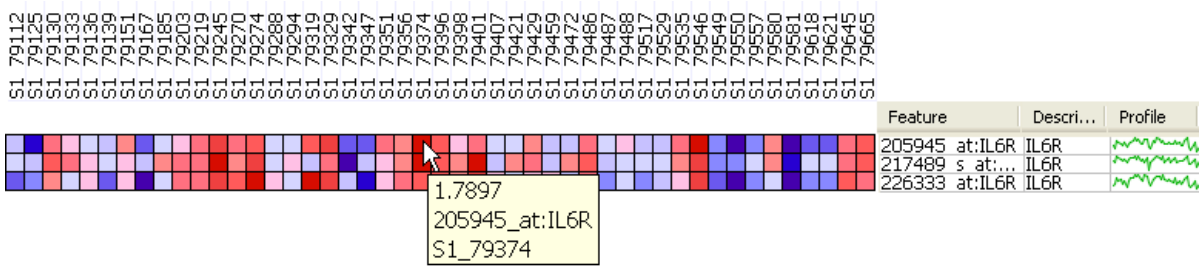
The following sections illustrate some of the features and visualizations available with an interactive heat map.

> **Note:** A standard heat map is used in the following examples, but the same features apply to all interactive heat maps you generate with Dataset Explorer.

View a Particular Data Point

To view a particular data point value:

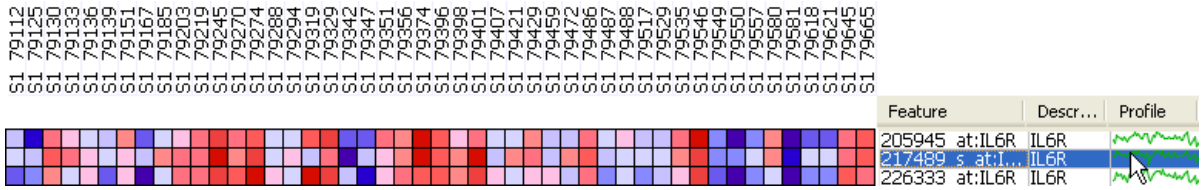
- Hover the mouse pointer over the cell representing the data point of interest:



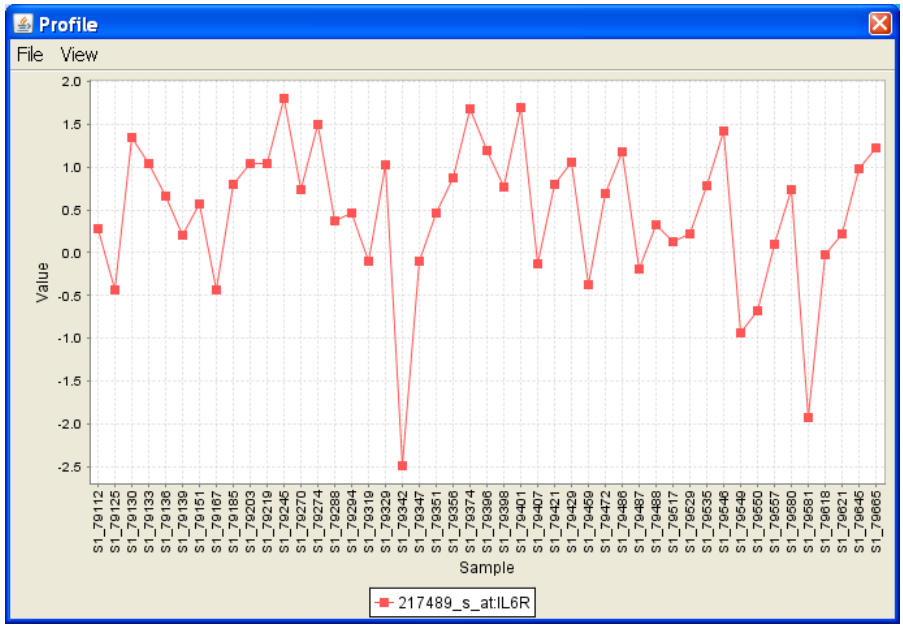
View a Profile of All Data Points for a Probe Set

To view a profile of all data points for a given probe set:

1. In the **Feature** column to the right of the heat map, click probe set **217489_s_at** to select it.
2. Click the green line in the **Profile** column for the selected probe set:



The following profile chart appears:

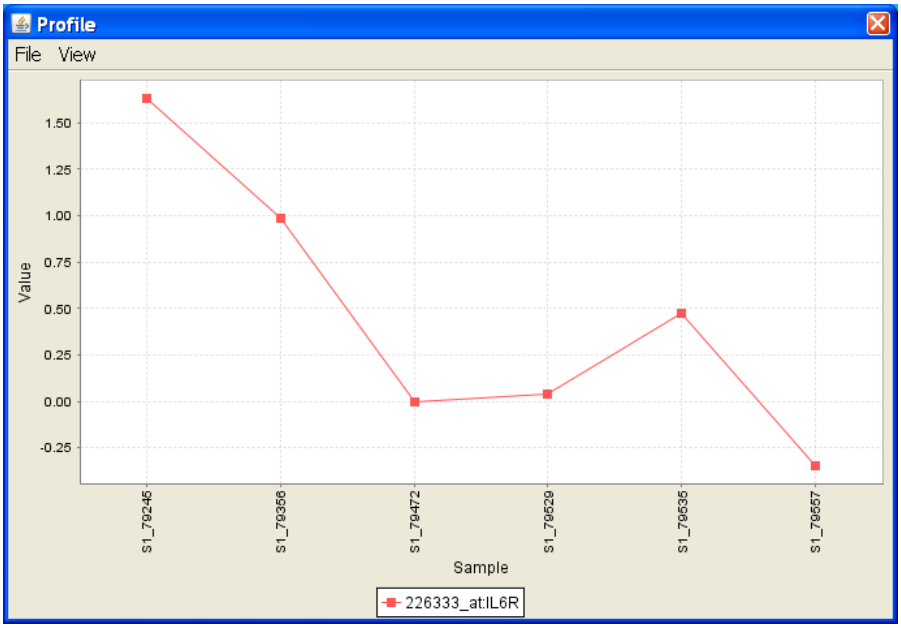


3. When finished analyzing the chart, close the chart window.

View a Profile of Selected Data Points for a Probe Set

To view a profile of selected data points for a given probe set:

1. Hold down the **Ctrl** key, then click the ID of each subject whose data point you want to include in the profile:
2. Click probe set **226333_at** to select it.
3. Click the green line in the **Profile** column for the selected probe set:

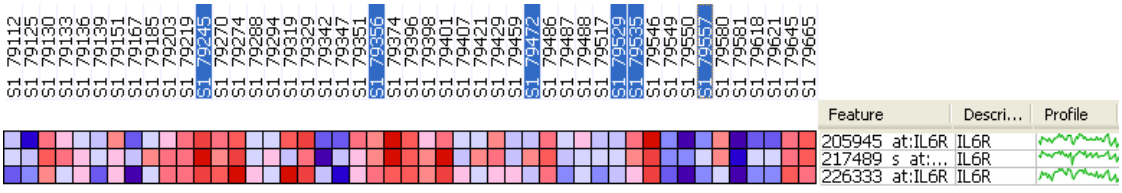


4. When finished analyzing the chart, close the chart window.

View a Profile of Selected Data Points for Multiple Probe Sets

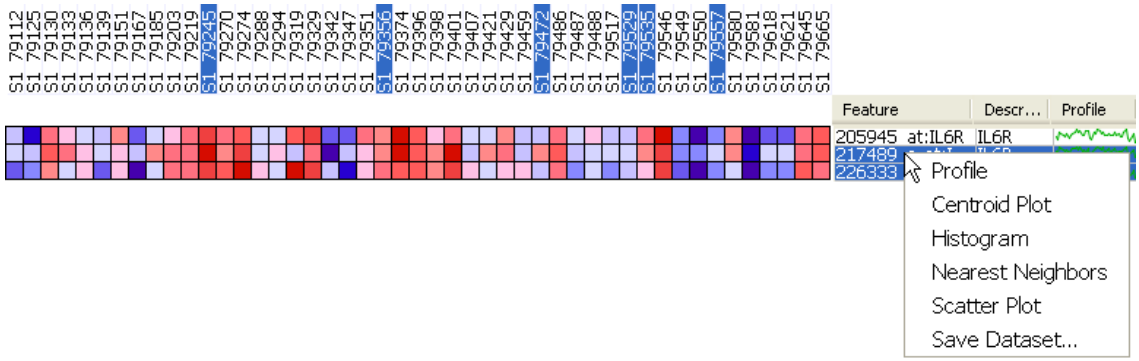
To view a profile of selected data points for multiple probe sets:

1. Hold down the **Ctrl** key, then click the ID of each subject whose data point you want to include in the profile:



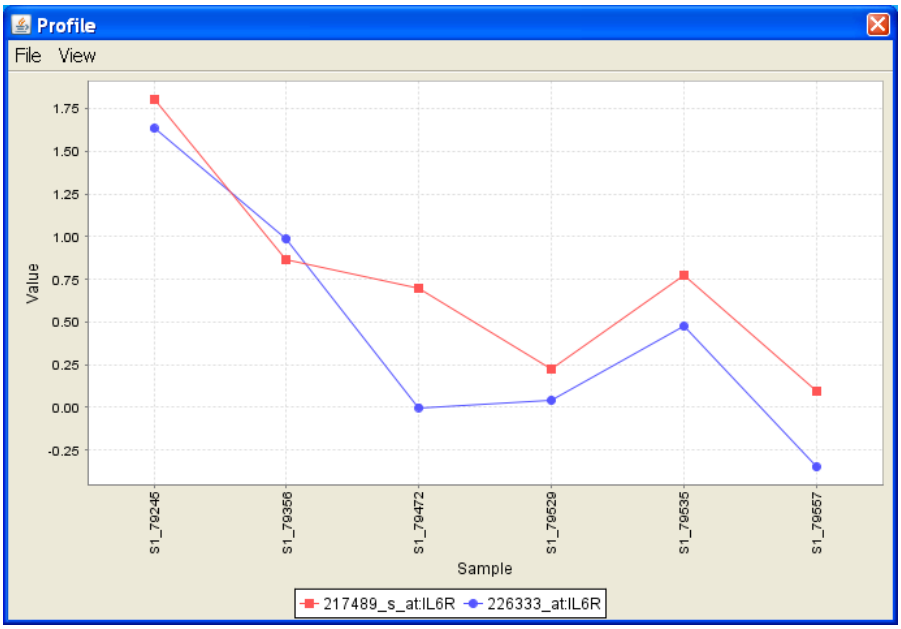
2. Click probe set **217489_s_at** to select it.
3. Hold down the **Ctrl** key, then click probe set **226333_at** to select it:

4. Right-click either of the selected probe sets:



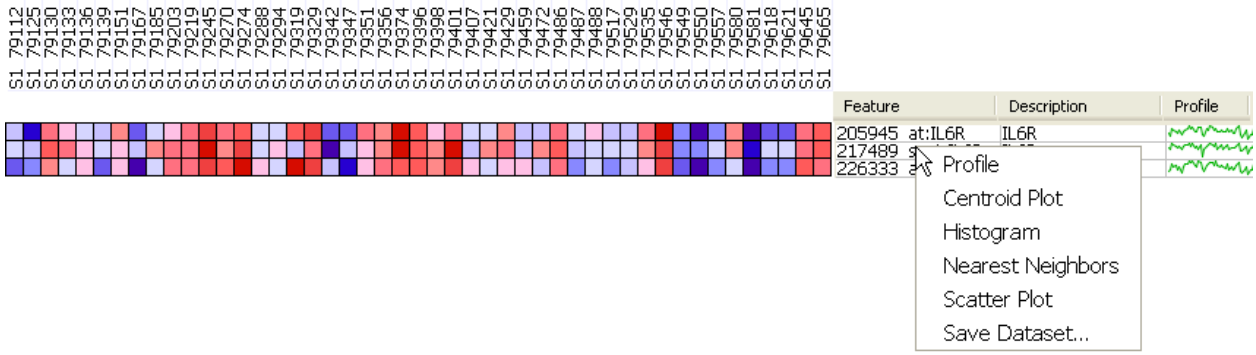
5. Click **Profile**.

   The selected data points for each probe set are profiled:



6. When finished analyzing the chart, close the chart window.

7. In the heat map window, right-click any probe set to view the other operations you can perform:



Lesson 6

Perform a Survival Analysis

Lesson Goal: Generate a survival analysis visualization and statistics.

Scenario: You hypothesize that breast cancer patients with positive estrogen receptors have a better overall survival rate than breast cancer patients with negative estrogen receptors. You want to test that hypothesis against data from a study in Dataset Explorer.

To generate a survival analysis in Dataset Explorer, you must introduce the following criteria into the two groups you are comparing:

- The observed survival times of the individuals in each group.
- The specific event (death) being tracked for the individuals in the study, and optionally, any censoring factors that occurred before the event took place.

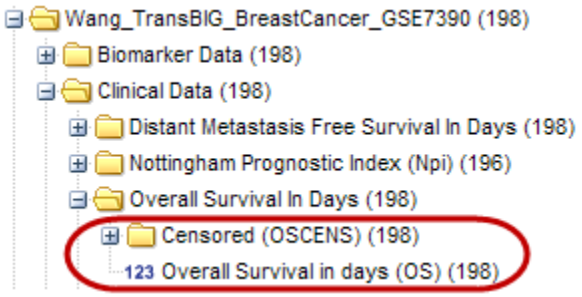
  A censoring factor might be the withdrawal of an individual from the study, or the conclusion of the study before the event occurred for a given individual.
- At least one variable that distinguishes the two groups – in this case, positive vs negative estrogen receptors.

To generate the survival analysis:

1. Click the Dataset Explorer **Comparison** tab, then click the **Clear** button to remove the subset definitions from the previous lesson.
2. In **Public Studies**, open the study **Wang_TransBIG_BreastCancer_GSE7390**.
3. Open the following nested nodes in the following order:

   a. Clinical Data
   b. Overall Survival in Days

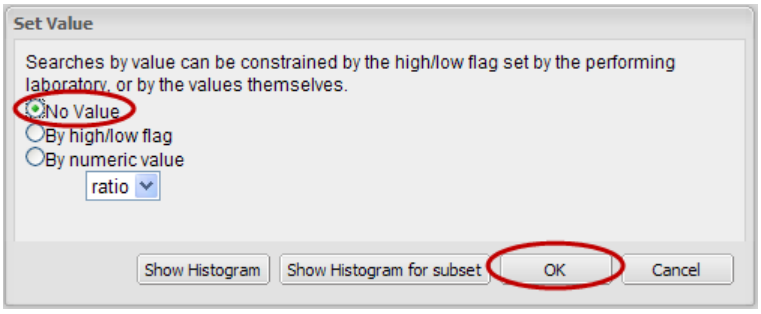
   The folder **Overall Survival in Days** contains two items. In the following steps, you will drag them both into subset boxes.



4. Drag the time-to-event dataset **Overall Survival in days (OS)** into Subset 1.

   Don't drag the folder into the subset box – just the dataset included in the circle above.

5. In the Set Value dialog, select **No Value**, then click **OK**:



   Specifying a value limits the values in the dataset. By not specifying a value, the entire time-to-event dataset is used.

6. Repeat step 4 and step 5 for Subset 2.

7. In the same **Clinical Data** node, drag the **Censored (OSCENS)** folder into empty boxes in Subset 1 and Subset 2.

   The contents of the **Censored (OSCENS)** folder are **No** and **Yes**. This concept introduces the **Event** and **Censored** datasets into the analysis.

   Now you will introduce the variable whose effect on survivability you want to test.

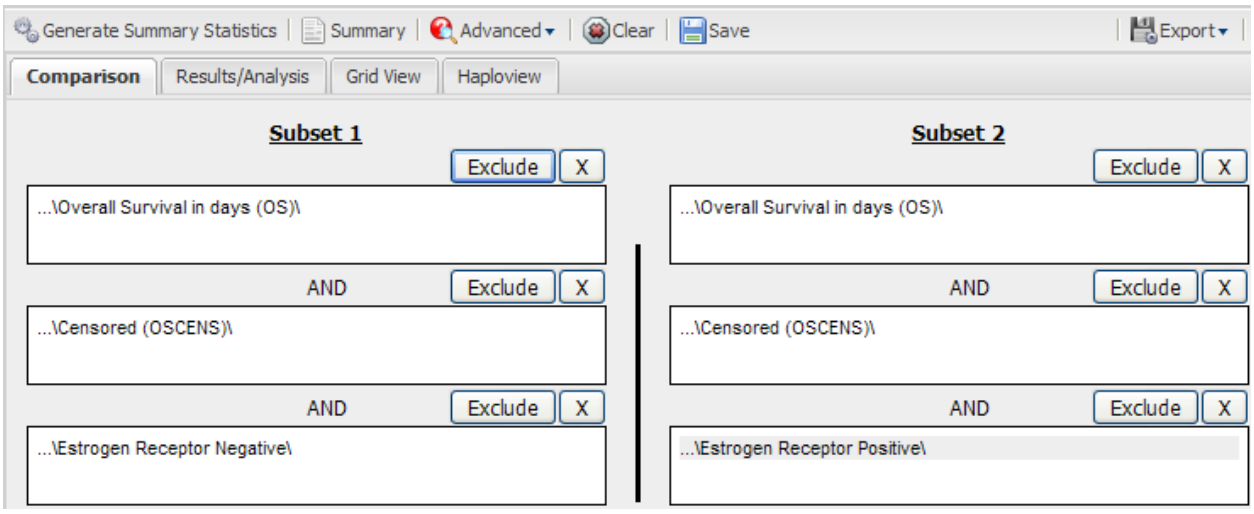
8. Open the following nested nodes in the following order:

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

9. Drag **Estrogen Receptor Negative** into an empty box in Subset 1.

10. Drag **Estrogen Receptor Positive** into an empty box in Subset 2.

    The subset boxes are now defined as follows:



11. Click the **Advanced** tab, then click **Survival Analysis**.

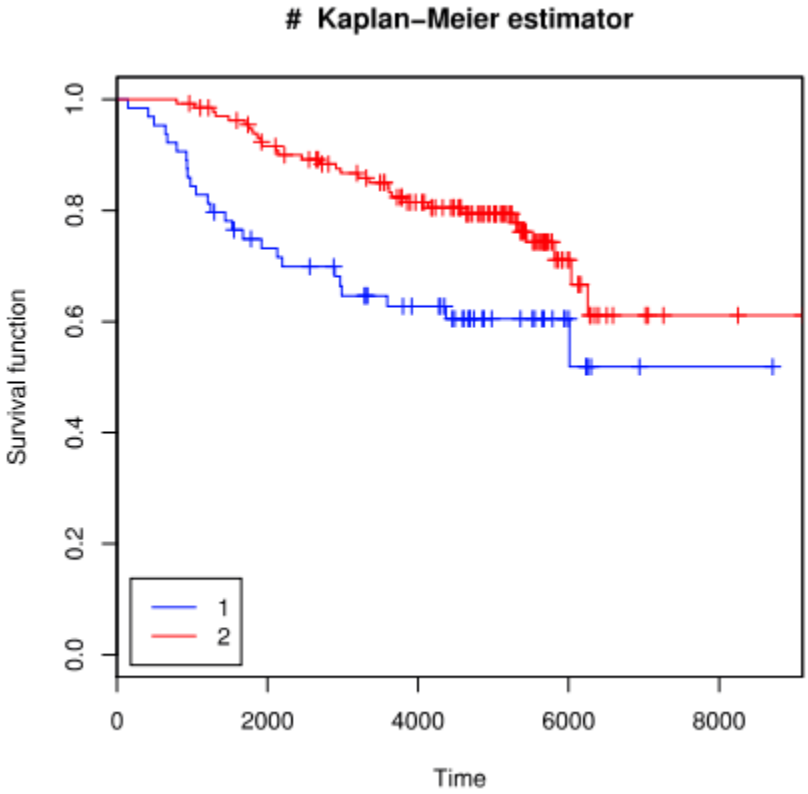
    In a few seconds, the Survival Analysis window appears. It has four sections:

    - At the top, a summary of the subset definitions.
    - Next are two tables that contain the survival analysis statistics.

      In the first table, the Hazard Ratio and Relative Risk statistics compare Subset 2 results against Subset 1 results. In this example, Subset 2 patients (those with positive estrogen receptors) have a lower hazard ratio and lower relative risk than those in Subset 1 (those with negative estrogen receptors).

Number of Subjects	198
Hazard Ratio (95% CI)	0.476 (0.281 - 0.806)
Relative Risk (p Value)	-0.743 (0.0058)

    - The final section of the Survival Analysis window contains Kaplan-Meier curves of the survival times of each group:



      In the figure, the x-axis represents survival time in days, and the y-axis represents the percentage of subjects who were still alive at a given point in time during the study.

      Note the hashmarks in the plot lines. These represent censored data – for example, subjects who dropped out of the study before the event (death) occurred.

12. When finished viewing the analysis, close the Survival Analysis window.

Lesson 7

Perform a Principal Component Analysis

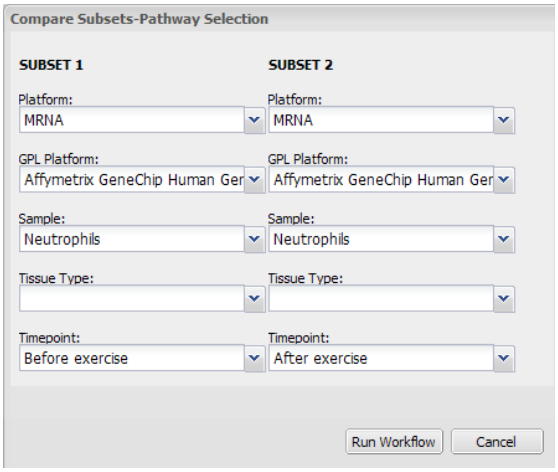
Lesson Goal: Generate a principal component analysis (PCA) visualization and statistics.

Scenario: You are interested in a study on the effect of strenuous exercise on neutrophils in the 12 healthy male subjects. You want to see if exercise causes significant change on the gene expression profiles of these 12 subjects, before and after exercise.

It is expected that for one subject, one set of genes will show some changes, and another set of genes will show changes for another subject. Direct comparisons cannot answer the question if the group's gene profiles change after exercise. PCA analyzes the multiple dimensional data and finds several linear combinations of the original dimensions to represent the most variance in the data. Each linear combination of the original dimension is named a "Principal Component." PCA can be used to provide multiple snapshots of the data, and show if the two groups of data have different distributions on the graph.

1. Click the Dataset Explorer **Comparison** tab, then click the **Clear** button to remove the subset definitions from the previous lesson.
2. In **Public Studies**, open the study **Radom-Azik_Exercise_GSE8668**.
3. Open the following nested nodes in the following order:
   a. Biomarker Data
   b. Affymetrix GeneChip Human Genome U133 Plus 20 Array
   c. Neutrophils
4. Drag **Before exercise** into an empty box in Subset 1.
5. Drag **After exercise** into an empty box in Subset 2.
6. Click the **Advanced** tab, then click **Principal Component Analysis**.

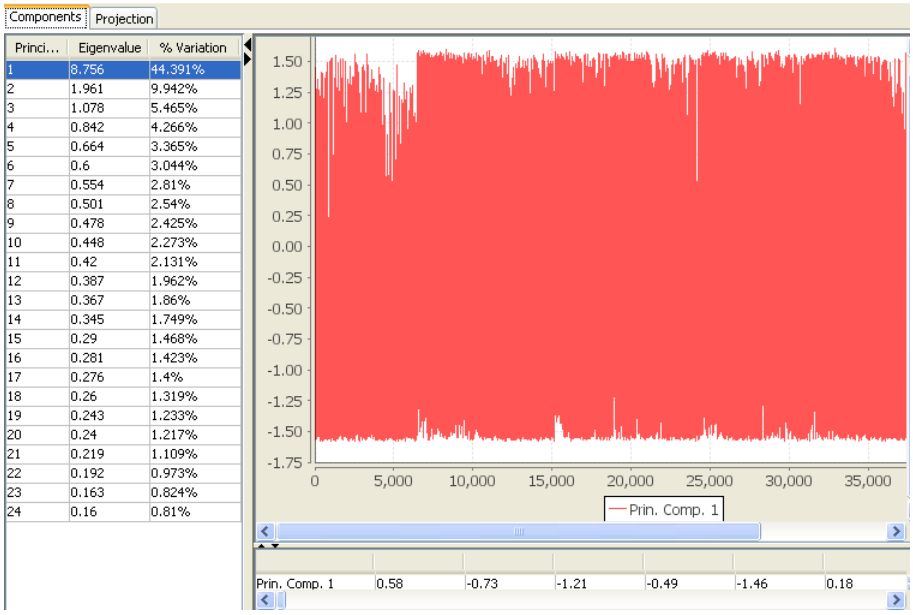
   The Compare Subsets-Pathway Selection dialog appears, with its fields already filled out with the default values for the analysis:



7. Click **Run Workflow**.

   Due to the large amount of gene expression data being processed, it may take a minute or two for the PCA Viewer window to appear. When the window does appear, you see the **Components** tab by default. Take a minute to familiarize yourself with its contents:

   - The left pane contains a table listing the principal components. The components are listed in order of variability of data along this vector (the combination of original dimensions), from the greatest variability of these measurements within a component to the least.
   - The right pane contains a visualization of the eigenvectors for each principal component selected in the table on the left.
   - Below the visualization is a table of the gene expression measurements represented in the visualization for each selected component.

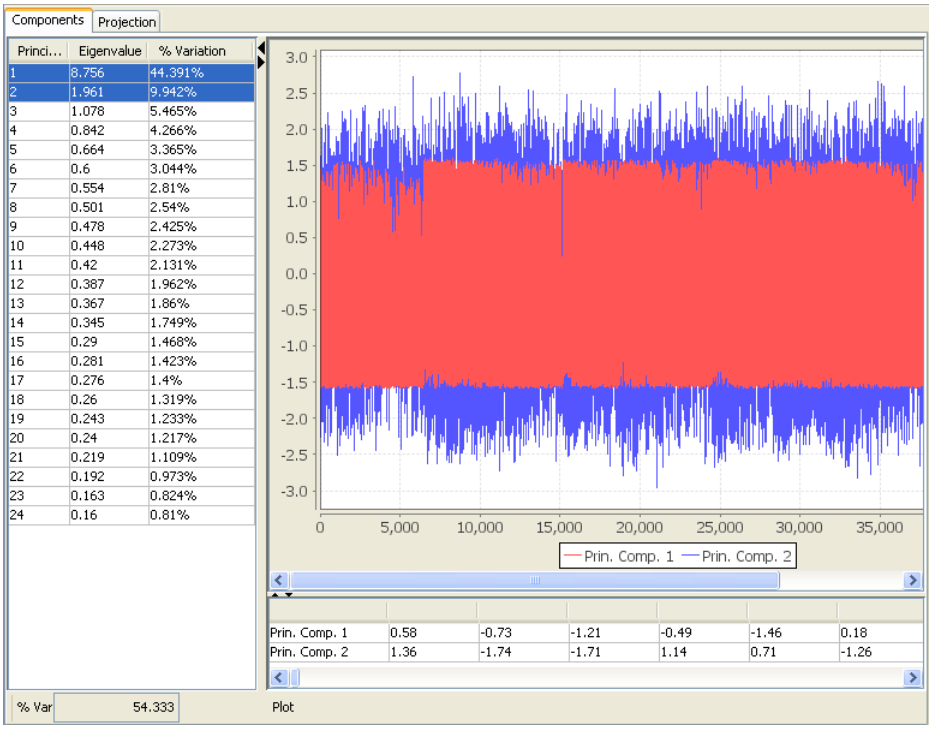


   By default, the visualization shows the values for the first principal component only. You want to view a visualization of the first and second principal components.

8. While holding down the **Ctrl** key, click the second principal component in the table on the left.

9. Click the **Plot** button at the bottom of the window.

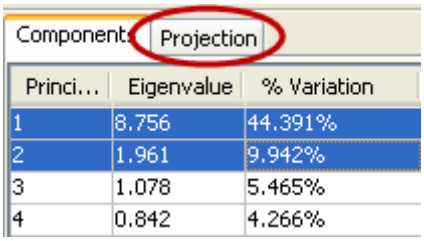
   The visualization presents the selected components in different colors, and the table below the visualization now contains data from both components:



   You can generate a visualization of any combination of principal components that you select in the table on the left.

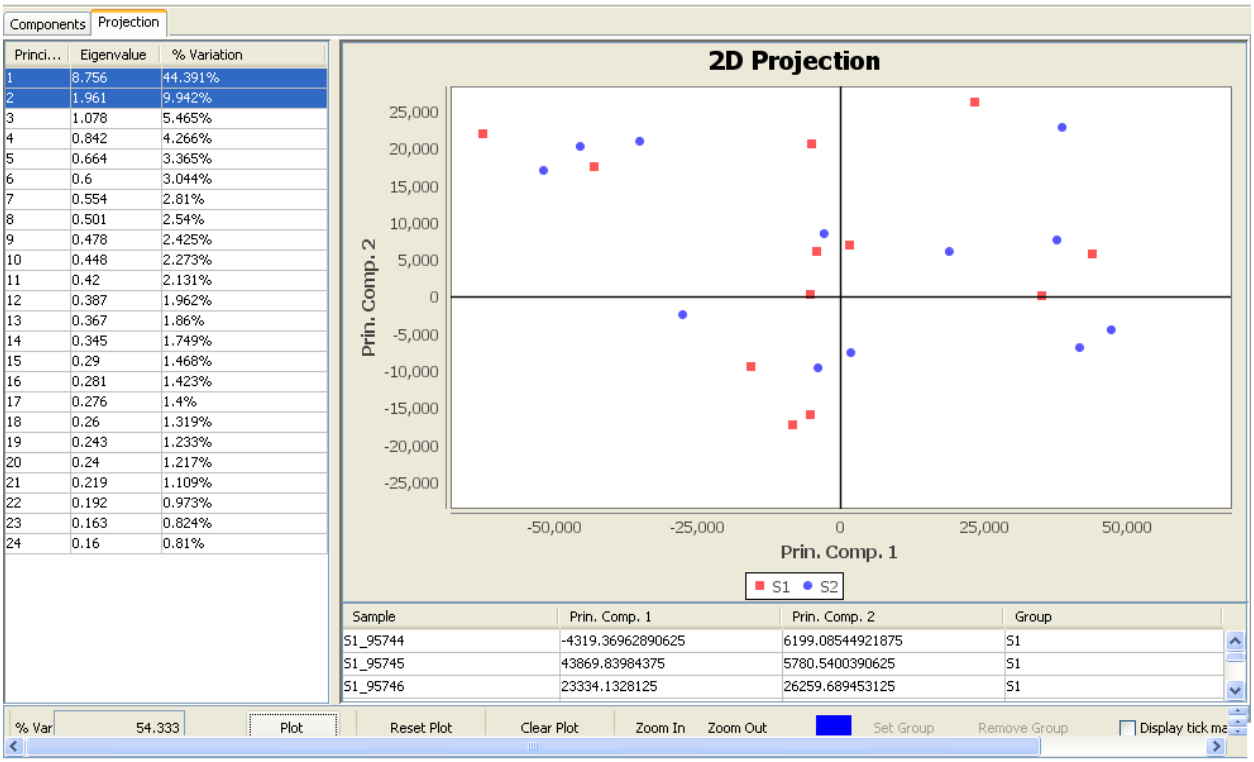
   Now you want to see the distribution of the 12 subjects' data, before and after exercise, on these new dimensions. If the before and after data groups occupy completely different sectors in the 2D or 3D view, it indicates that these two groups are significantly different.

10. Click the **Projection** tab:



11. Select the first two principal components, then click the **Plot** button at the bottom of the PCA Viewer window.

    A two-dimensional projection like the following appears:



> **Note:** You can also generate a three-dimensional projection by selecting three principal components. However, three-dimensional projections require that you have Java 3D installed.

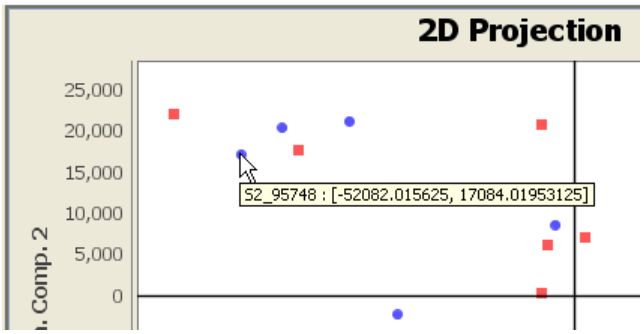
In examining the two-dimensional projection and data, you observe that:

- The two new dimensions of this 2D view represent 44.4% and 9.9% variability of the data. If there is significant change caused by exercise, there should be separation of two groups of subject data (blue and red data points) in this view.
- Even distribution of red data points in this view indicates no outlier among the 12 gene expression profiles before the exercise. Similarly, blue data points indicate no outlier for those after the exercise.
- The evenly mixed red and blue data points indicate there is no significant change in gene expression profiles for these 12 patients before and after exercise.

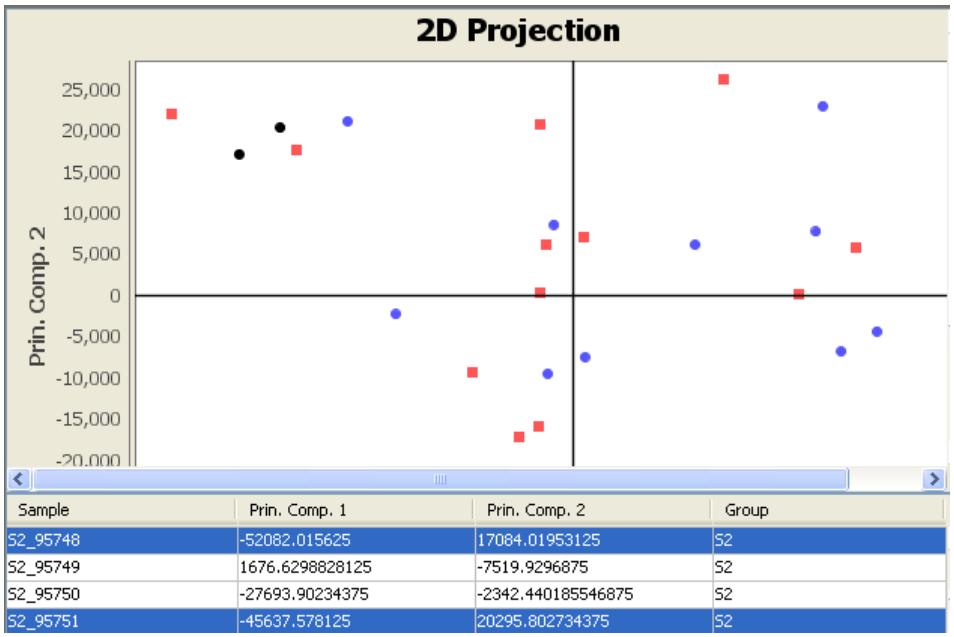
The observations above are based on the capability and limitation of PCA method. Other analysis methods may yield different observations and conclusions.

12. Familiarize yourself with the interactive features of the visualization – for example:

    - Hover the mouse pointer over a data point to display its data:



    - Click on one or more data points (selected data points become black) to highlight the corresponding sample's data in the table below the visualization:



Sample	Prin. Comp. 1	Prin. Comp. 2	Group
S2_95748	-52082.015625	17084.01953125	S2
S2_95749	1676.6298828125	-7519.9296875	S2
S2_95750	-27693.90234375	-2342.440185546875	S2
S2_95751	-45637.578125	20295.802734375	S2

    - Conversely, select one or more samples in the table to locate the corresponding data points in the two-dimensional projection (the data points become black).

13. When finished viewing the projection, close the PCA Viewer window.

Lesson 8

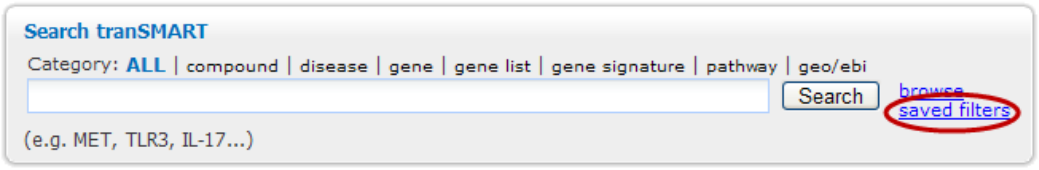
Submit tranSMART Search Results to Pictor

Lesson Goals: Run a saved search, and submit the genes in the search results to Pictor for further analysis.

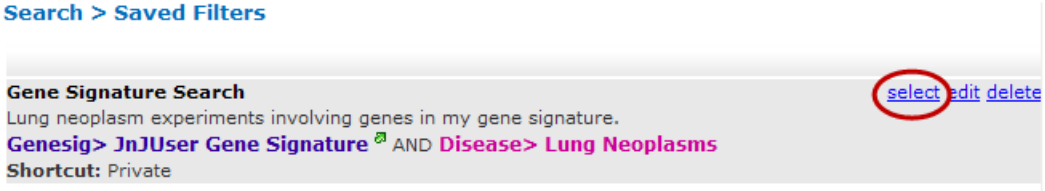
> **Note:** To complete this lesson, you must have access to Pictor, the Johnson & Johnson web site that searches pathway databases such as KEGG, GeneGO and Ingenuity.

Scenario: In previous lessons, you created a gene signature based on the most highly regulated gene expressions in an experiment involving lung adenocarcinoma, and then you used the gene signature in a search related to lung neoplasms. You now want to submit the results of that search to Pictor, to find pathways that correspond to the genes in your gene signature.

1. Click the tranSMART **Search** tab to display the Search window.
2. Click **saved filters** in the Search window:

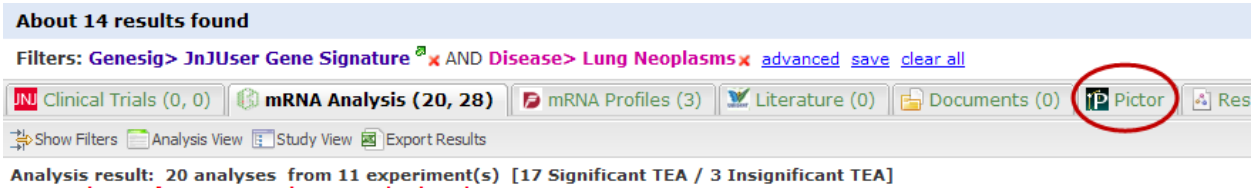


3. In the **Saved Filters** list, click **select** for the saved filter **Gene Signature Search**:

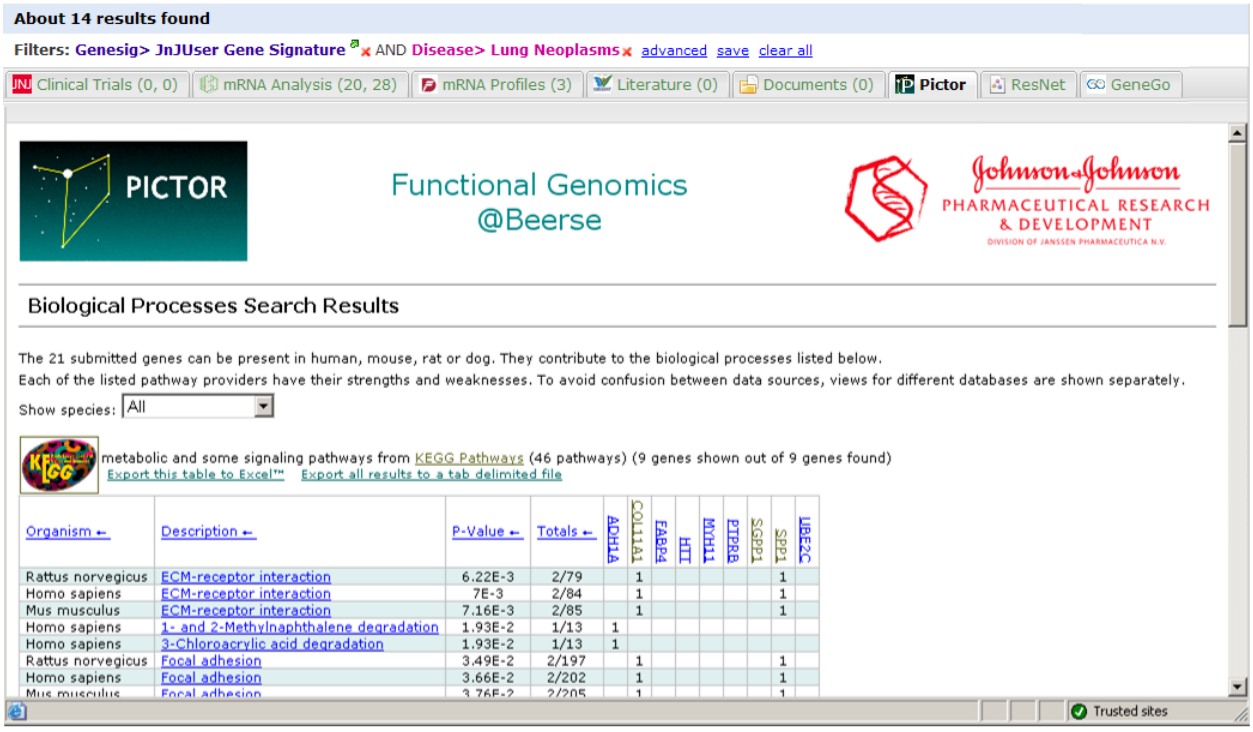


The search begins immediately, and a search result is returned.

4. Click the **Pictor** tab:



The Pictor pathways that most correspond to the genes in your gene signature are displayed in the tranSMART search results area:



5. Select **Homo sapiens** in the **Show species** field:

Scroll down to view the databases that returned results to Pictor – KEGG, GeneGO, and Ingenuity.

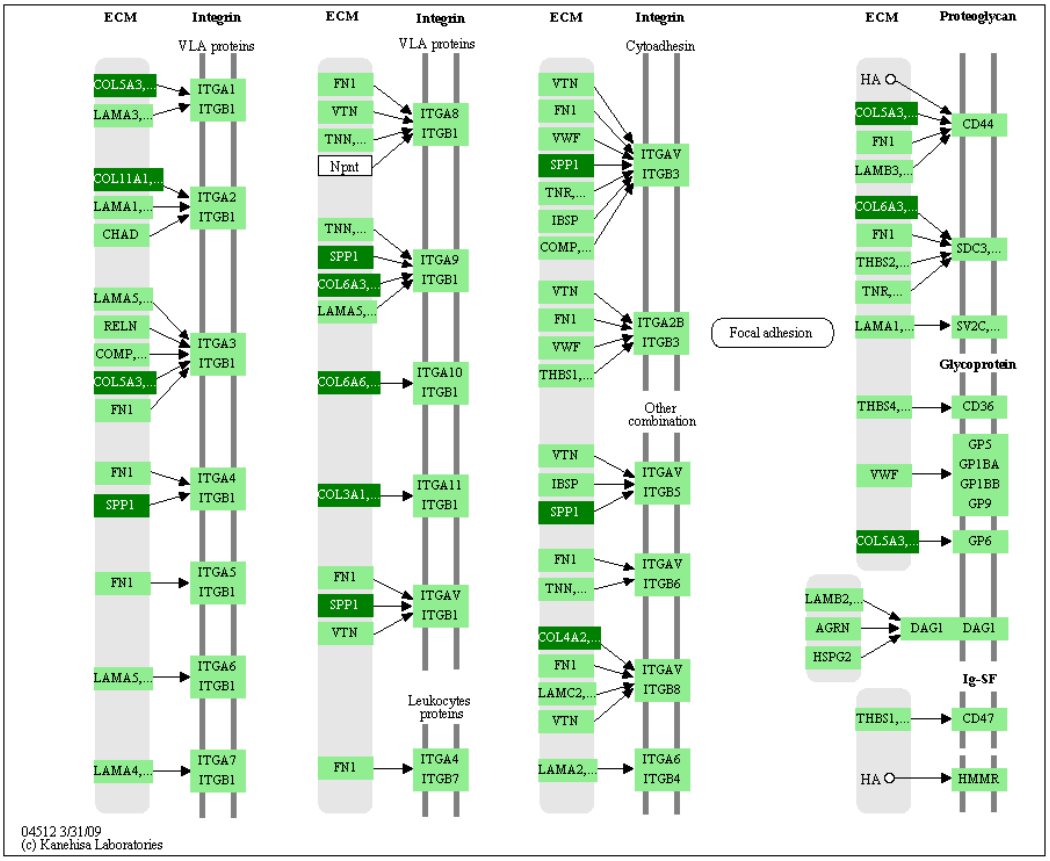
Next you will scan the lists of matching pathways from these databases.

6. In the list of matching KEGG pathways, click **ECM-receptor interaction**.

Two browser windows open – a KEGG visualization of ECM receptor interactions for homo sapiens, and a KEGG map listing the associated genes.

In the KEGG map, notice that the genes that match those in your pathway – COL11A1 and SPP1 – are highlighted and marked with an asterisk.

The following figure shows the KEGG visualization:



7. Explore the interactions of the genes in the visualization, clicking on any areas of interest.

8. When finished, close all browser windows except for the tranSMART window containing Pictor.

9. In the next set of pathway data, GeneGO curated human metabolic and signaling pathways, click **Hedgehog and PTH signaling pathways participation in bone and cartilage development**:

curated human metabolic and signaling pathways from MetaCore™ (12 pathways) (6 genes shown out of 6 genes found)
Export this table to Excel™ Export all results to a tab delimited file

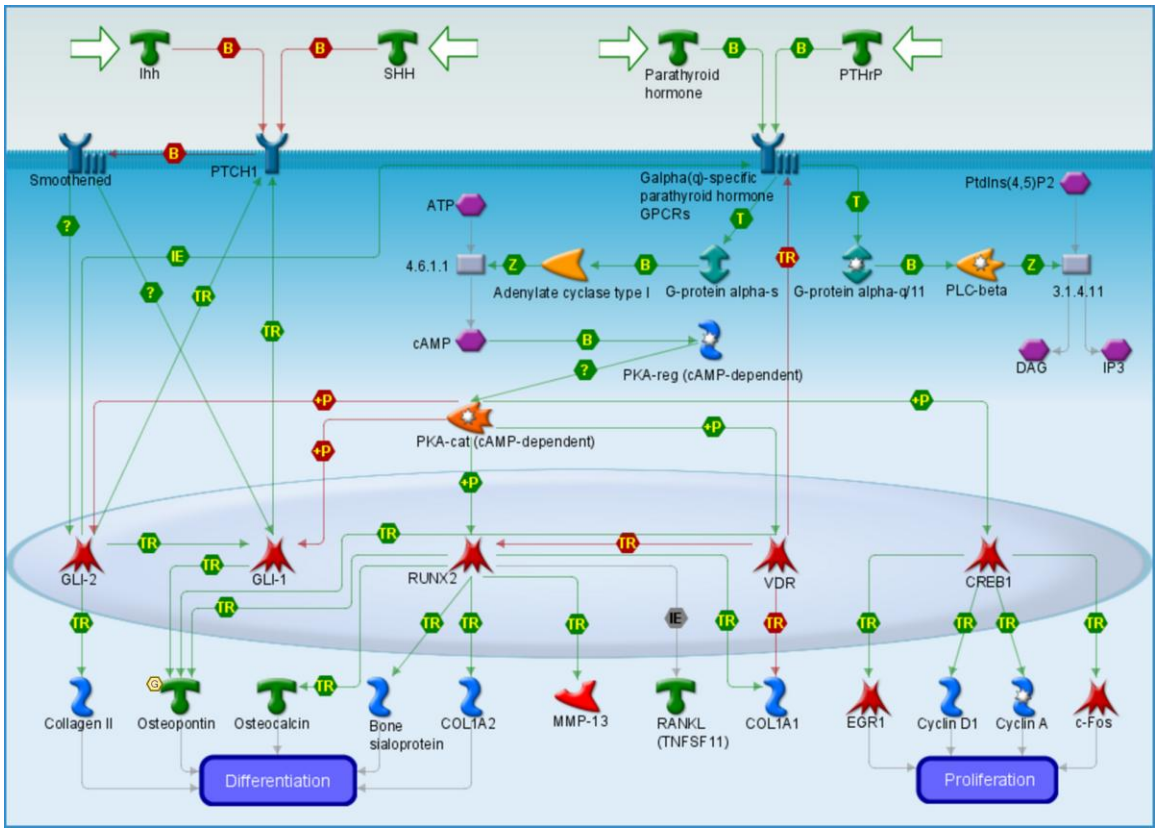
Organism	Description	P-Value	Totals	ADH1A	CXXC1	FABP4	MYH11	SGPP1	SPP1
Homo sapiens	Hedgehog and PTH signaling pathways participation in bone and cartilage development	1.84E-3	2/45		1				1
Homo sapiens	Role of VDR in regulation of genes involved in osteoporosis	3.47E-3	2/62		1				1
Homo sapiens	Pyruvate metabolism	3.44E-2	1/23	1					
Homo sapiens	PPAR regulation of lipid metabolism	4.31E-2	1/29			1			
Homo sapiens	Bile Acid Biosynthesis	4.46E-2	1/30	1					
Homo sapiens	Sphingolipid metabolism	5.48E-2	1/37					1	
Homo sapiens	Regulation of actin cytoskeleton by Rho GTPases	7.2E-2	1/49				1		
Homo sapiens	MAG-dependent inhibition of neurite outgrowth	7.34E-2	1/50				1		
Homo sapiens	Leukotriene 4 biosynthesis and metabolism	7.91E-2	1/54	1					
Homo sapiens	Integrin-mediated cell adhesion	8.33E-2	1/57				1		

more...
Genes not present in any GeneGO pathways: AGER C10orf116 COL11A1 EEF1A1 EEF1A2 HTT LPHN2 MFAP4 PTPRB SLC6A4 SOSTDC1 SPINK1 TOX3 UBE2C

Two browser windows open – a Hydra window containing a visualization of the pathways, and a GeneGO window listing the associated genes.

If prompted to log into Hydra, use **demo** for both the user name and password.

The following figure shows the Hydra visualization:

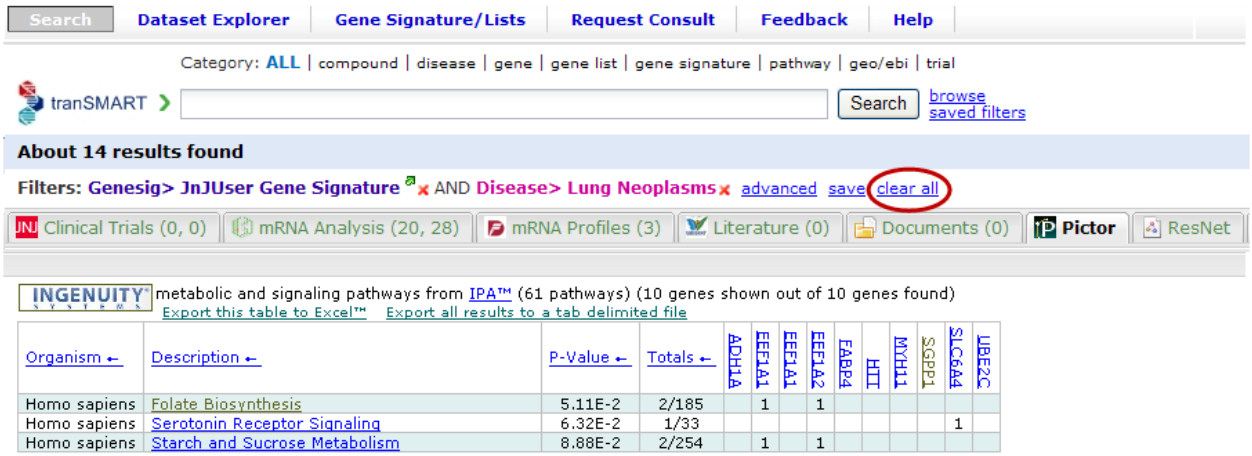


10. Explore the visualization of the Hedgehog and PTH signaling pathways in Hydra, clicking on any areas of interest.

11. When finished, close all browser windows except for the tranSMART window containing Pictor.

    The final two lists of matching pathways in Pictor are the GeneGO curated human diseases list and the Ingenuity list. After browsing these list, you decide there is no new information there that interests you.

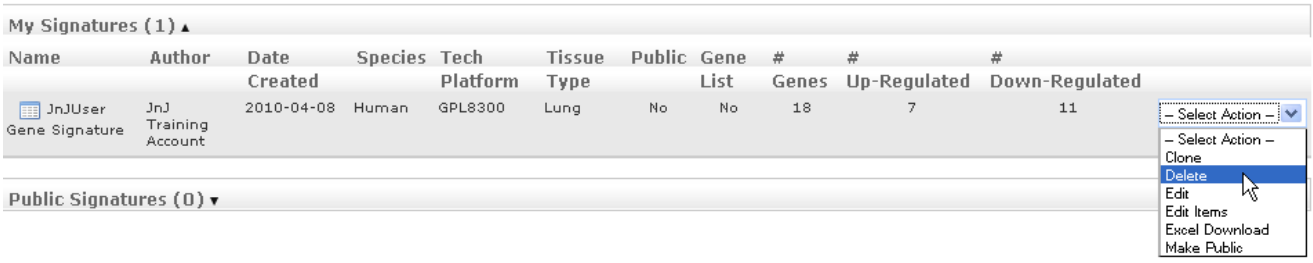
12. In tranSMART, click **clear all** to remove Pictor from the tranSMART window:



You now decide to delete your gene signature.

13. Click the tranSMART **Gene Signature/Lists** tab.

14. In the **Select Action** dropdown for the gene signature you created, click **Delete**:



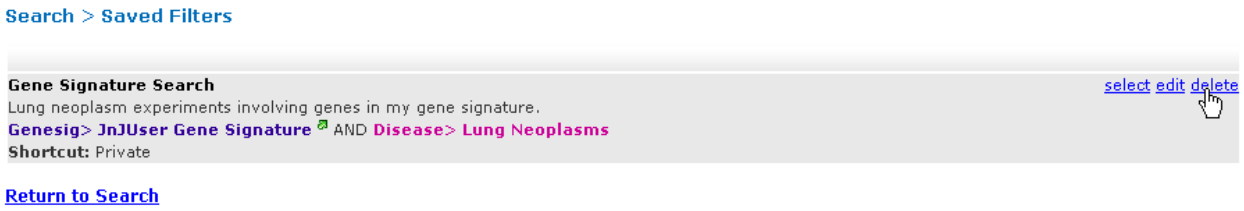
15. Click **OK** to confirm the deletion.

Finally, you decide to delete the saved search filter that included your gene signature.

16. Click the tranSMART **Search** tab.

17. Click **saved filters**.

18. Click **delete** for the saved filter **Gene Signature Search**:



19. Click **OK** to confirm the deletion.

20. Click **Return to Search** to prepare for the next lesson.

Lesson 9

View a Gene's Relationships to Other Entities in a Pathway

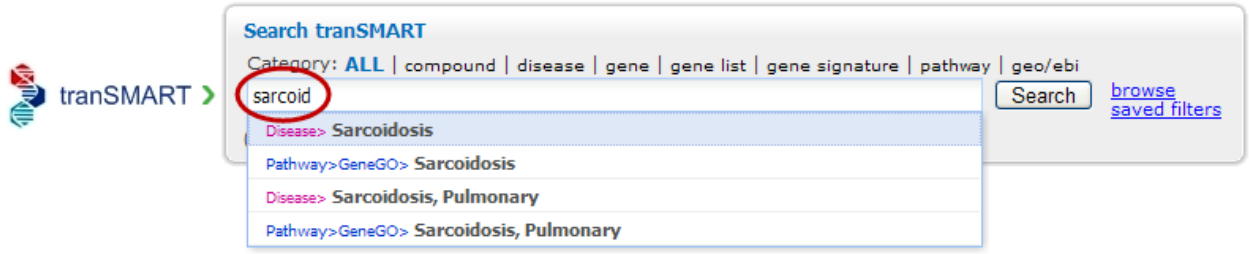
Lesson Goals: (1) Export analysis data from a gene expression experiment from tranSMART into Ariadne Genomics Pathway Studio. (2) Find a pathway that has been enriched with a gene of interest, and display a visualization of the pathway.

> **Note:** To perform this lesson, you must have Pathway Studio installed, and you must have downloaded the ResNet Mammalian Database.
>
> **Note:** This lesson uses the demo version of the Mammalian database. The demo database contains a subset of the data in the Mammalian database, and should not be used for actual research.

Scenario: In a search for mRNA experiments involving the pathway for pulmonary sarcoidosis, you notice that the gene DDR1 was the only gene that was up-regulated in one particular experiment. You are curious about other pathways that are enriched with this gene.

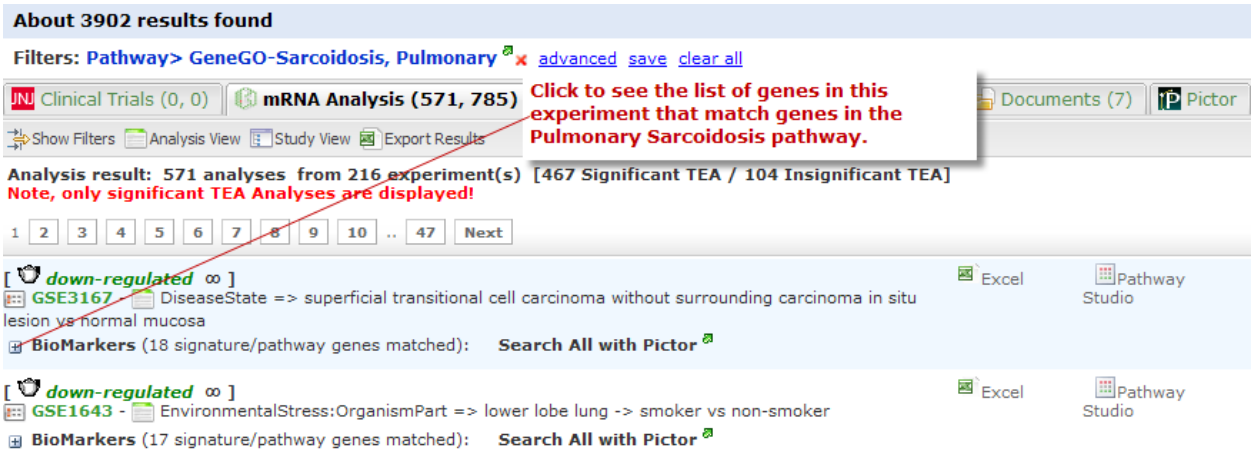
Import Analysis Data into Pathway Studio

1. Type **sarcoid** in the tranSMART Search field:



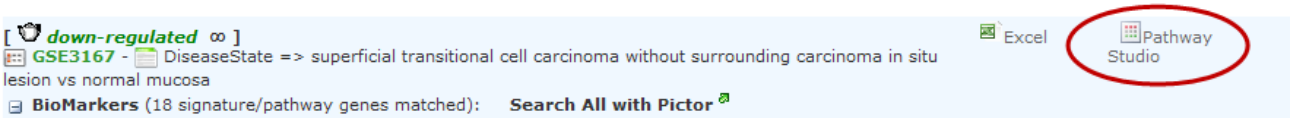
2. Click **Pathway>GeneGO> Sarcoidosis, Pulmonary** in the dropdown list.
3. In the **mRNA Analysis** search results, find the experiment **GSE3167**.

4. Click the **+** icon (⊞) to the left of the label **BioMarkers** to display a list of genes in the experiment that match genes in the pathway:



You notice that the gene DDR1 is the only matching gene that was up-regulated in the experiment.

5. Click the **Pathway Studio** button to the right of the experiment name:



6. When the File Download dialog appears, click **Open**.

   When you click **Open**, Pathway Studio starts up and launches the Experiment Import Wizard, allowing you to specify the experiment data to import.

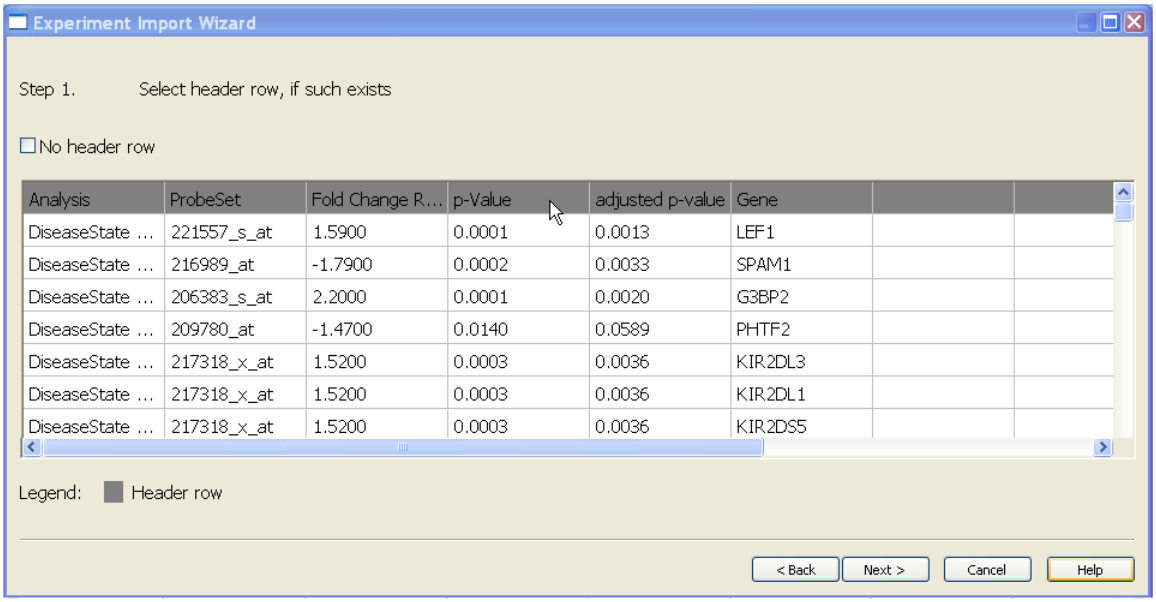
   Note: After Pathway Studio starts up, it may take a few seconds for the Experiment Import Wizard to appear (also, the default Pathway Studio window appears before the wizard is displayed). When the wizard appears, you may need to maximize the window to see the entire contents.

7. Select **Gene Expression** in the **Experiment Type** field.
8. Select **Microsoft Excel** in the list of **File format** choices.
9. Click the **Browse** button to the right of the **Destination folder** field.
10. In the Select Folder dialog, click **Experiments**, then click **OK**.
11. Click **Next** to proceed to the next wizard page.

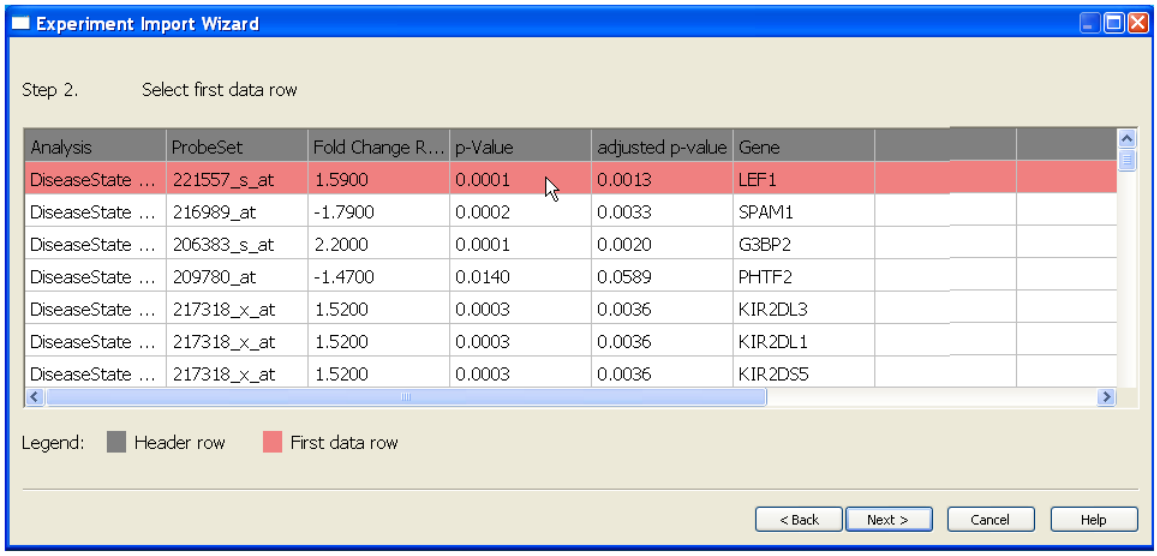
    This page displays the analysis data exported from tranSMART. The wizard will now guide you through several steps in which you will select the data to import, and define the format in which the imported data will be displayed.

    Whenever possible, Pathway Studio will select the data for you.

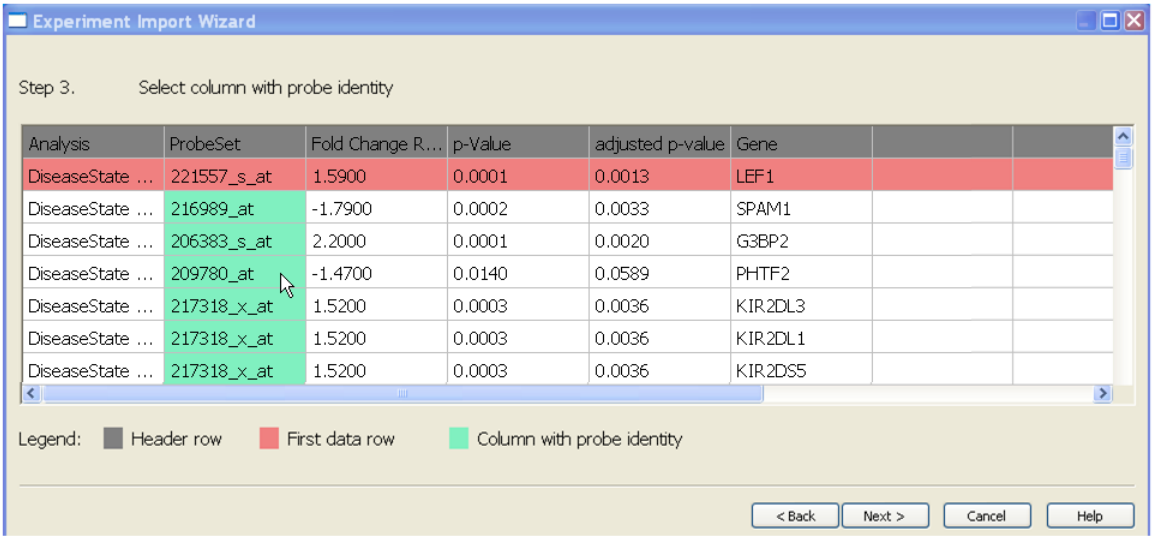
12. In **Step 1**, click the header row at the top of the data rows to select it, then click **Next**.



13. In **Step 2**, click the first data row to select it, then click **Next**.



14. **Step 3**, click the **ProbeSet** column to select it, then click **Next**.



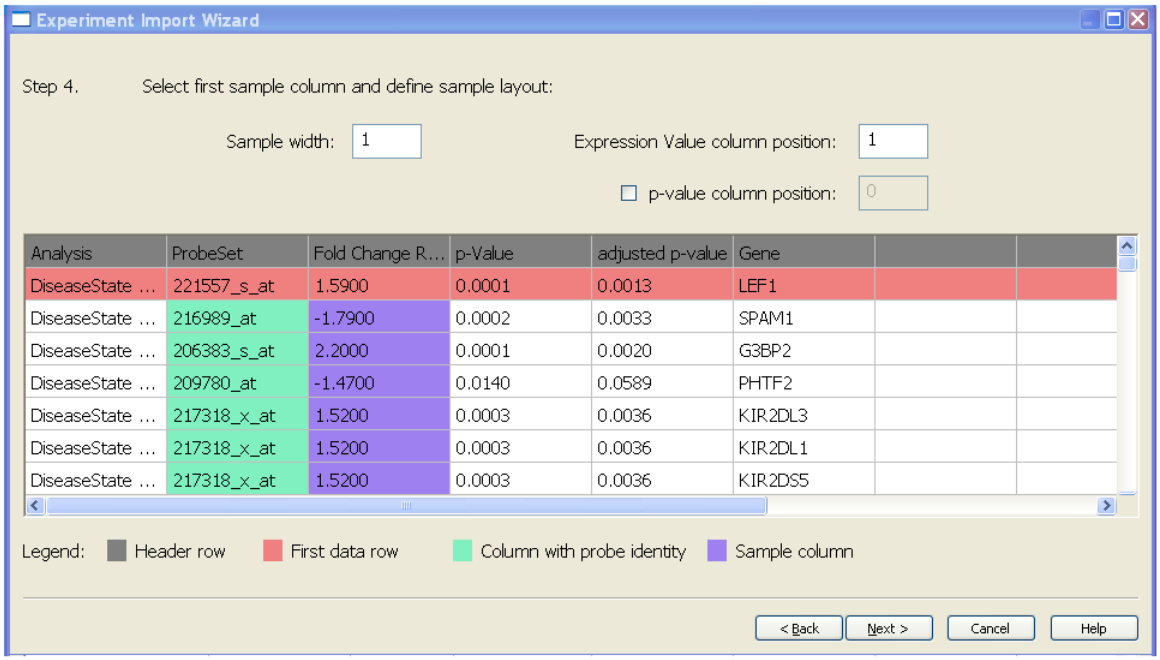
In **Step 4**, you will select the sample data to import with the probe set names. You will also inform Pathway Studio of the location of the columns of sample data to import.

15. Click the **Fold Change Ratio** column to select it.
16. Accept the default value **1** in the **Sample width** field.
17. Accept the default value **1** in the **Expression Value column position** field.

    This value specifies that Pathway Studio can expect to find the gene expression data in the first column after the probe set names.

18. Make sure that the **p-value column position** check box is cleared.

Step 4 of the wizard should now appear as shown below:

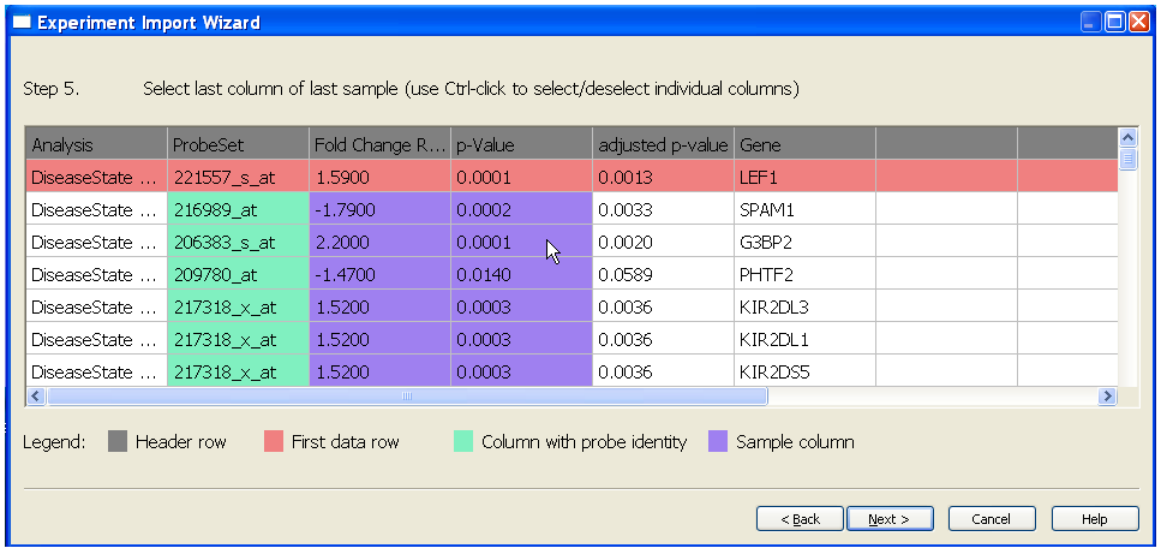


19. Click **Next** to proceed to **Step 5**.

    In **Step 5**, notice that all the remaining columns are now selected for inclusion as sample data. However, you only want to add one more column of sample data – the **p-Value** column.

20. Press and hold down the **Shift** key, then click the **p-Value** column to select it.

    This action selects the p-value data as the second column of sample data, and de-selects all the columns to the right of the **p-Value** column.



21. Click **Next** to proceed to **Step 6**.

    This step allows you to import additional data into Pathway Studio, to serve as annotations of the sample data. Pathway Studio does not use annotation data in the operations it performs on the sample data.

22. Click **Next** to decline to import annotation data and to proceed to **Step 7**.

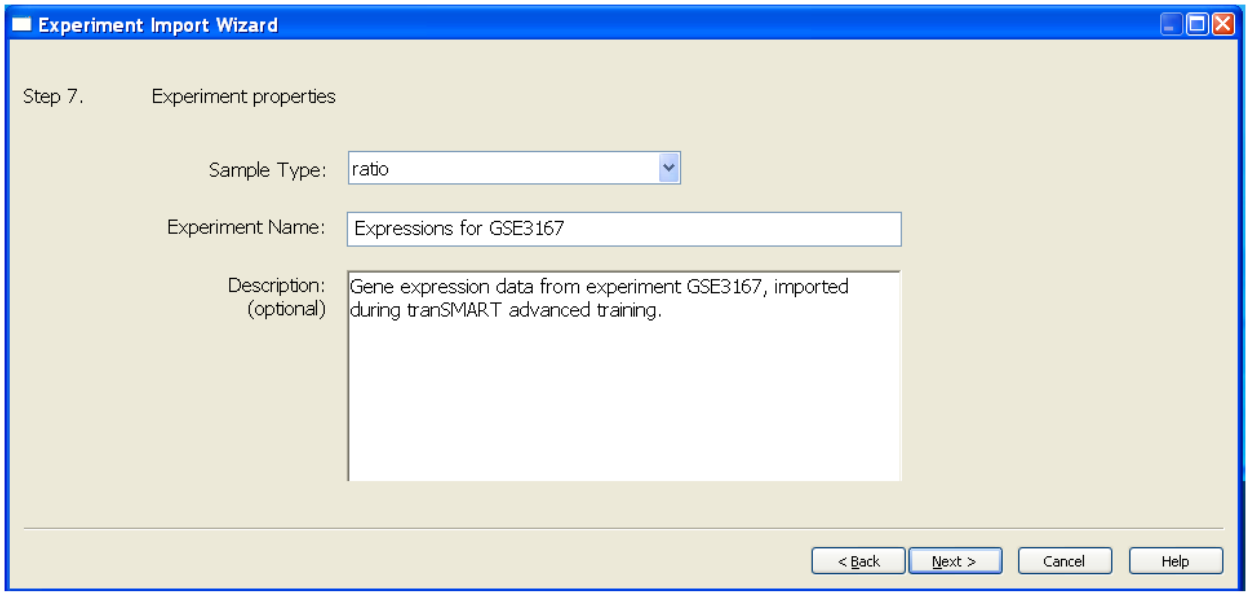
23. Provide the following information in **Step 7**:

    - In the **Sample Type** field, select the data type **ratio**.
    - In the **Experiment Name** field, type the following name for the data you are importing:

      `Expressions for GSE3167`

    - In the **Description** field, type the following description of the data you are importing:

      `Gene expression data from experiment GSE3167, imported during tranSMART advanced training.`



24. Click **Next** to proceed to **Step 8**.

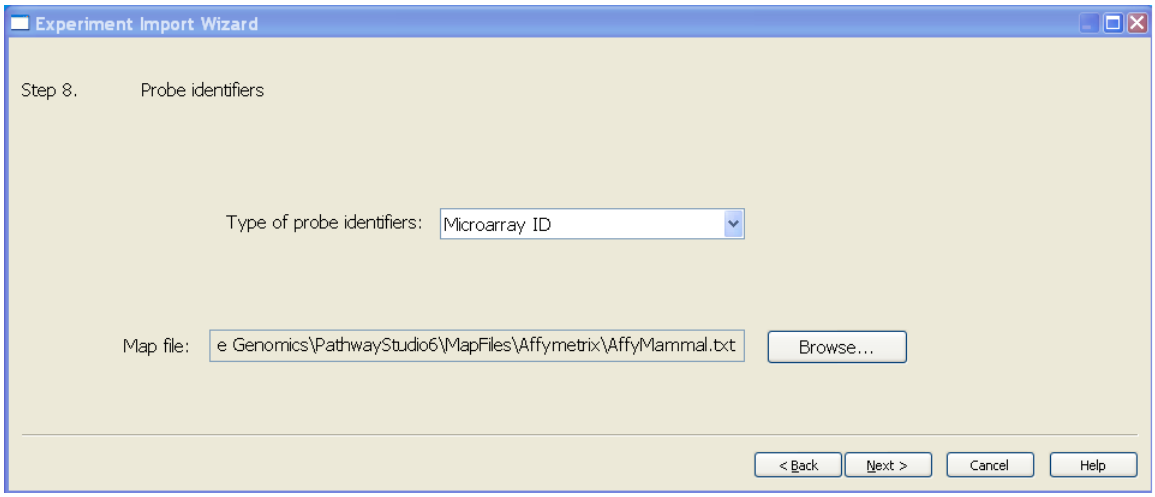
25. Provide the following information in **Step 8**:

    - In the **Type of probe identifiers** field, select **Microarray ID**.
    - In the **Map File** field, specify the path and name of the map file to use. Pathway Studio uses the map file to find the Entrez Gene IDs that correspond to the imported Microarray probe IDs.

      Optionally, click **Browse** to navigate to the map file.

      In this lesson use the map file **AffyMammal.txt** in the following default location:

      ```
      C:\Documents and Settings\All Users\Application Data\Ariadne
      Genomics\PathwayStudio6\MapFiles\Affymetrix\AffyMammal.txt
      ```

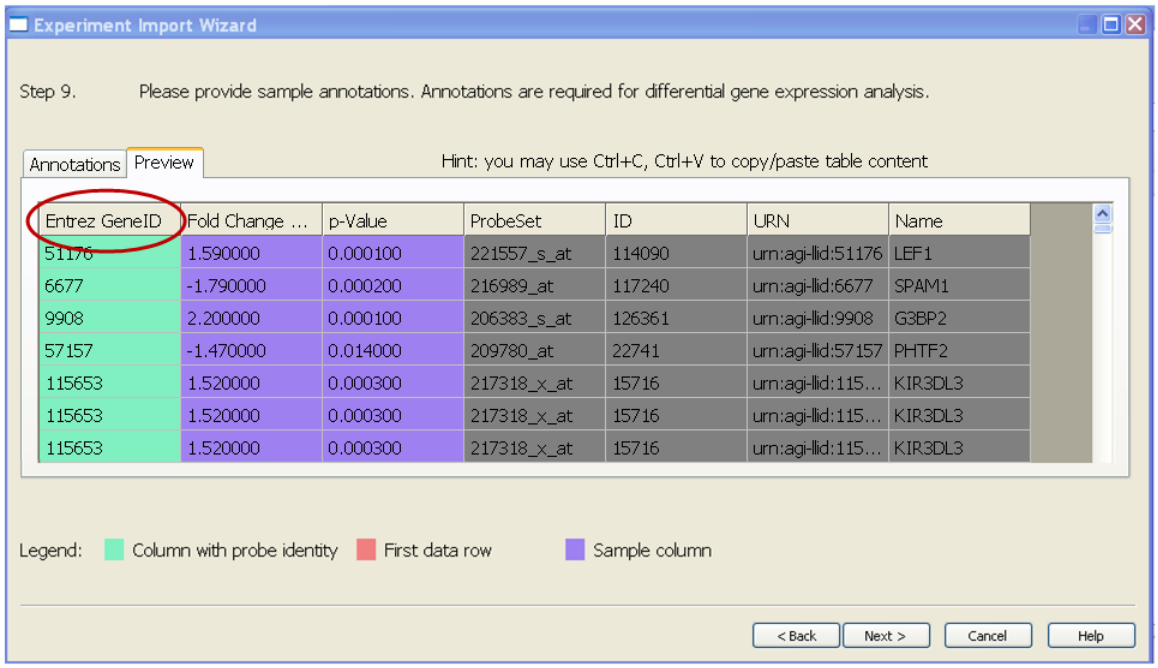


26. Click **Next**.

    Pathway Studio performs the mapping operation (which may take a minute or so), then proceeds to **Step 9**.

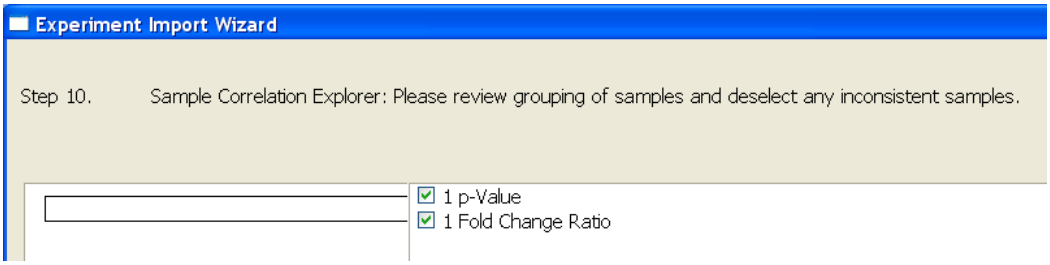
27. Click the **Preview** tab in **Step 9**.

    Notice that a new column containing the mapped Entrez Gene ID has been added:



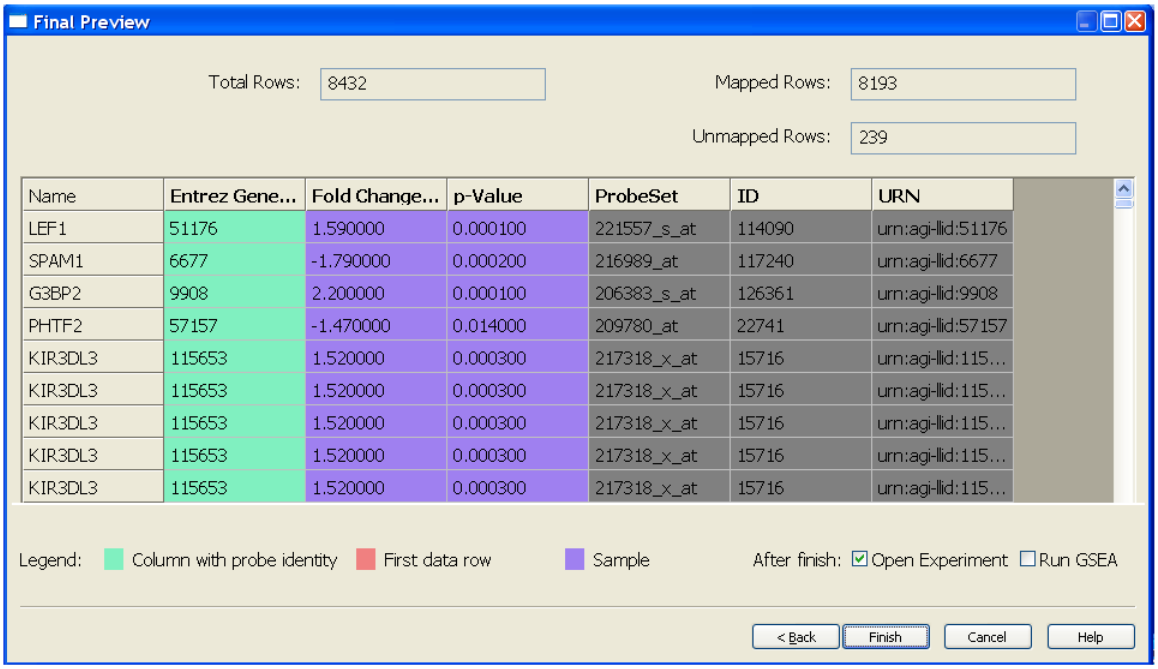
28. Click **Next** to proceed to **Step 10**.

29. Accept the defaults in **Step 10** as shown below, then click **Next**.



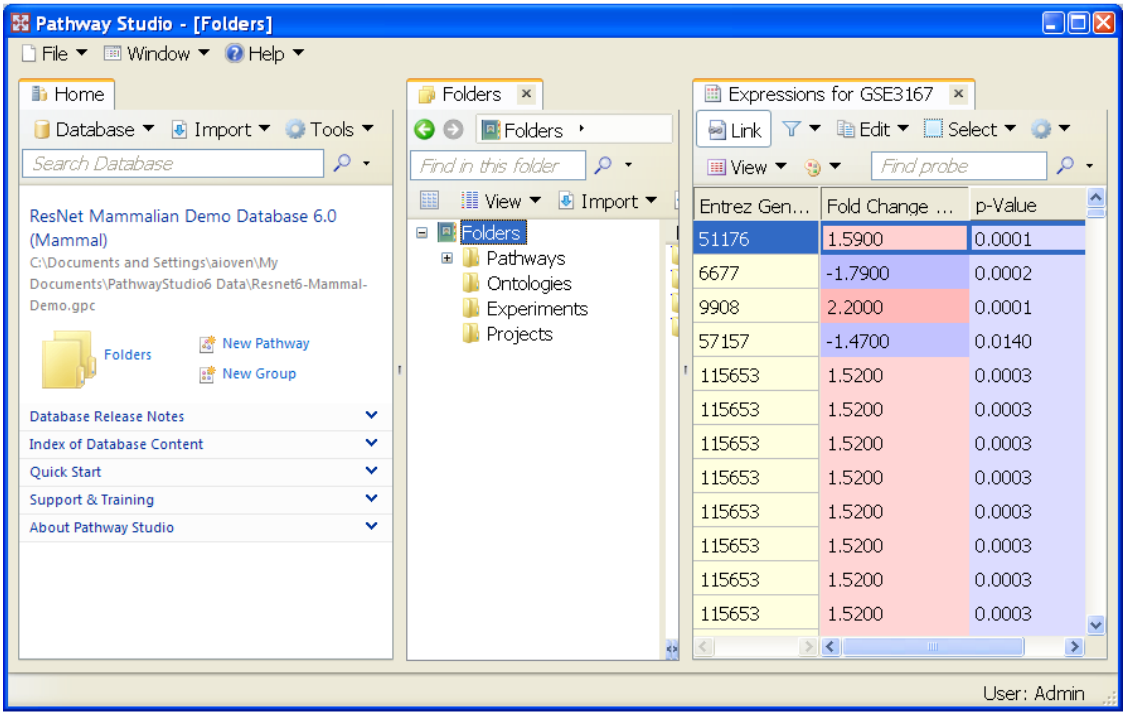
30. Click **Next** in the Find Differentially Expressed Genes dialog without making any changes.

The Final Preview dialog appears, as shown below:



31. Click **Finish** to import the data.

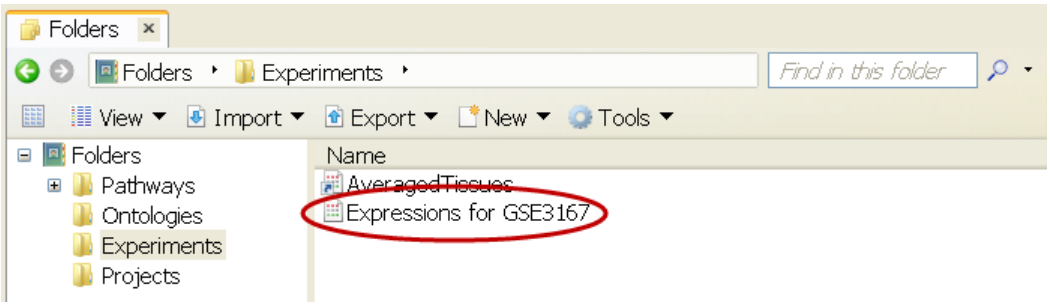
After a few seconds, the imported data appears in the right pane of the Pathway Studio as a heat map:



32. Maximize the Pathway Studio window.

33. Double-click either of the **Experiments** folders.

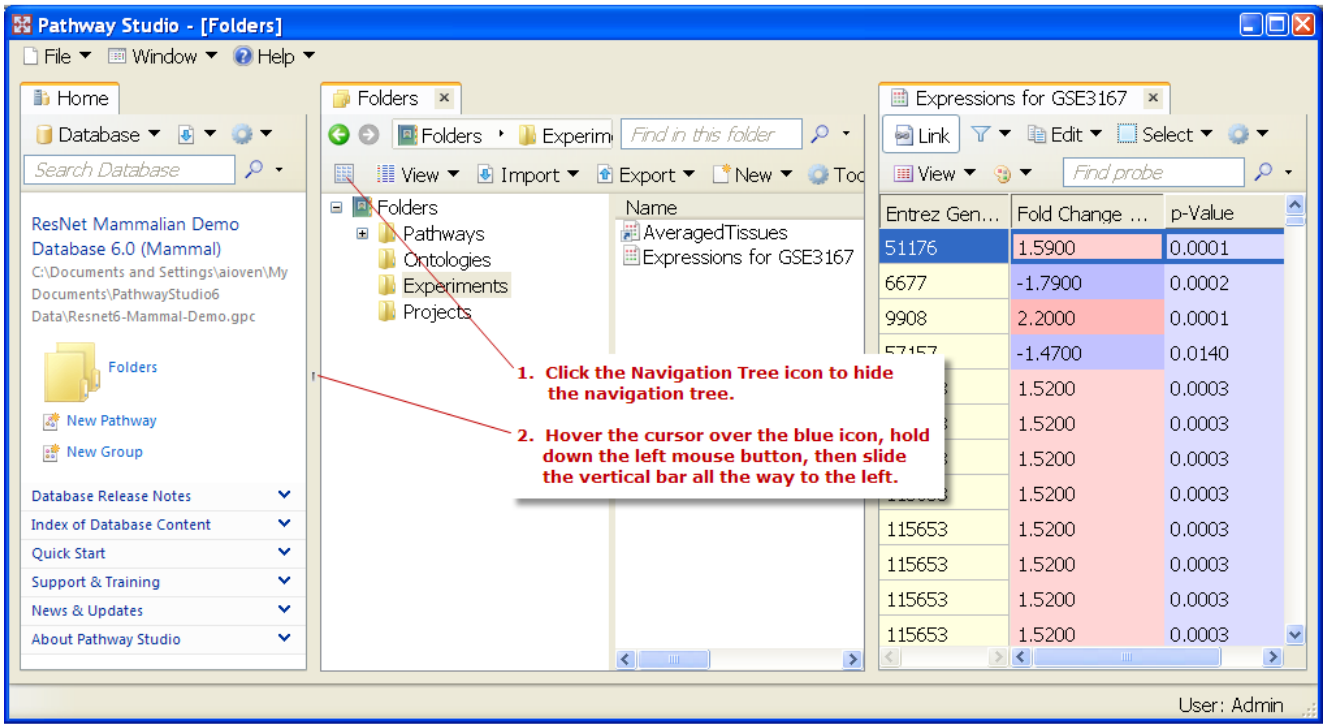
    Notice that the imported data, **Expressions for GSE3167**, has automatically been saved to a file and stored in the Experiments folder:



In the next lesson, you will search for all the pathways in the Pathway Studio database that have been enriched with a gene you select from the imported sample data.

Find and Display a Visualization of a Pathway

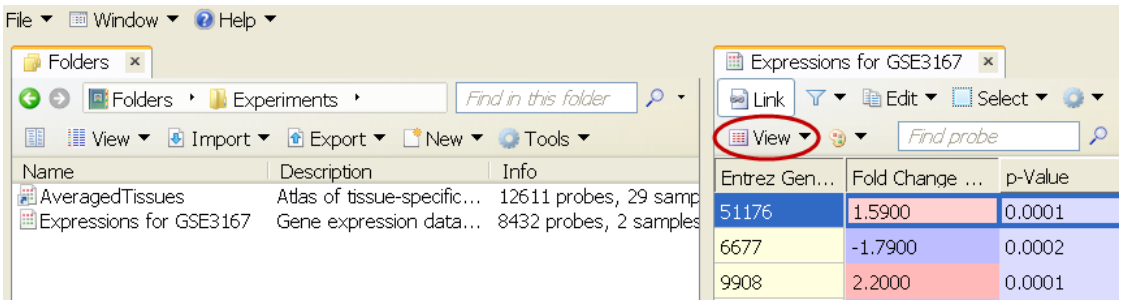
Note: Before you begin, you can create more room for the visualization by performing the two steps shown in the figure below:



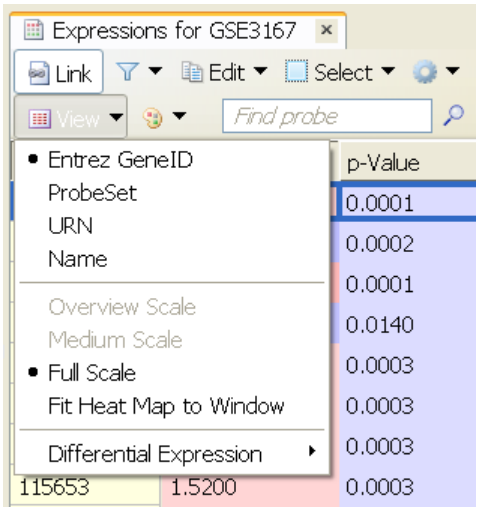
The gene DDR1 interests you because it was the only gene in the pathway for pulmonary sarcoidosis that was up-regulated in the experiment you imported from tranSMART. You want to find any pathways that are enriched with DDR1, and then display a visualization of one of the pathways to study the relationships of the genes and other entities in the pathway.

To find and display a visualization of a pathway enriched with DDR1:

1. In the heat map for the Expressions for GSE3167 sample, click **View**:

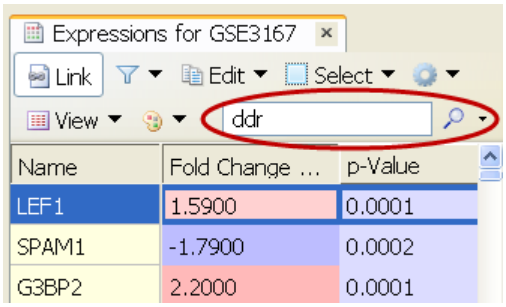


The View menu lets you select the type of identifier to use in the heat map – Entrez GeneID, the probe set name, the URN, or the gene name:

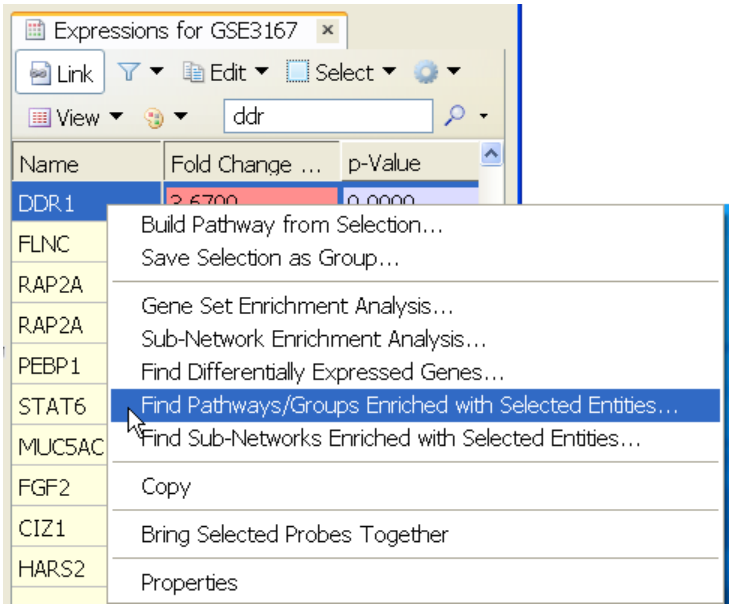


2. Click **Name** to use the gene name as the identifier.

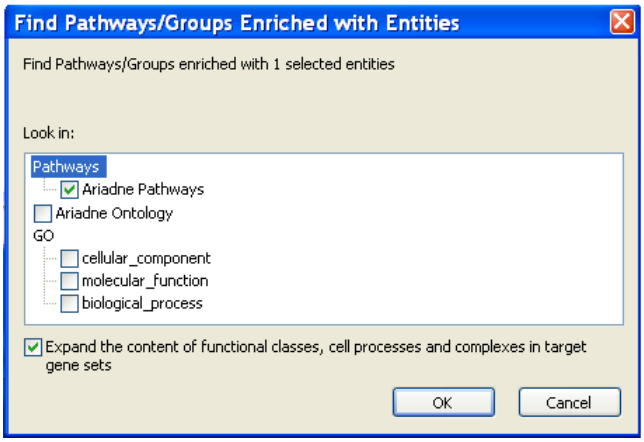
3. Type **ddr1** in the search field to the right of the **View** menu, then click the magnifying glass icon ( ):



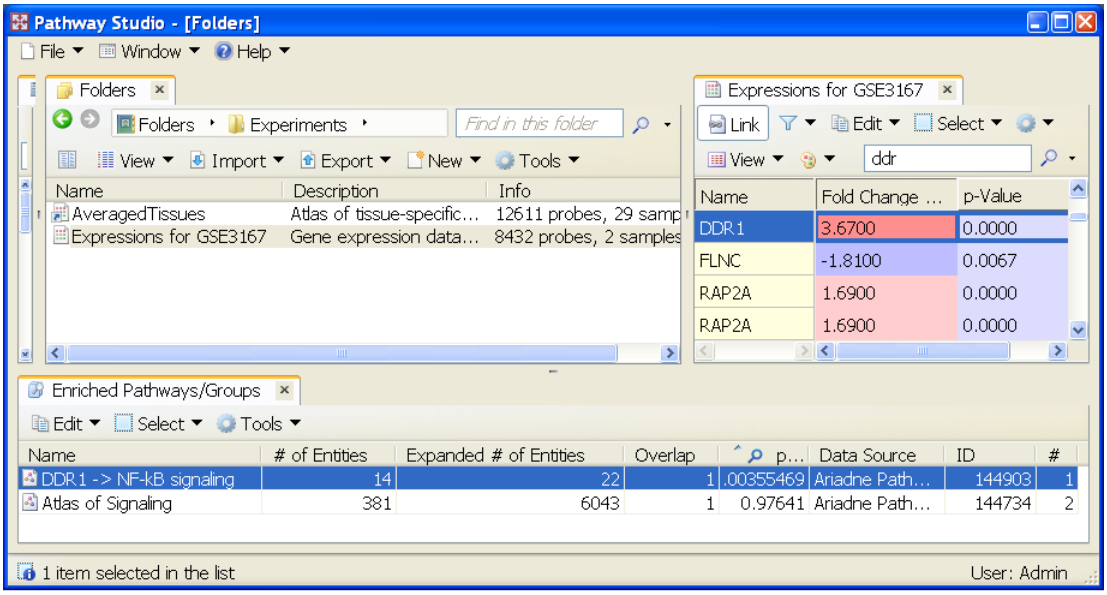
4. Right click the selected DDR1 entry, then select **Find Pathways/Groups Enriched with Selected Entities**:



5. In the pop-up dialog, check the **Ariadne Pathways** check box, then click **OK**.



Pathway Studio displays any pathways it finds inside the bottom pane of the window:

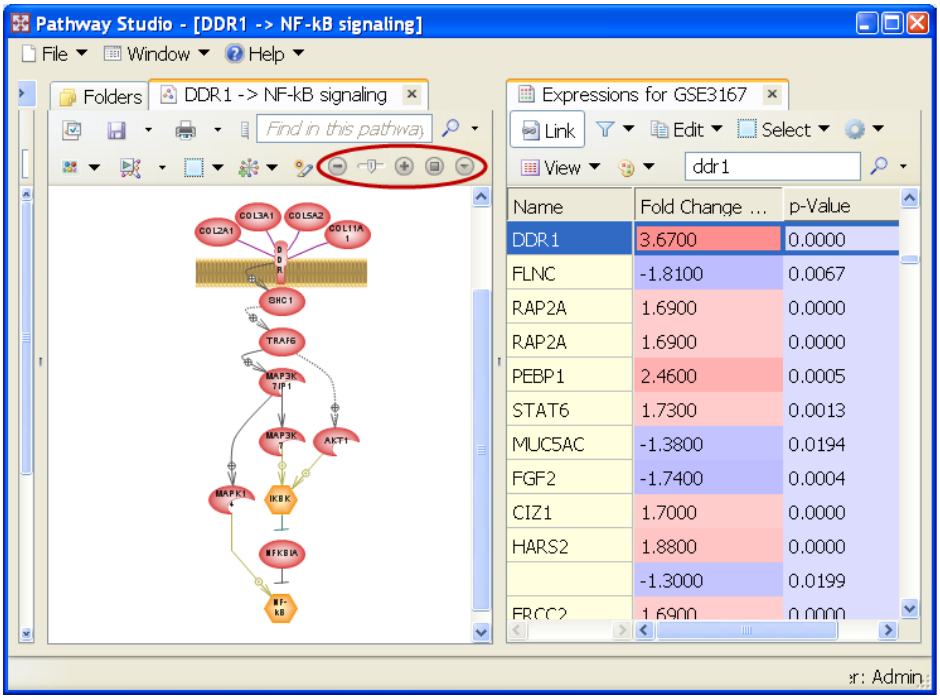


6. Double-click the pathway **DDR1 -> NF-kB signaling**.

   In a few seconds, the pathway visualization appears in the pane to the left of the heat map.

7. Click the **X** icon (✕) in the **Enriched Pathways/Groups** tab in the bottom pane to remove the pane from the Pathway Studio window and provide more room for the visualization.

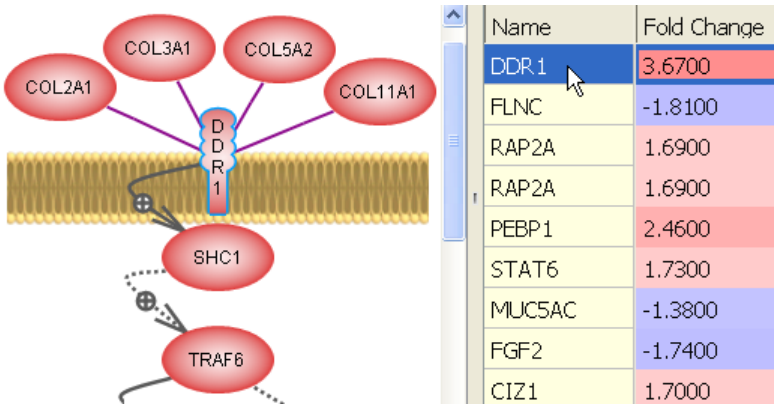
8. Maximize the window, then adjust the size of the visualization with the controls indicated in the figure below:



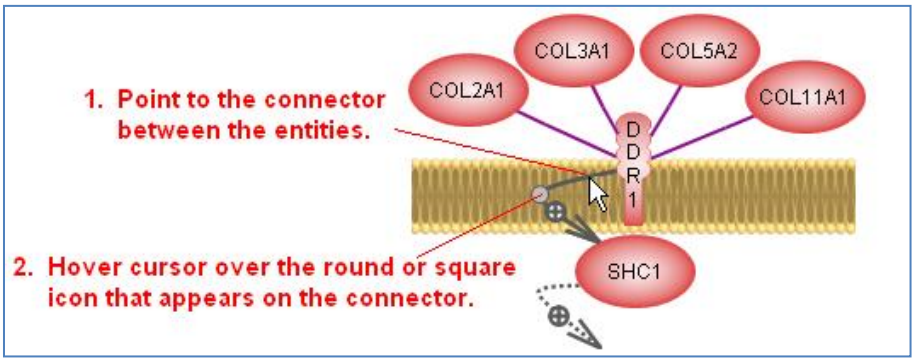
9. Explore the entities and the relationships in the visualization. For example:

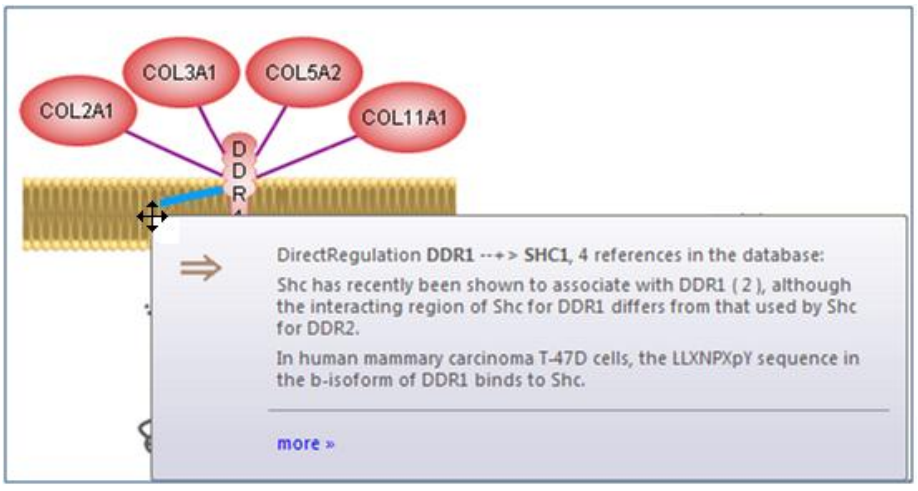
   - To locate the DDR1 gene in the visualization, click one of the DDR1 entries in the heat map.

     The selected gene is highlighted by a blue border, as shown below:

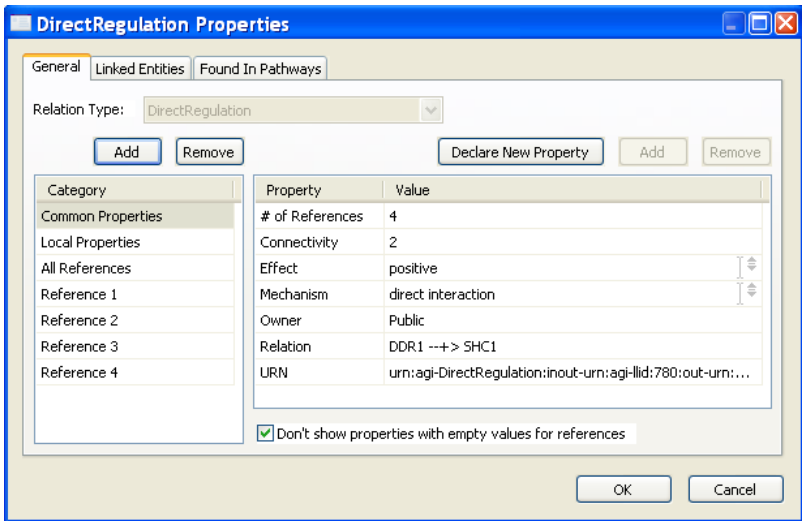


   - Display brief information about the relationship between two entities, as illustrated in the figures below:

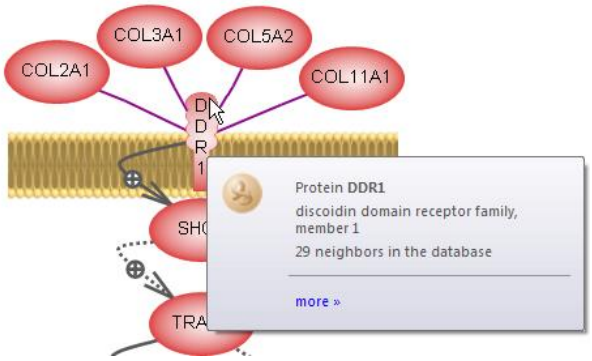




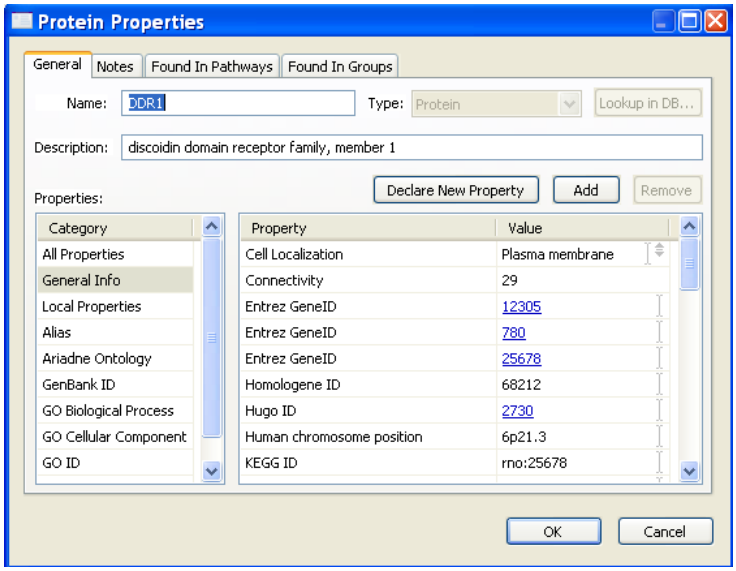
- Double-click a connector between two entities to display detailed information about the relationship between the entities:



- Hover the cursor over a gene or other entity in the visualization to display brief information about the entity:



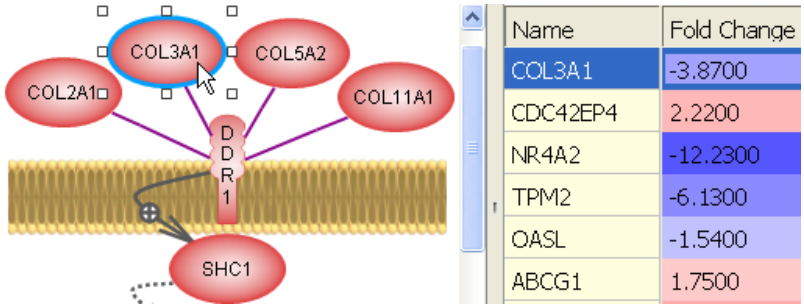
- Double-click a gene or other entity to display detailed information about the entity:

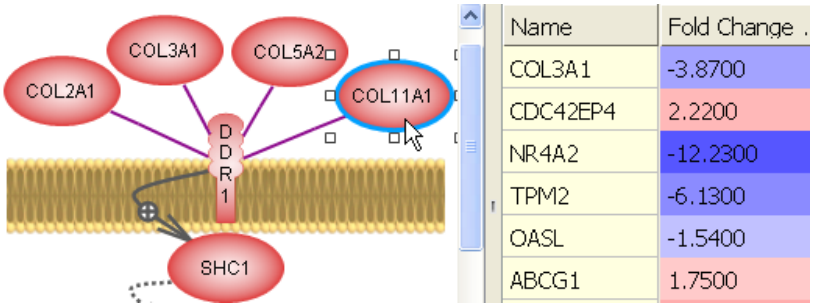


- Click a gene in the visualization to see if it is one of the genes imported from the experiment.

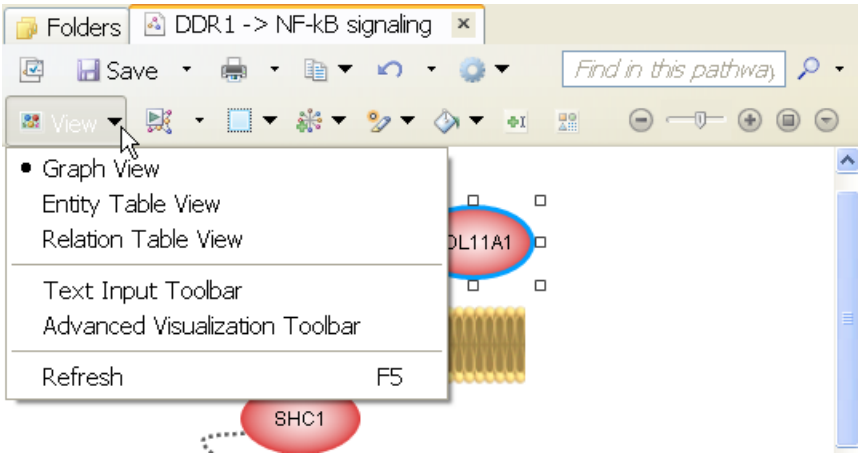
  If the selected gene is one of the imported genes, the gene becomes highlighted in the heat map.

  For example, in the following figures, COL3A1 is one of the imported genes, and COL11A1 is not:

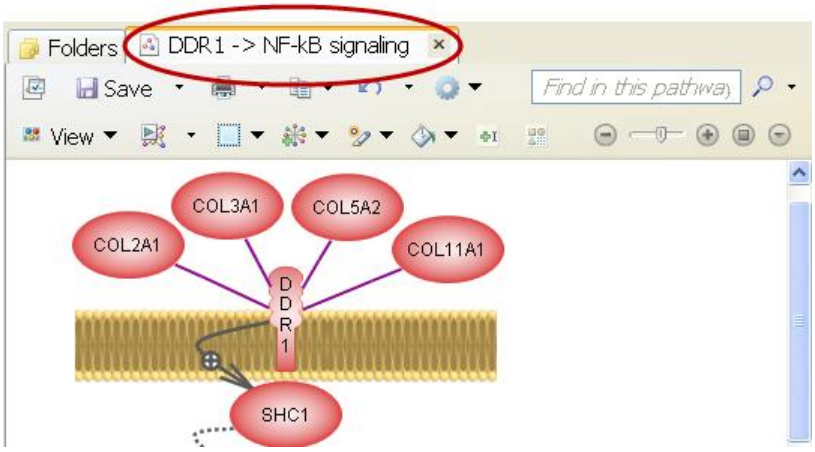




- Click the **View** menu above the visualization to see the pathway data in **Entity Table View** and **Relation Table View** (you may need to expand the window):



10. When finished familiarizing yourself with the visualization, click the **X** icon (×) in the **DDR1 -> NF-kB signaling** tab to close the visualization.



11. When prompted to save the pathway to the database, click **No**.
12. Leave Pathway Studio running. You will use it in the next lesson.

Lesson 10

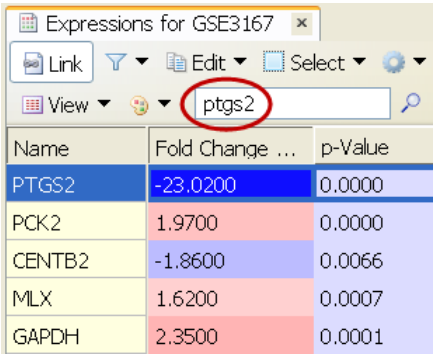
Create a Pathway from Two Genes of Interest

Lesson Goals: Create a new pathway in Pathway Studio.

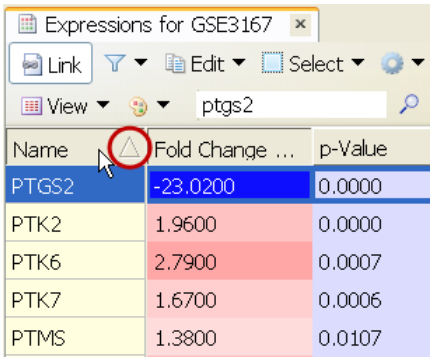
> **Note:** To perform this lesson, you must have Pathway Studio installed, and you must have downloaded the ResNet Mammalian Database.
>
> **Note:** This lesson uses the demo version of the Mammalian database. The demo database contains a subset of the data in the Mammalian database, and should not be used for actual research.

Scenario: In addition to the DDR1 gene, you are also interested in the gene PTGS2, and how they both relate to cancer. PTGS2 was included with the sample data you imported into Pathway Studio in the previous lesson. To find the relationship, if any, between these genes and cancer, you decide to create a new pathway based on these genes.

1. In the right pane of Pathway Studio, type **PTGS2** in the search field above the heat map, then click the magnifying glass icon ( ):

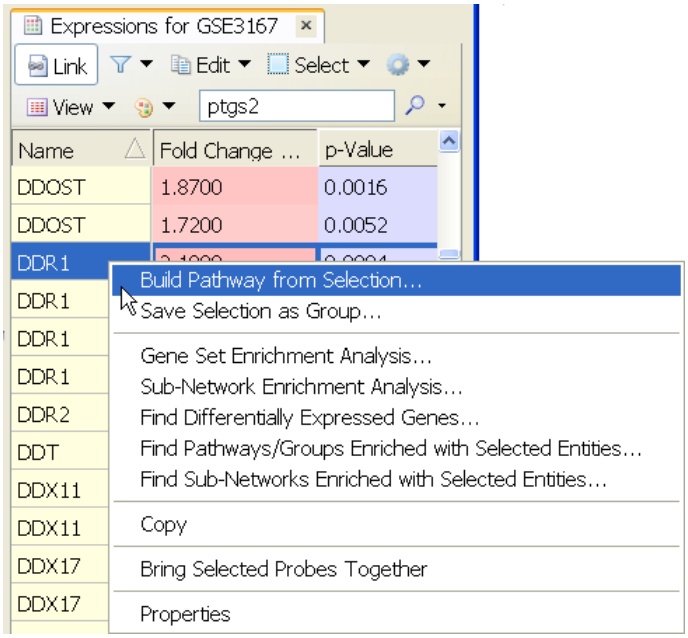


2. Look at the heading of the **Name** column. If you do not see an upward-pointing triangle after the word **Name**, click the heading until you do. This action sorts the gene names alphabetically in ascending order:



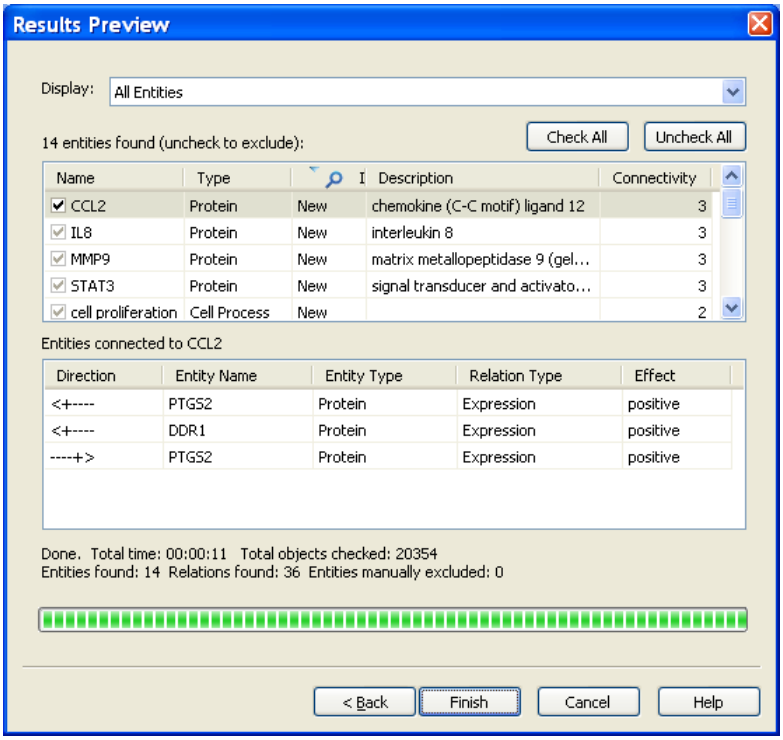
3. Scroll the heat map up until the gene **DDR1** appears.
4. Press and hold down the **Ctrl** key, then click one of the **DDR1** entries.

   The genes PTGS2 and DDR1 are now both selected.
5. Right-click **DDR1**, then select **Build Pathway from Selection**:



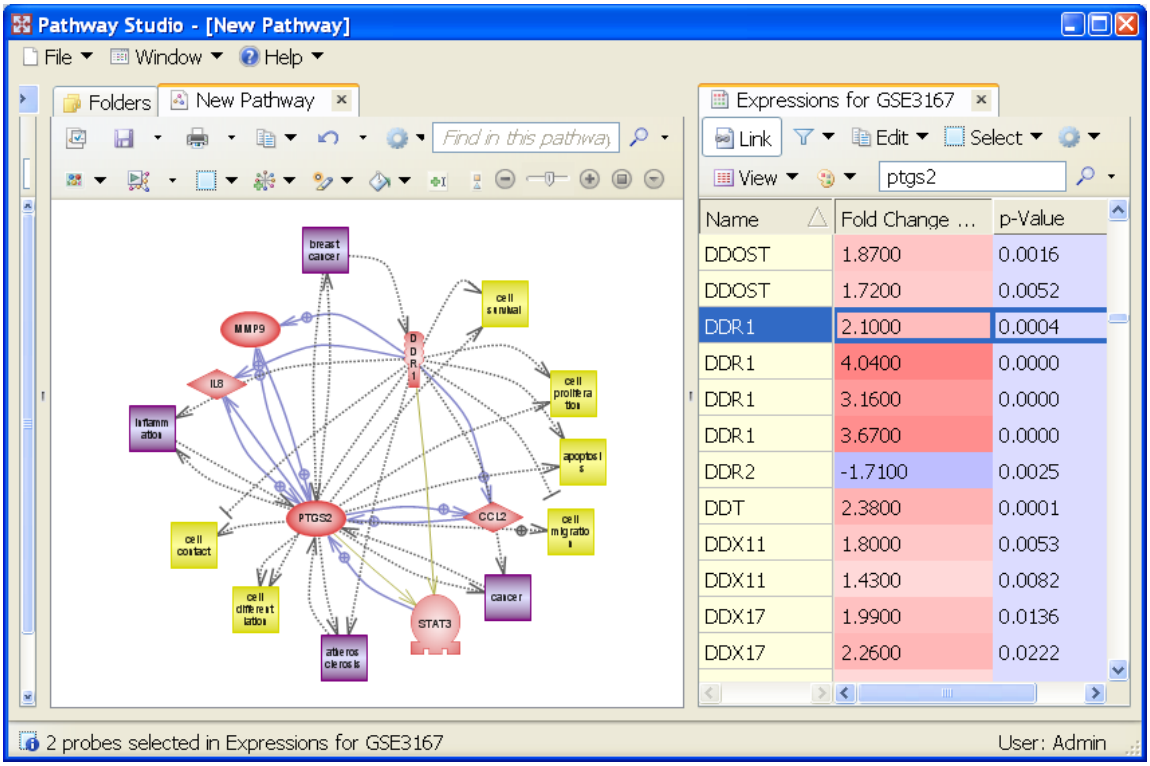
6. Select **Add shortest path** in the Select Algorithm Type dialog, then click **Next**.
7. Make sure that both **PTGS2** and **DDR1** appear in the Select Directions for Each Entry dialog. If they do not, repeat Step 1 through Step 4.
8. Click **Next**.
9. Make sure that all **Entity Type** entries in the Set Filter Parameters dialog are checked, then click **Next**.
10. Make sure that all **Relation Type** entries are checked, then click **Next**.

The Results Preview dialog appears, and Pathway Studio builds your pathway. When the build is finished, the dialog appears as shown below:



11. Click **Finish**.

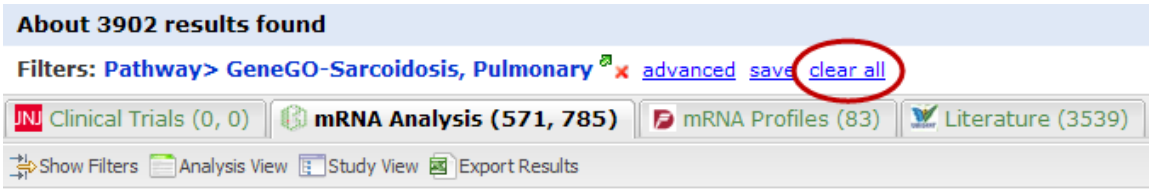
In a few seconds, the visualization of the new pathway appears:



12. Explore the entities and the relationships in the pathway, using the methods described in the previous lesson.
13. When finished, click the **X** icon in the **New Pathway** tab.
14. When prompted to save the new pathway to the database, click **Yes**.

    The Save Pathway dialog appears.
15. Click the ellipsis button (...).
16. Select the **Experiment** folder, then click **OK**.
17. Assign the pathway the name **PTGS2 and DDR1**, then click **Save**.
18. Close Pathway Studio.

    After Pathway Studio closes, the tranSMART window reappears.
19. Click **clear all** to prepare for the next lesson:



Lesson 11

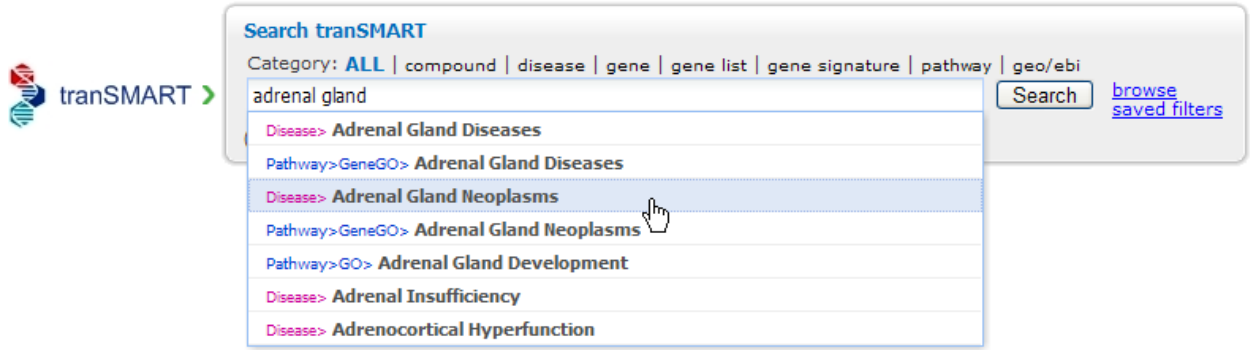
Create a Visualization of a Literature Search Result

Lesson Goals: Create a graphical representation of the interactions between the biomarkers returned from a literature search.

> **Note:** To perform this lesson, you must have Pathway Studio installed, and you must have downloaded the ResNet Mammalian Database.
>
> **Note:** This lesson uses the demo version of the Mammalian database. The demo database contains a subset of the data in the Mammalian database, and should not be used for actual research.

Scenario: You want to search for curated literature related to adrenal gland neoplasms. You are interested in seeing what relationships exist between the biomarkers returned from the search.

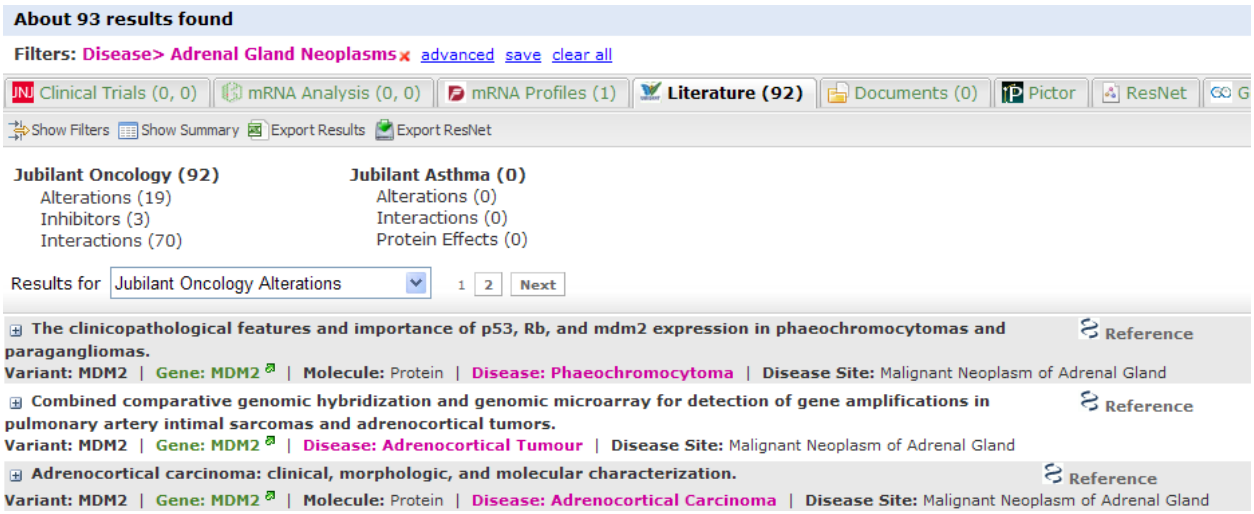
1. Type **adrenal gland** in the tranSMART Search field.
2. Click **Disease> Adrenal Gland Neoplasms** in the dropdown list:



The search begins immediately, and a search result is returned.

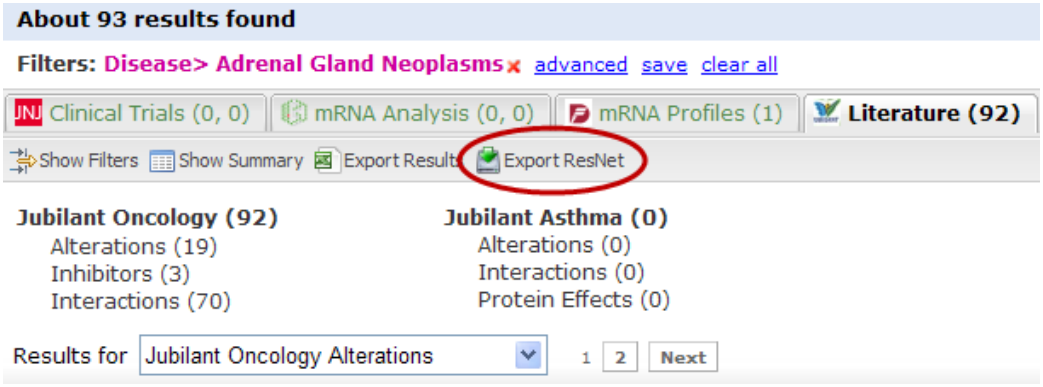
3. Click the **Literature** tab.

   tranSMART displays the search results for Jubilant-curated literature:



   Notice that there are oncology results for alterations, inhibitors, and interactions. However, tranSMART only supports the export of alterations and interactions data to Pathway Studio.

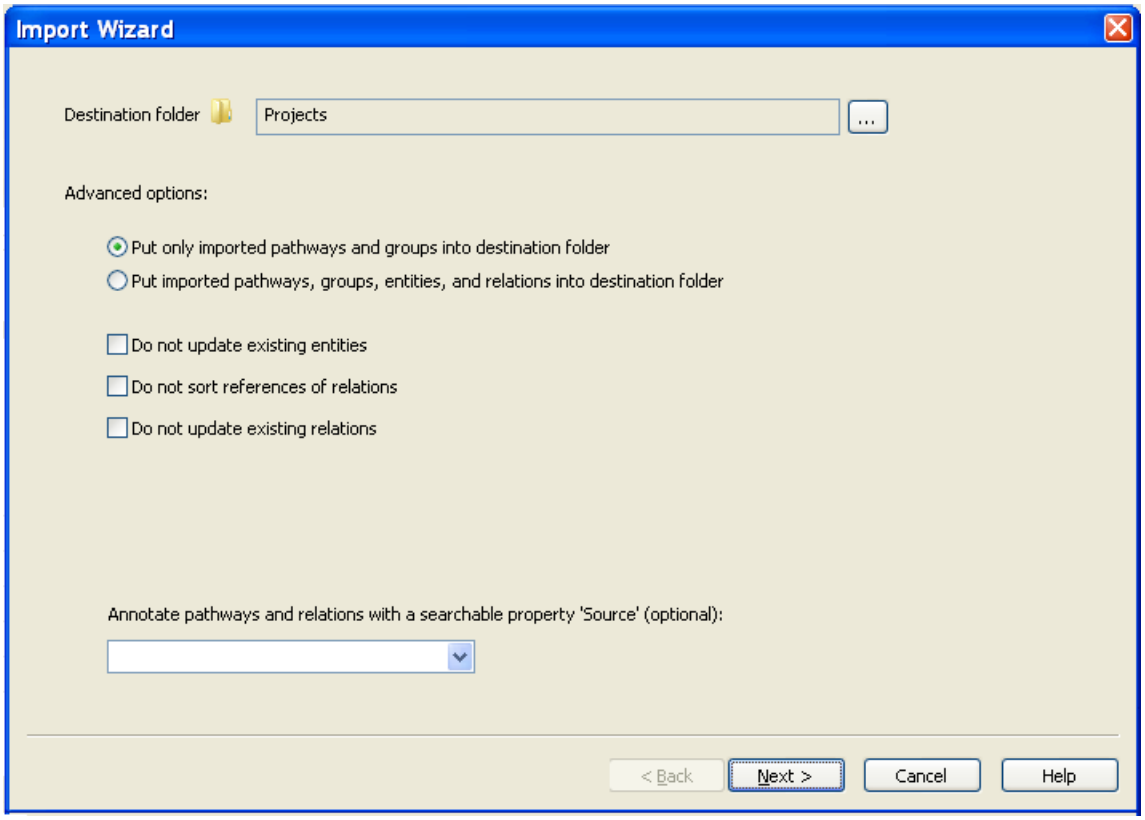
4. Click the E**xport ResNet** button:



5. When the File Download dialog appears, click **Open**.

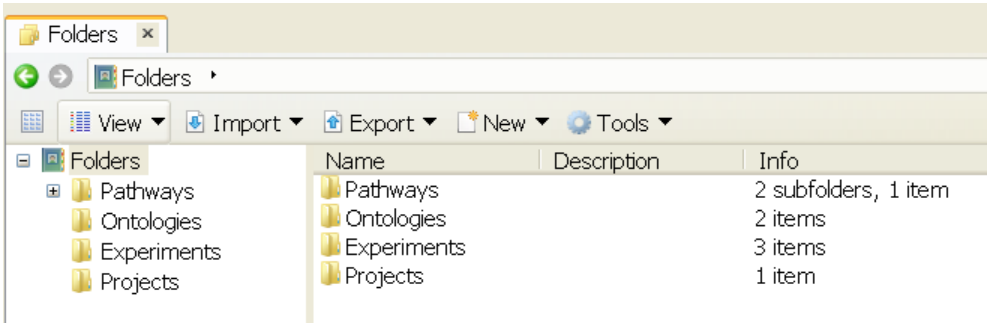
   When you click **Open**, Pathway Studio starts up, and then launches the Import Wizard. It may take a few seconds for the Import Wizard to appear.

6. Accept all the default settings on the Import Wizard, as shown below:

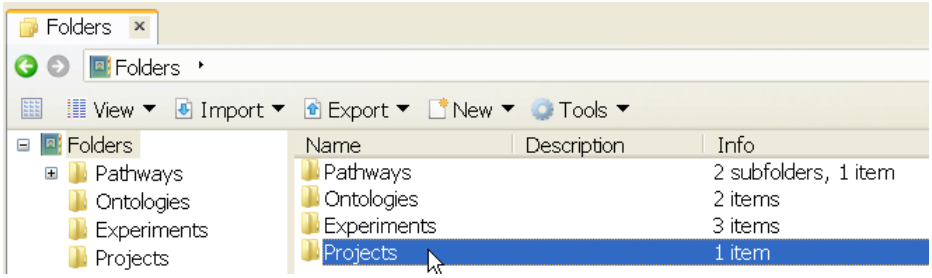


7. Click **Next**.
8. When Pathway Studio is finished building the visualization, click **Finish**.

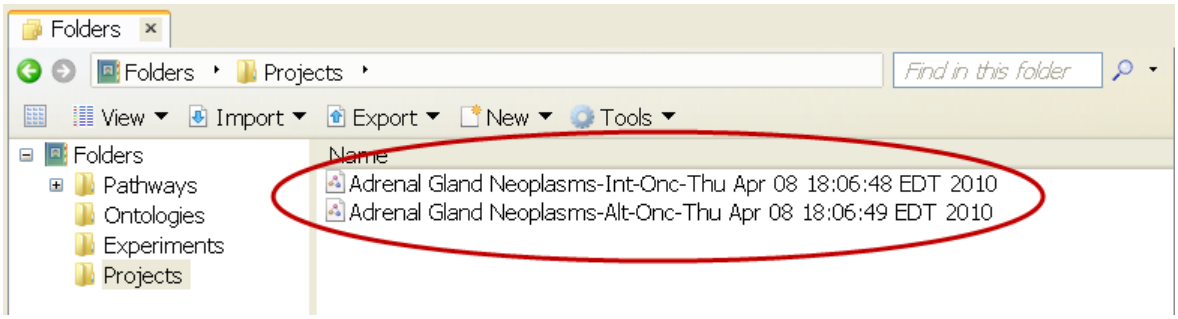
   After a few seconds, the list of folders appears:



9. Double-click the **Projects** folder.



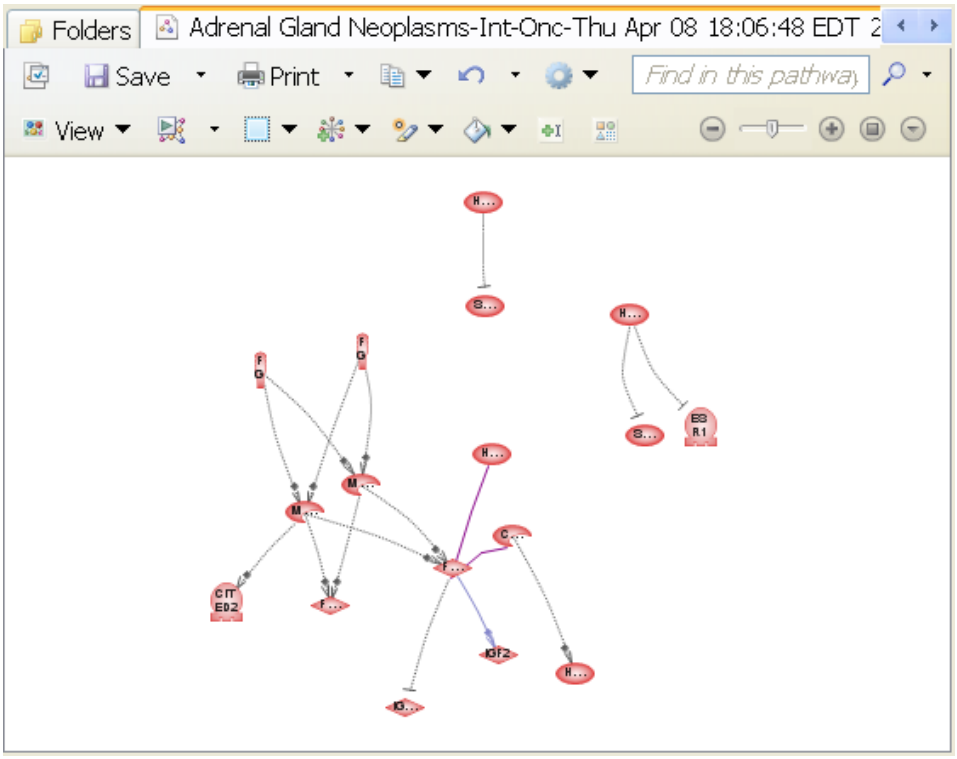
The new entries for the curated literature results are listed in the Projects folder. The names assigned to the entries consist of the name of the tranSMART search filter, abbreviations of the data types, and the date the entries were created – for example:



You are interested in the interactions data (abbreviation **Int-Onc** in the name above).

10. Double-click the entry for the interactions data you just exported.

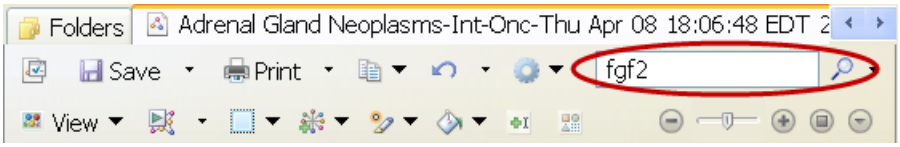
In a few seconds, the visualization appears:



Notice the relationships and the non-relationships between the biomarkers cited in the curated literature.

You are interested in the interaction between FGF2 and IGF2 and want to read source documentation about it.

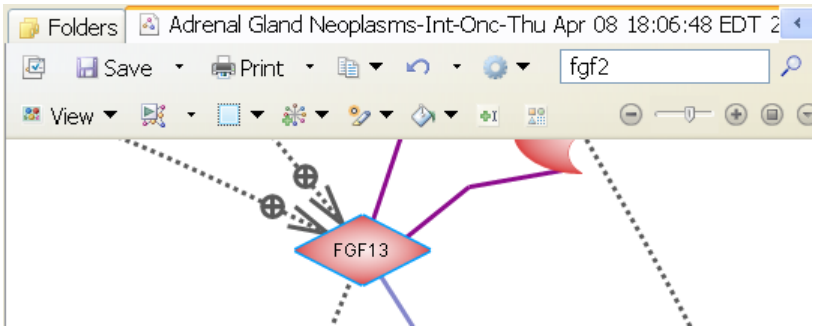
11. Type FGF2 in the ***Find in this pathway*** box, then click the magnifying glass icon ( ):



If FGF2 is in the visualization, it will have a blue border around it.

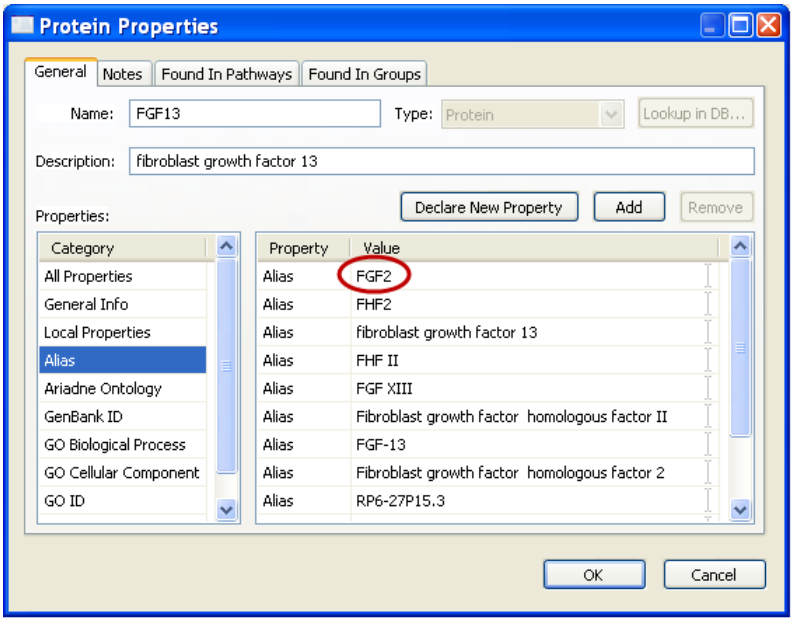
12. Enlarge the visualization to find the biomarker with the blue border (it's in the lower right cluster).

You notice that the biomarker with the blue border is not FGF2, but FGF13:



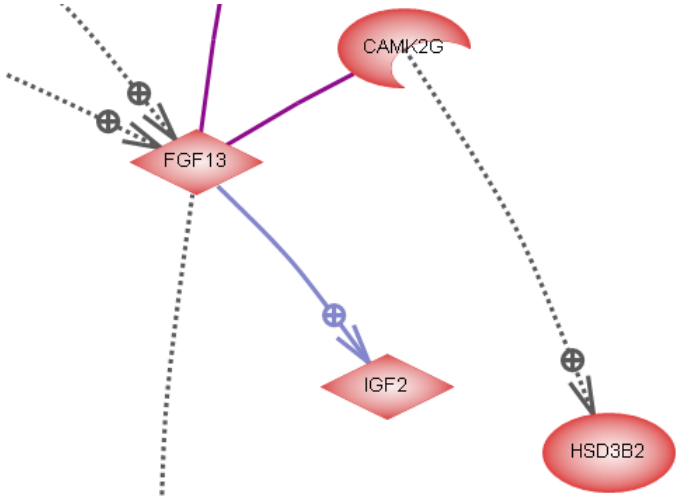
13. Double-click FGF13, then click the property **Alias** in the Protein Properties dialog.

You see that FGF2 is an alias for FGF13:



14. Click **OK** to close the Protein Properties dialog.

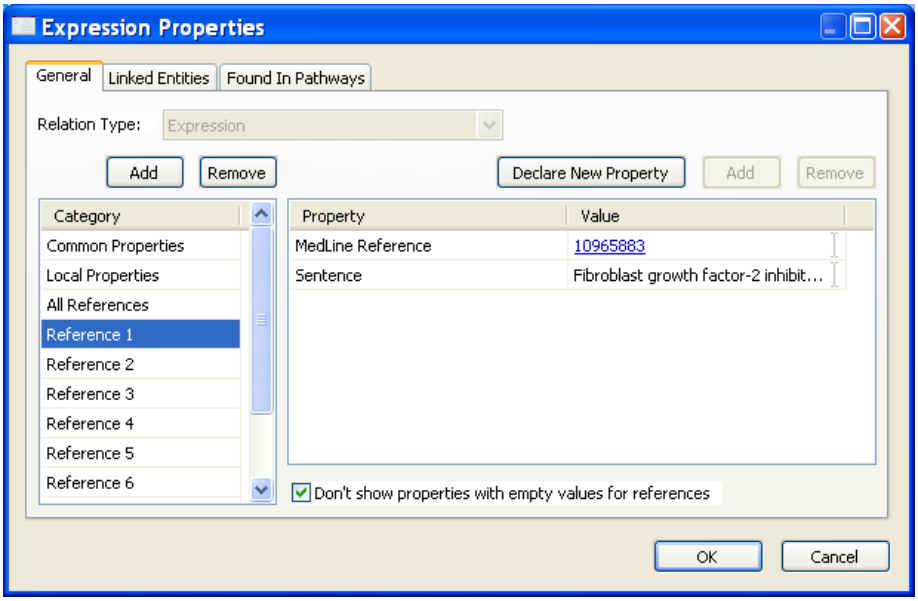
    Looking at the biomarkers around FGF13 in the visualization, you see that IFG2 is below it, connected by an expression relationship line:



15. Double-click the connecting line between FGF13 and IGF2:

    The Expression Properties dialog appears.

16. Click **Reference 1**:



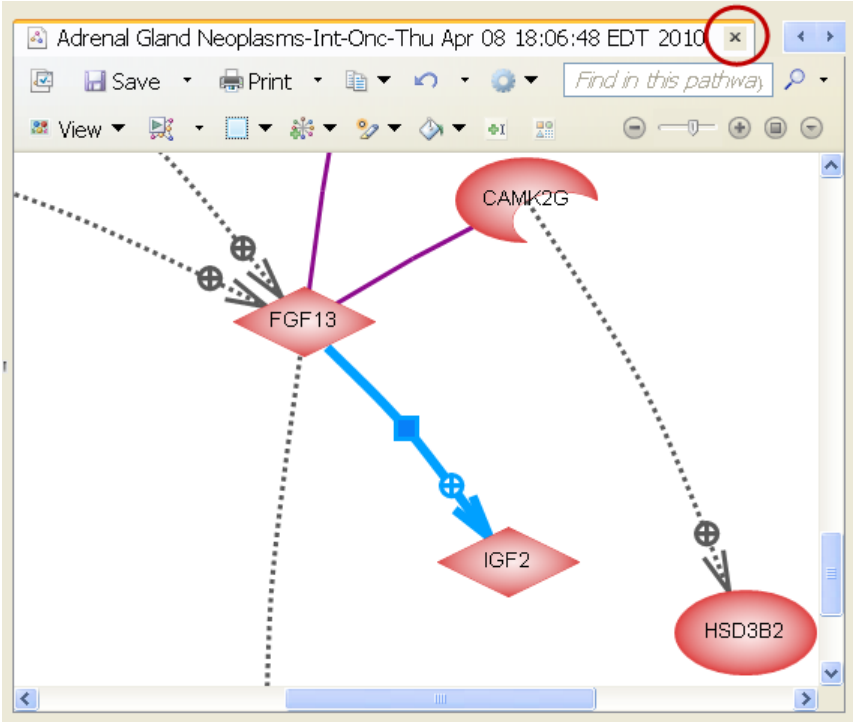
17. Click the value shown for the **MedLine Reference** property (10965883 in the figure above).

    The referenced PubMed article is displayed.

18. When finished reading, close the PubMed article.

19. Click **OK** on the Expression Properties dialog to close it.

20. Close the visualization by clicking the **X** icon (×) on the visualization tab:



21. If you have made any changes to the visualization, click **Yes** to save them when prompted.
22. Close Pathway Studio.
23. Close tranSMART.