

Pathway Studio[®] Desktop 7.1
Pathway Studio PC-Client 7.1 (Enterprise)
MedScan Reader[®] version 3.1

Training Manual

Version 4.1

Preface

This manual is applicable to Pathway Studio users of both the Desktop edition (using a local database) and to the Pathway Studio PC-Client edition (Pathway Studio Enterprise, server-based product.) Applications and functions that are specific for the Enterprise version of Pathway Studio are indicated in *italics*.

Instructions for the installation of Pathway Studio are not included in this manual. Please see the Pathway Studio Installation Guide for information on the installation of Pathway Studio 7.1, MedScan Reader 3.1 and the ResNet Mammalian 7.0 database or the ResNet Plant 2.1 database on your computer.

While this manual is based on the ResNet 7.0 Mammalian database, the basic tool functionality is the same if one uses the Plant database or the ChemEffect® database as well.

How to Use this Manual

This manual is designed to walk you through some major workflows in Pathway Studio and MedScan Reader utilizing the ResNet 7.0 Mammalian Database. Each section has an introduction to the tools in Pathway Studio and MedScan Reader and examples of how they can be used, followed by a hands-on work example. Each example builds on information from the previous exercises, and sometimes utilizes files generated from previous examples. The topics covered in this manual approximate what would be covered in a day's hands-on training course.

Two example data files accompany this training manual:

- GDS2126.soft – This is a GEO dataset soft file containing a microarray experiment
- SCLC genes – this is an MS Excel spreadsheet containing genes differentially expressed between normal lung and small cell lung tumor samples.

You will need to download these files to be able to reproduce the hands-on exercises. You can find these files in the same location in the Support section of Ariadne's website as this Training Manual.

ResNet® Mammalian Database, ResNet® Plant Database, ChemEffect® Database, MedScan® Reader, Pathway Studio®, Desktop Edition software, Pathway Studio®, PC Client software are registered trademarks of Ariadne Genomics, Inc.

MS Excel® is a registered trademark of Microsoft Corporation.

Affymetrix GeneChip® is a trademark of Affymetrix, Inc.

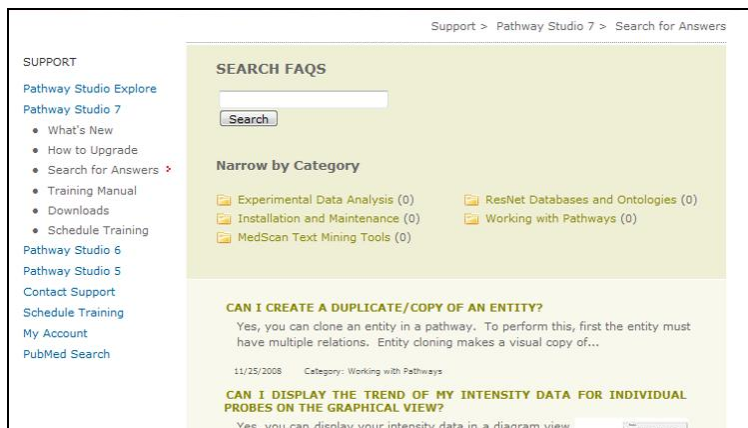
GenomeStudio™ is a trademark of Illumina, Inc.

Other Training and Support Resources

The Ariadne Technical Support site has additional resources available to enable Pathway Studio users including:

- Key word searchable FAQs:

<http://www.ariadnegenomics.com/support/pathway-studio-7/search-for-answers/>



Analysis Guides (examples of workflows to answer biological questions):

<http://www.ariadnegenomics.com/support/pathway-studio-7/training-material/>

TITLE	DATE	TYPE	SIZE	FILE TYPE
PS RN6 Analysis Guide: How can I find diseases that are involved in the amplification of a specific gene? *	Jan 15, 2009	Analysis Guide	305 Kb	PDF
PS RN6 Analysis Guide: How can I find diseases that are involved in an activation mutation of specific gene? *	Jan 15, 2009	Analysis Guide	191 Kb	PDF
PS RN6 Are fusions of my gene (from chromosomal translocations) involved in any diseases? *	Jan 15, 2009	Analysis Guide	185 Kb	PDF
PS RN6 Analysis Guide: Are splicing variants of my gene involved in diseases? *	Jan 15, 2009	Analysis Guide	256 Kb	PDF
PS RN6 Analysis Guide: What genes are transcriptionally regulated by my protein? *	Jan 15, 2009	Analysis Guide	171 Kb	PDF
PS RN6 Analysis Guide: What proteins regulate the activity or abundance of my protein? *	Jan 15, 2009	Analysis Guide	200 Kb	PDF
PS RN6 Analysis Guide: Does my protein	Jan 15, 2009	Analysis Guide	213 Kb	PDF

For users of the ChemEffect® database, a Quick Start Guide:

<http://www.ariadnegenomics.com/support/pathway-studio-7/training-manual/>

ChemEffect Training			
TITLE	TYPE	SIZE	FILE TYPE
> Quick Start Guide for ChemEffect *	Basic Documentation	486 Kb	PDF

For users interested in adding miRNA predicated relationships to their mammalian database:
<http://www.ariadnegenomics.com/support/pathway-studio-7/downloads/>

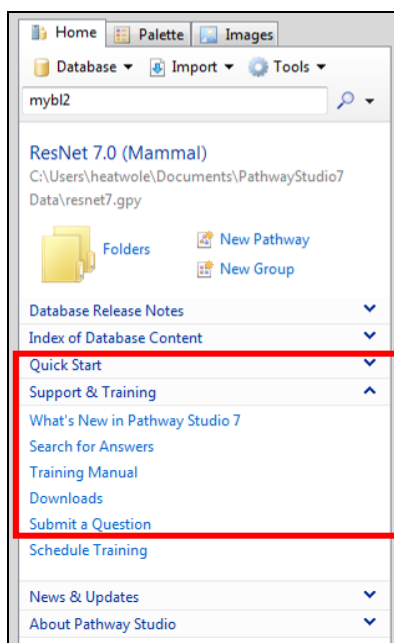
ADDITIONAL DATABASES			
TITLE	TYPE	SIZE	FILE TYPE
› Introduction to miRNA Analysis in Pathway Studio *	Basic Documentation	580 Kb	PDF
› Predicted miRNA Relationships *	Database	525 Kb	ZIP/Text

Purchase Instructor Led Training

In addition instructional training may be purchased from Ariadne, for an individual or a group, as either on-site or web based sessions. If you are interested in purchasing training, please contact your local sales representative or sales@ariadnegenomics.com.

Access to Technical Support

Technical Support for Ariadne's products can be easily accessed from Pathway Studio:



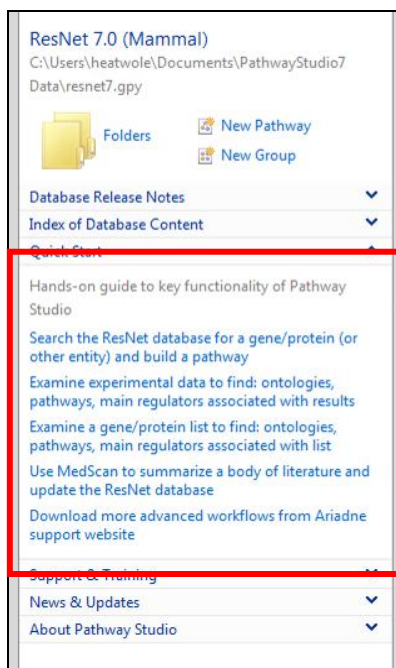
Links within the Information Pane in Pathway Studio take you directly to the support website. Registration and log-in to this website is required.

In addition, you can access the support site directly at:

<http://www.ariadnegenomics.com/support/pathway-studio-7/>

A link is provided here to submit questions directly to Technical Support.

Quick Start Guides



In addition this Training Manual, you can access the Quick Start Guides through the Information Pane. These Quick Start Guides will walk you through the steps of some common analysis workflows for Pathway Studio and MedScan Reader.

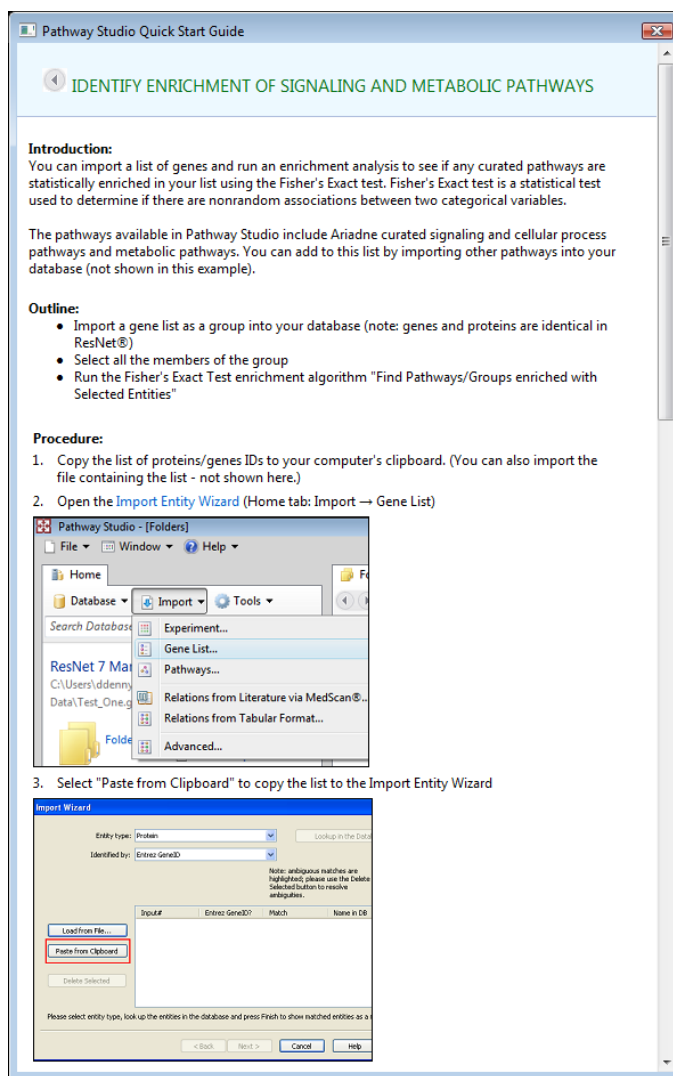


Table of Contents

Preface	2
Section 1: Introduction to Pathway Studio, MedScan Reader and ResNet Mammalian Database	9
What can Pathway Studio do for you?	9
MedScan Reader	10
ResNet 7 Mammalian database	11
Curated Pathway Collection.....	12
Ariadne Gene Ontology/Gene Ontology.....	12
Quarterly Updates for ResNet 7 Mammalian Database	13
Section 2: Introduction to the Pathway Studio interface	14
Opening a Database	15
Downloading a database.....	15
Information Pane	16
Folders View.....	17
• Exercise One: Open the local ResNet Mammalian Database (Desktop Users Only)	18
• <i>Exercise Two: Open the server-hosted ResNet Mammalian Database (PC-Client, Enterprise only)</i>	19
Section 3: Building Pathways	20
Search ResNet for an Entity	20
Import a list of Entities.....	20
Import Network Data	21
• Exercise Three: Search ResNet for Entities and Relationships and Import a Protein List	23
Build Pathway Tools in Pathway Studio.....	26
Build Pathway Tool – Quick Menus	26
Build Pathway Tool – Advanced Menu Desktop.....	27
<i>Build Pathway Tool – Advanced Menu Enterprise</i>	28
Graph View/ Entity Table view / Relation Table view.....	29
Customizing tables, filtering by significance	29
Viewing Details about Entities/Relationships	30
Creating new Entities and Relationships	31
Customization of Pathway Layouts.....	32
• Exercise Four: Building Pathways	34
Saving Pathway Images.....	41

Exporting Pathways.....	41
<i>Publishing Pathways (PC-Client, Enterprise only)</i>	42
Section 4: Importing Experimental Data	43
Data formats that can be imported into Pathway Studio.....	43
Mapping Files for Microarrays	44
Using the Import Wizard to import data	44
• Exercise Five: Import data	45
Section 5: Experimental data analysis tools.....	51
Fisher’s Exact Test.....	51
• Exercise Six: Experimental Data Analysis – Fisher’s Exact Test for Enriched Groups and Pathways.....	55
Gene Set Enrichment Analysis	58
• Exercise Seven: Experimental Data Analysis – Gene Set Enrichment Analysis	60
Sub-Network Enrichment Analysis.....	63
Defining the Sub-Networks.....	63
Both enrichment algorithms, Fisher’s Exact Test and Gene Set Enrichment Analysis, can be used to identify enriched Sub-Networks.	65
• Exercise Eight: Experimental Data Analysis – Sub-Network Enrichment Analysis (with GSEA)	66
<i>Build Dense Expressed Networks (PC-Client, Enterprise only)</i>	70
• <i>Exercise Nine: Experimental Data Analysis – Build Dense Expressed Networks (PC-Client, Enterprise only)</i>	72
Section 6: MedScan Reader	73
Using MedScan to search internet resources:	73
Selecting Reading Settings: dictionaries and relation extraction	73
Select Reading Interests.....	74
Search PubMed through MedScan Reader browser	77
• Exercise Ten: Use MedScan to extract facts from the results of a PubMed query – Popular Destinations	82
Quick Import of PubMed Search.....	85
• Exercise Eleven: Use MedScan to extract facts from the results of a PubMed query – Quick Import	86
• Exercise Twelve: Query the MedScan Dictionary for an Entity	87
Saving a MedScan Project	89
Using MedScan to search local documents:	89
Import a document or directory	89
Appendix A: Definitions	90
Appendix B: User Roles for Enterprise Accounts	92

Appendix C: How to Import Network Data	93
Appendix D: Deleting Entities and Relations from a Local Database (Pathway Studio Desktop)	94

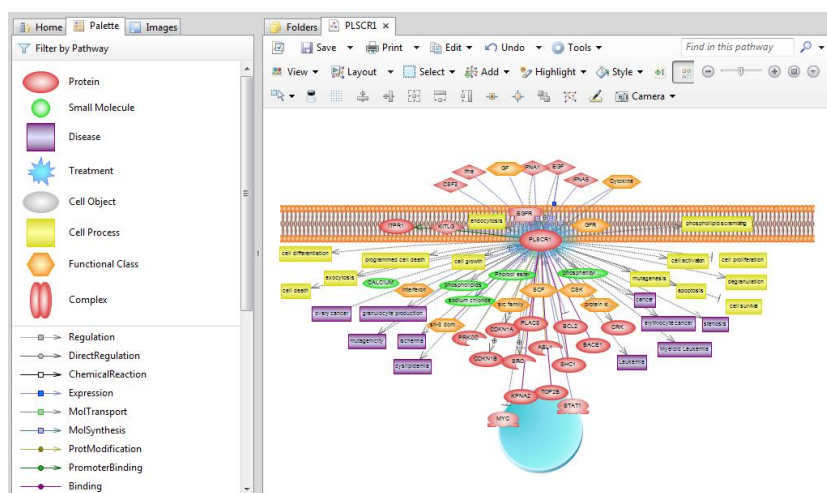
Section 1: Introduction to Pathway Studio, MedScan Reader and ResNet Mammalian Database

Congratulations on your decision to utilize Pathway Studio as an important tool to support your biological research. Pathway Studio helps you to interpret experimental data in the context of pathways, gene regulation networks and protein interaction maps; interpret microarray and proteomics data, classify and prioritize proteins, draw pathway diagrams, and automatically update your pathways with newly published facts using MedScan Technology.

What can Pathway Studio do for you?

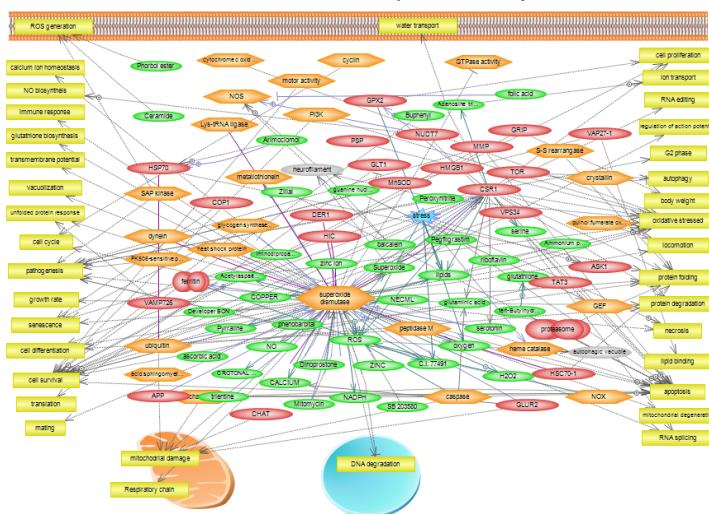
- Identify relationships among proteins, small molecules, cell processes and diseases

What is known to interact with my protein? What processes are associated with my protein?



- Automatically extract biological interactions from the scientific literature

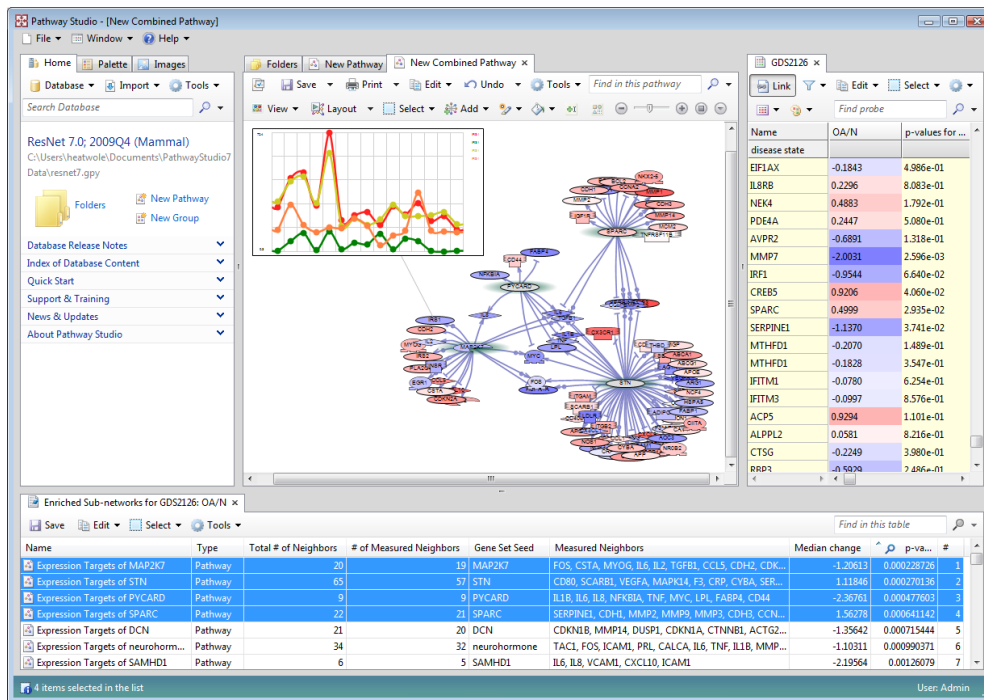
Create a Literature Network for a specific subject



Literature relation summary for all PubMed abstracts that mention SOD1 (cytoplasmic superoxide dismutase) and ALS (amyotrophic lateral sclerosis)

- **Build and reconstruct pathways from your microarray and other high-throughput data**

Analysis Tools to find enriched networks in experimental data



MedScan Reader

MedScan technology is a Natural Language Processing Technology used to extract relationship information from biomedical literature. Ariadne utilizes MedScan technology with ConceptScan®, Ariadne's latest generation of Natural Language extraction technology, to build the ResNet7 Mammalian Database, the mammalian (human, mouse and rat) database, of relationship information summarized from all abstracts in PubMed as well as information contained in 61 free full-text journals.

Here is an example of how MedScan works. MedScan reads sentences and identifies entities (proteins, complexes, small molecules, diseases etc.) in sentences, here depicted in red.

Sentence in scanned literature: **"Axin binds beta-catenin and inhibits GSK-3beta."**

- Identify Proteins in Dictionary (in red): **[Entities]**

"Axin binds beta-catenin and inhibits GSK-3beta."

Next MedScan utilizes grammar rules to identify described relationships between the sentences

- Identify Interaction Type (in blue): **[Relationships]**

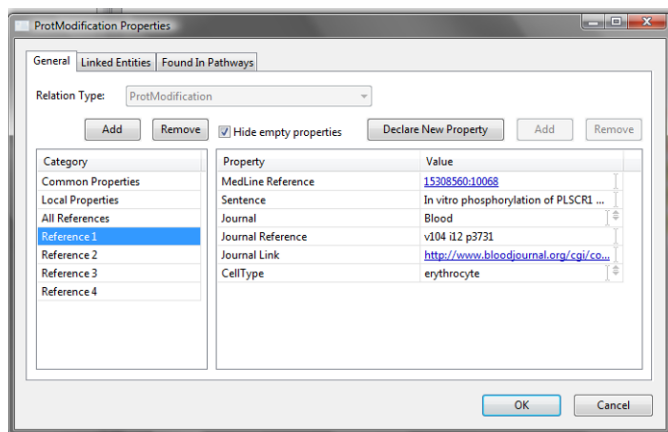
"Axin binds beta-catenin and inhibits GSK-3beta."

The extracted facts are the relationships, which are added to the ResNet database:

Extracted Facts:

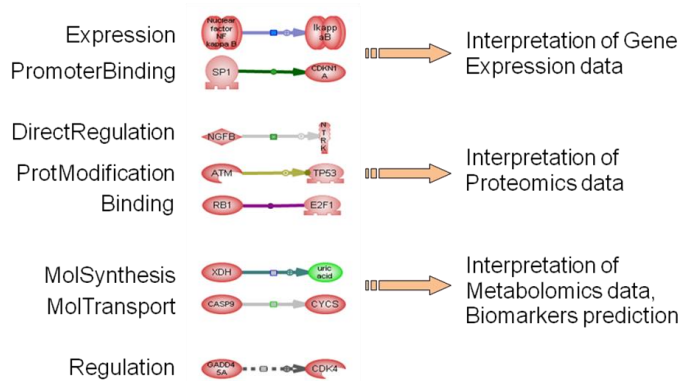
Axin - beta-catenin relation: **Binding**
Axin -> GSK-3beta relation: **Regulation, effect: Negative**

The sentences containing identified relationships are available for examination within the Pathway Studio interface as well as the literature reference.



The MedScan dictionary contains lists of entity names. (There are eight types of relationships between entities identified by MedScan. To see a definition of entity types and relationships types, see Appendix A. All relations have directionality except Binding.)

Types of Relationships Identified by MedScan



The Pathway Studio software includes the MedScan Reader component, which you can use to add additional information to your ResNet database, thus allowing you to utilize additional literature and information sources to augment your pathway building activities.

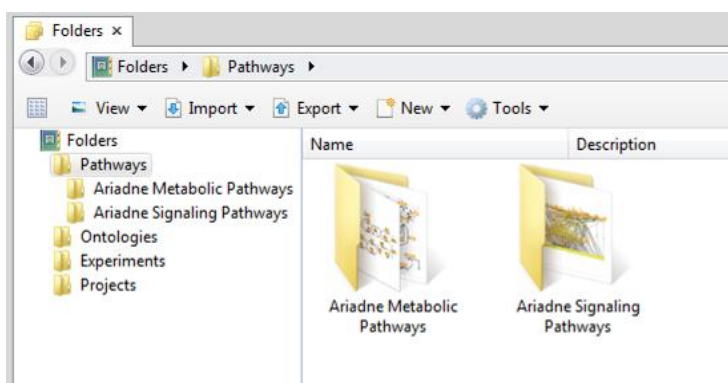
ResNet 7 Mammalian database

The ResNet Mammalian database includes over 800,000 unique relationships derived from over 19 million PubMed abstracts as well as 61 full-text journals. The extracted information is processed through the ResNet Curator to condense information and remove some technical false positives. In addition to information extracted by MedScan, ResNet includes MeSH terms for diseases based on the Medical Subject Headings (MeSH) from the National Library of Medicine (<http://www.nlm.nih.gov/mesh/meshhome.html>), GO terms from The Gene Ontology Consortium (<http://www.geneontology.org/>), Ariadne Ontology terms, and a collection of Ariadne curated reference pathways (227 receptor signaling pathways, 21 cellular process pathways and 39 metabolic pathways). For more information about what is included in the ResNet 7 database, please see the Technical Support Bulletin 0905 "Release of the ResNet® Mammalian 7.0 Database" which can be downloaded from the Ariadne Technical Support site.

Quarterly updates to ResNet are available for download from Ariadne's support website. Instructions for applying the update are also available on the support website.

Curated Pathway Collection

The ResNet database includes a collection of reference pathways including a large number of receptor signaling pathways and cellular process pathways. These can be found in the Pathways folder. Ariadne scientists have built these pathways based on general knowledge in order to provide you with useful building blocks to extend with your own specific expertise.



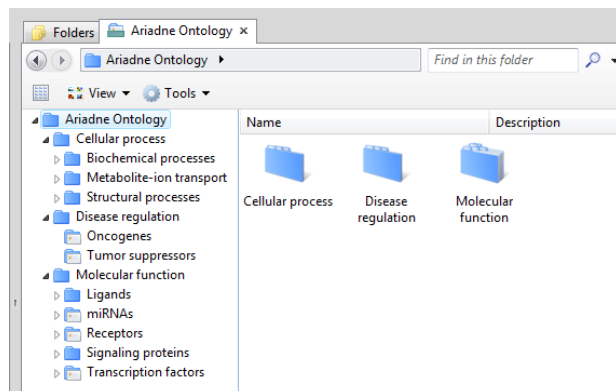
Receptor pathways curated into 227 signaling pathways

21 Cellular Process pathways

39 Metabolic Pathways

Ariadne Gene Ontology/Gene Ontology

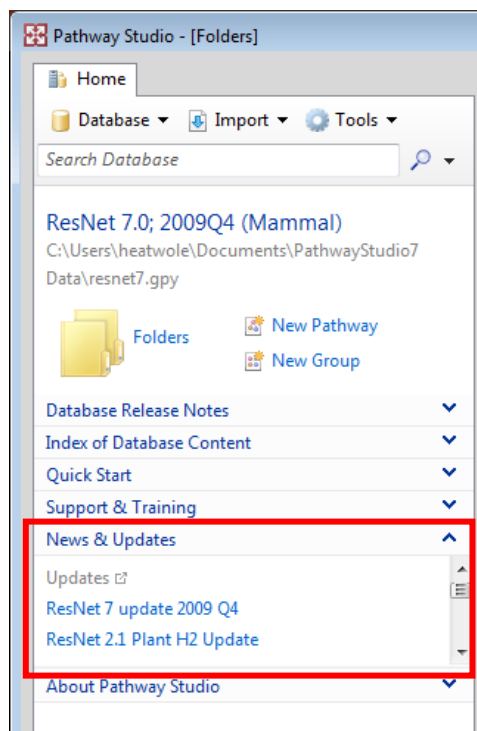
In addition to the well-known “GO” vocabulary from the Gene Ontology Consortium, Ariadne has prepared a robust ontology that has been optimized to provide the best results from analysis with Pathway Studio. Ariadne’s Gene Ontology is designed to sort genes into the appropriate category, rather than to assign multiple categories to a gene. The ontology is relatively flat, containing only three groups, cellular process, molecular function and disease regulation. Additionally, the Ariadne Gene Ontology has been designed to minimize redundancy in the classification, and consequently avoids much of the redundancy found in analysis results produced using the Gene Ontology Consortium’s vocabulary.



Organizes almost 9000 genes into 507 groups with three-tier biological hierarchy (created by Ariadne scientists)

This more succinct ontology can be used with Gene Set Enrichment Analysis and Fisher’s Exact test to obtain meaningful results.

Quarterly Updates for ResNet 7 Mammalian Database

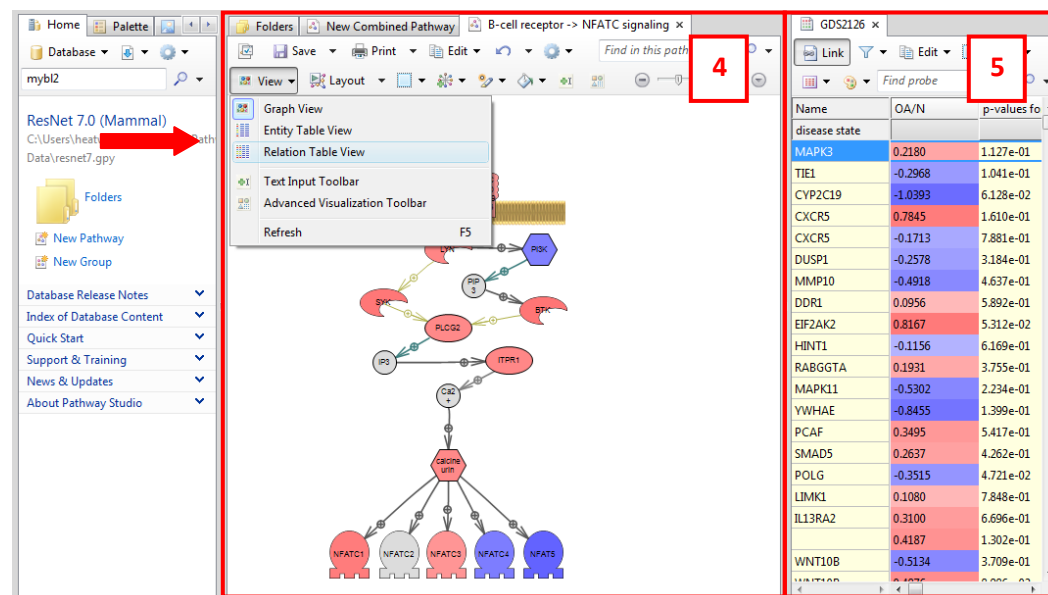
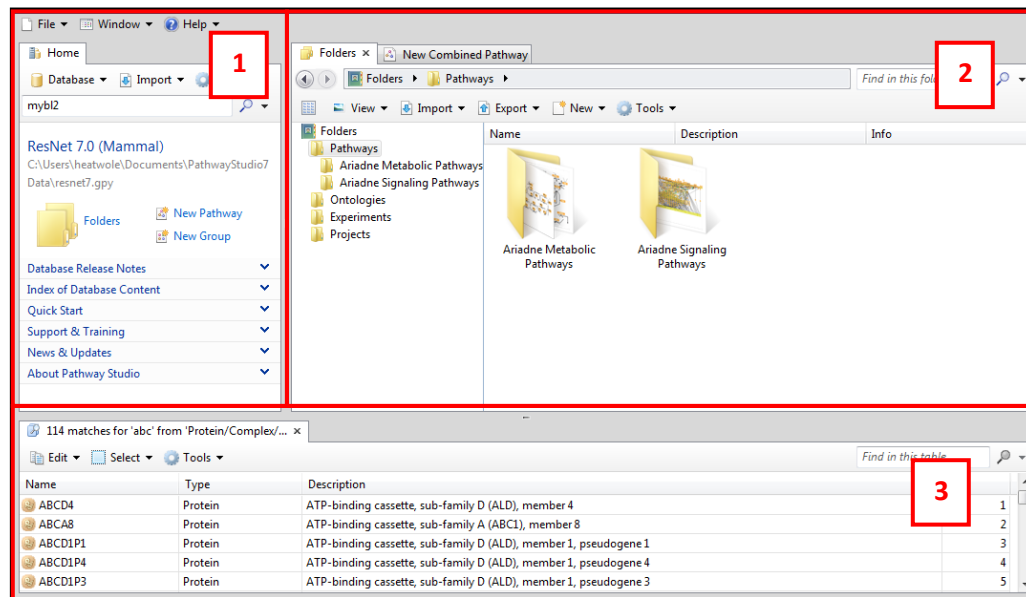


Quarterly updates for the ResNet 7 Mammalian database are available for download from Ariadne's Support website. The News & Updates section, located in the Information Pane, provides quick access to the update link.

Note: After you download the update file, Ariadne recommends that you allow the application of the update to run overnight, as the process may take several hours.

Section 2: Introduction to the Pathway Studio interface

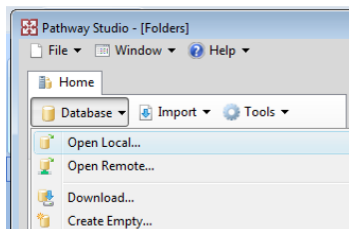
The major panes in the interface include: Information Pane (1), Folders Views (2), List Pane (3), Graph (and Relationship and Entity) View (4), Experiment Pane (5) (see following two figures).



Opening a Database

To before working with Pathway Studio, you must first select the database that is appropriate for your research.

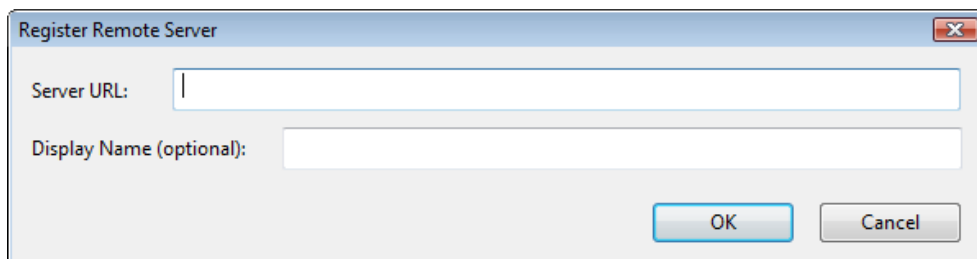
Desktop users: Pathway Studio's databases are hosted locally on your computer's hard drive.



To open a local database go to the Information Pane and select: Database > Open local.

Enterprise users only: You have the option of creating and working with locally hosted ResNet databases as described above, or working with a shared central database hosted on the Pathway Studio Enterprise server at your institution.

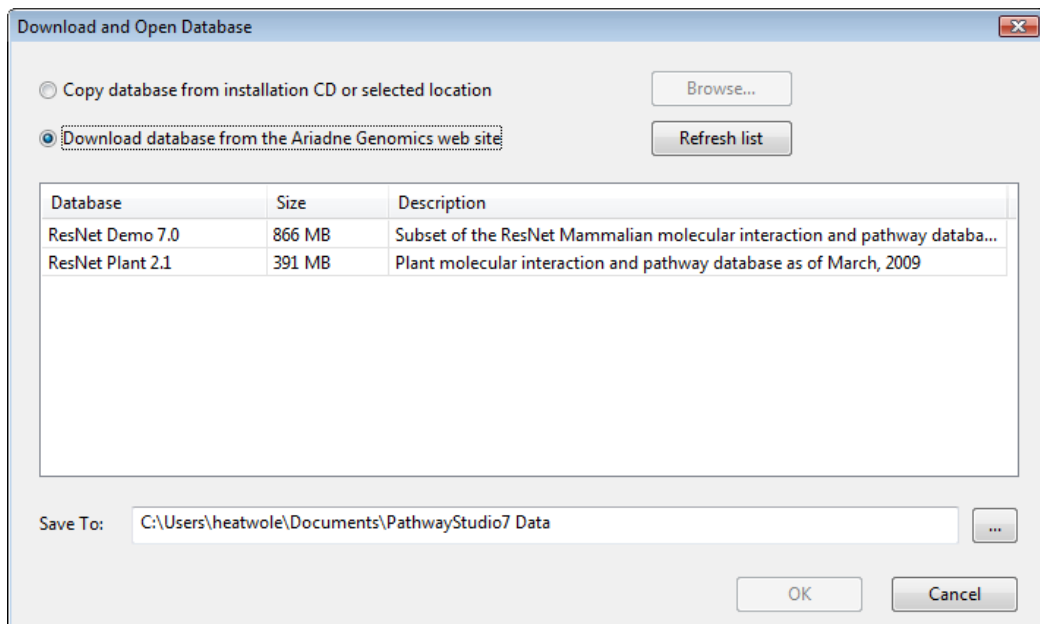
To open a remote (server based) database go to the Information Pane and select: Database > Open Remote. An Open Remote Database dialog appears. The first time you access the server database you must first register it. Select Register Remote Server button and add the appropriate information in the dialog. Obtain the server URL from your system administrator. The Display Name is a “nickname” that you can use to make identification of a specific server easier.



Downloading a database

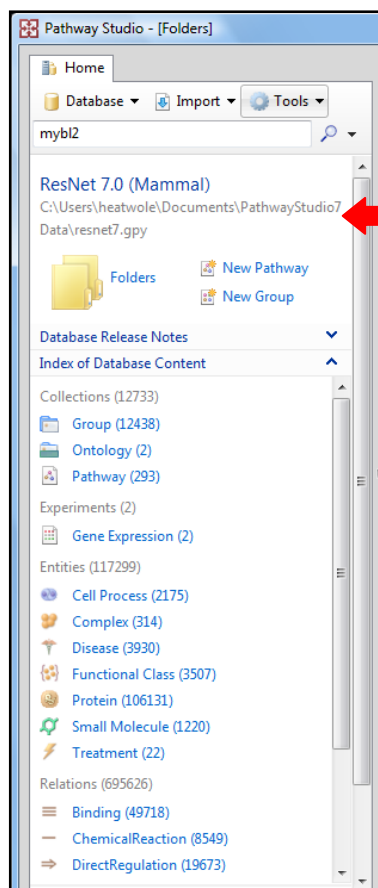
The local ResNet 7 Mammalian database is installed from a DVD (as it is quite large). Smaller databases, such as the ResNet Plant 2.1 database, are available for download by selecting Database > Download in the Information Pane.

Note: the ResNet Mammalian 7.0 Demo database contains only about 20% of the relationships that exist in the full ResNet Mammalian 7.0 database. **Be careful not to use the Demo database for your research projects!**



Information Pane

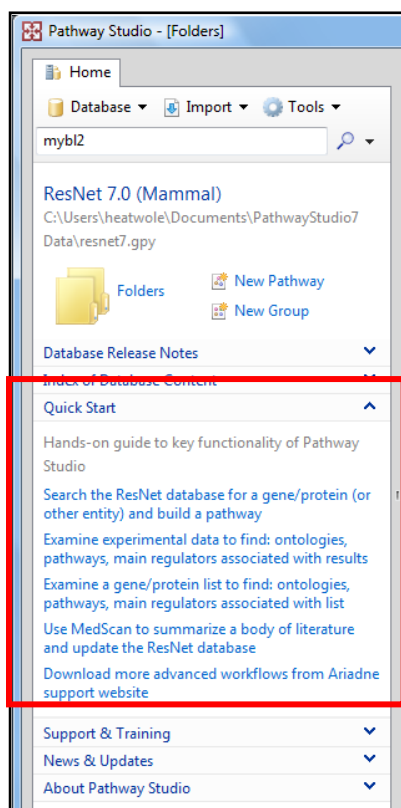
The Information Pane contains subheadings that provide useful links and information.



The header in the Information Pane indicates the database that is currently active in the Pathway Studio application.

Opening the “Index of the Database Content” tab reveals a summary of the information contained in the current database, such as totals for entity and relationship types, numbers of groups, pathways and experiments as well as ontologies.

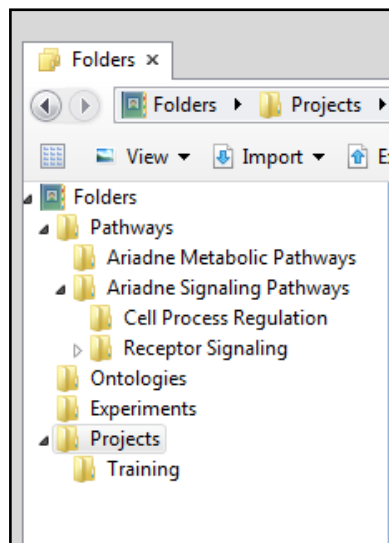
As you work with the ResNet database and utilize the MedScan Reader tool, the total number of objects in your database will change.



The Quick Start Guide is a good place for a new user of Pathway Studio to discover and utilize common workflows. Select the links for the desired analysis and have the Guide walk you through the steps.

Links to the Support site are found under Support & Training. From there, you can find more information about "What's New in Pathway Studio 7" as well as search an FAQ database and submit questions to Technical Support.

Folders View



The Folders View allows you to browse through the complete folder tree to find specific files. Pathways and Ontologies are provided in their respective folders. In the Experiments Projects folder you can create folders and subfolders as needed, and you now have the flexibility to save related items (group, pathway and experiment) in the same folder. This allows you easily organize your work.

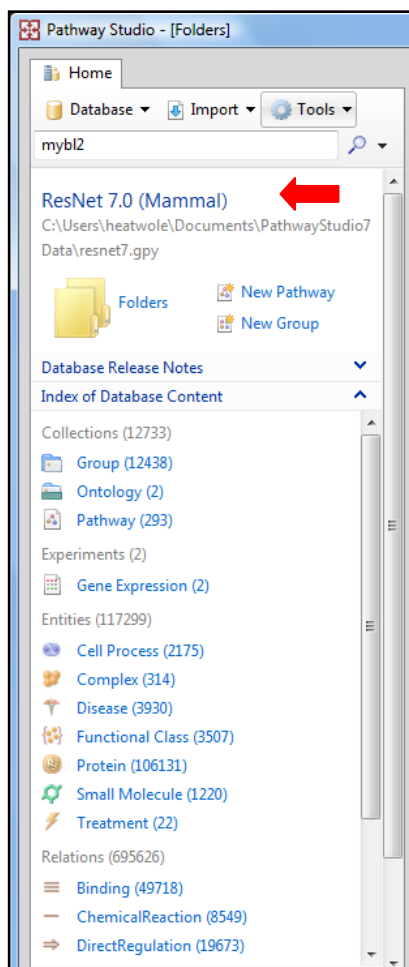
Note: When working with a local database you have complete freedom to manipulate the folders. This includes the ability to (accidentally) delete reference pathways and ontology groups. It is advisable to keep all your work in the Experiments and Projects folders.

● Exercise One: Open the local ResNet Mammalian Database (Desktop Users Only)

Begin Exercise:

Objective: To learn how to work with a specific database with Pathway Studio.

Let's begin using Pathway Studio by opening the local ResNet Mammalian Database. (To do this, the database must first be installed on your local computer, either from an Ariadne supplied DVD for Desktop users or by your local computer administrator for Enterprise users.)



1. In the Information Pane, go to Database > Open Local.
2. Select resnet7.gpy to open the ResNet 7 Mammalian database.
3. In the top of the Information Pane see "ResNet 7.0(Mammal)"
4. Select "Index of Database Content" to see a summary of the content in the local database: types and numbers of collections, experiments, entities and relationships.

Note: The numbers in your local ResNet database will change as you use Pathway Studio and MedScan Reader to add additional data to the local database. In addition, remember to apply the quarterly updates when they become available to keep your ResNet database current.

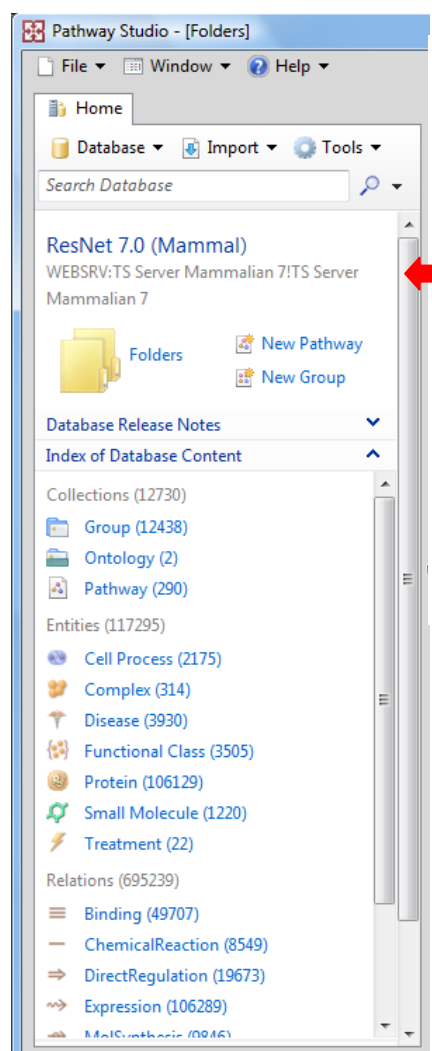
End Exercise: Open the local ResNet Mammalian Database

● **Exercise Two: Open the server-hosted ResNet Mammalian Database (PC-Client, Enterprise only)**

Begin Exercise:

Objective: To learn how to work with a server-hosted database with Pathway Studio.

Let's begin using Pathway Studio by opening the server-hosted ResNet Mammalian Database. (To do this, you first must obtain the address of the server where the database hosted.)



1. In the Information Pane, go to Database > Open Remote.
2. If you have not already done so, select "Register Remote Server"
3. Input the server URL obtained from your administrator.
4. Input a display name for the server, then select "OK".
5. In the Registered Databases view, select the display name of the server, then select "Open."
6. Select "Index of Database Content" to see a summary of the content in the database: types and numbers of collections, experiments, entities and relationships.

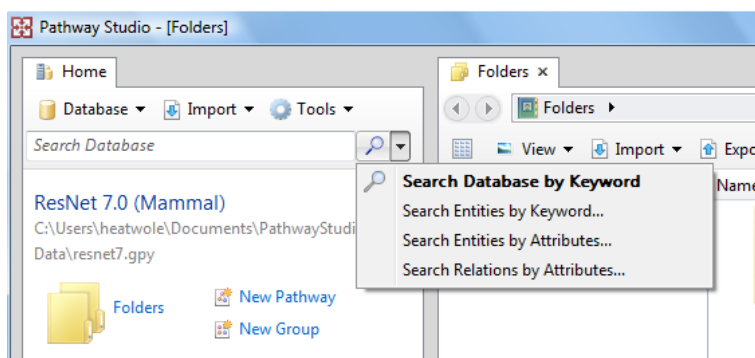
End Exercise: Open the server-hosted ResNet Mammalian Database (PC-Client, Enterprise only)

Section 3: Building Pathways

Pathway Studio provides the tools to build pathways (networks) from the entities and relationships found in the ResNet database. There are many ways to start to build pathways and only some examples will be shown here. You can start by searching the database to find an entity of interest and then query for relationships to that entity. Alternatively you can import a list of entities and query the relationships they have with each other and what other relationships in ResNet connect to them.

Search ResNet for an Entity

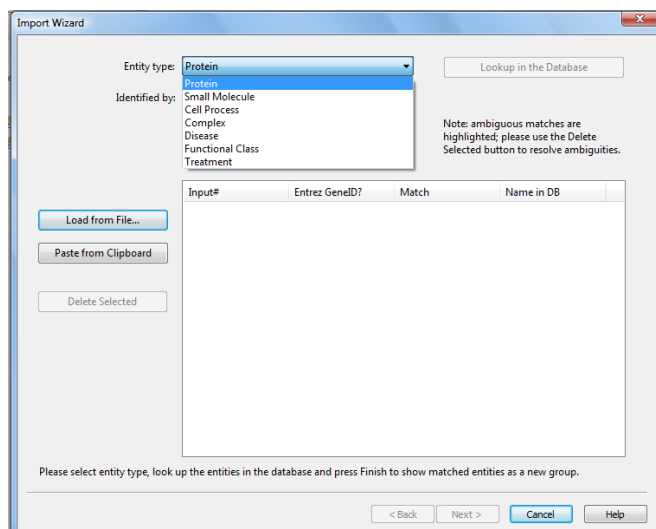
The Quick Search Box in the Information Pane allows for a keyword search of the entire database. Simply type the keyword in the box and select the magnifying glass to the right of the box. To define more specific searches, utilize the options in the drop-down menu to the right of the magnifying glass. Note: you can search for relationships as well as entities.



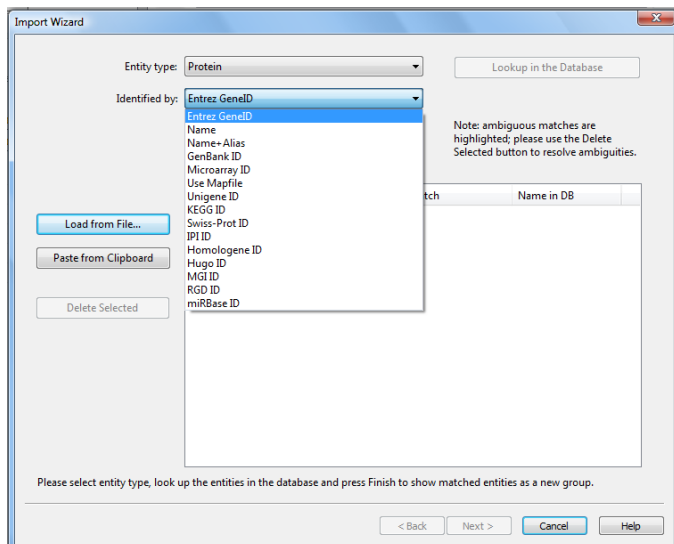
The results of the search will appear in the bottom list pane.

Import a list of Entities

You can import a list of entities by selecting the Import > Gene List. The Import Wizard appears

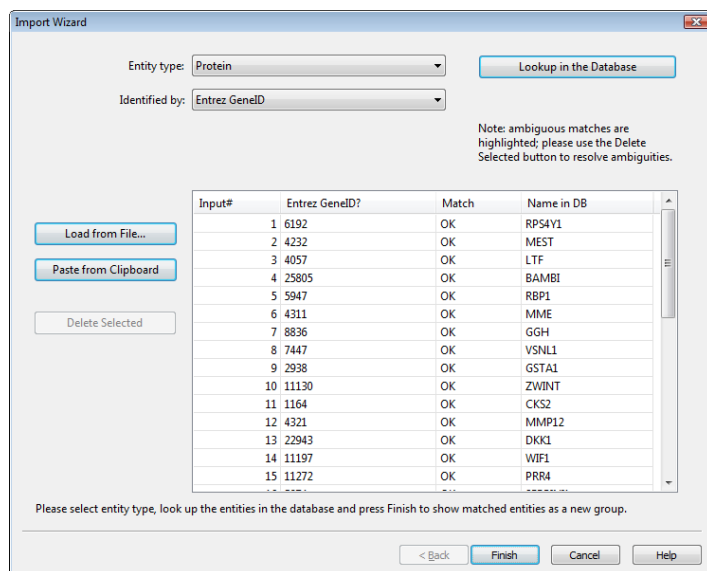


Note: Select this option for importing any entity list of IDs including: proteins/genes, small molecules, cell processes, complexes, diseases, functional classes and treatments.



Once the type of entity to import is selected, a second menu is available to identify the specific ID types.

You can simply copy a list of IDs from an Excel spreadsheet or txt file to your clipboard, and then select Paste from Clipboard. The list will appear in the window. Select Lookup in the Database to find the entities.

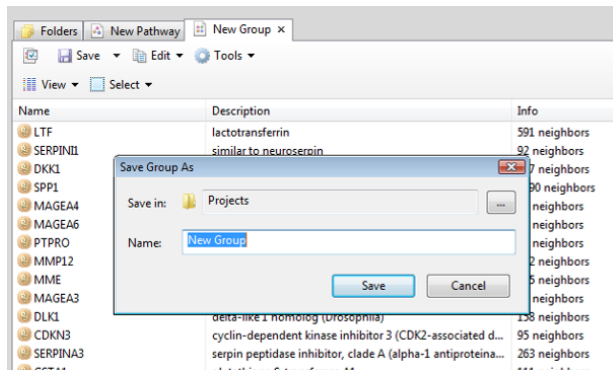


Note: if the imported ID list produces some ambiguous mapping, you have the opportunity to manually select the desired mapping. Use “Delete Selected” to remove any unwanted mapping.

Choose to save your imported list as a group.

Import Network Data

See **Appendix C: How to Import Network Data** for more information about bringing external network data into your ResNet database.



Once the entities that are the starting point in pathway building have been either identified by searching ResNet or imported and saved as a Group, you are ready to start to build a pathway.

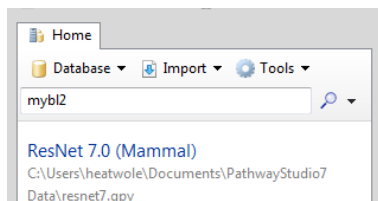
● Exercise Three: Search ResNet for Entities and Relationships and Import a Protein List

Begin Exercise:

Objective: To learn how to find entities and relationships contained in the ResNet database either by searching the database or by importing a list of entities.

Let's begin by searching the ResNet Mammalian 7.0 database for a specific protein:

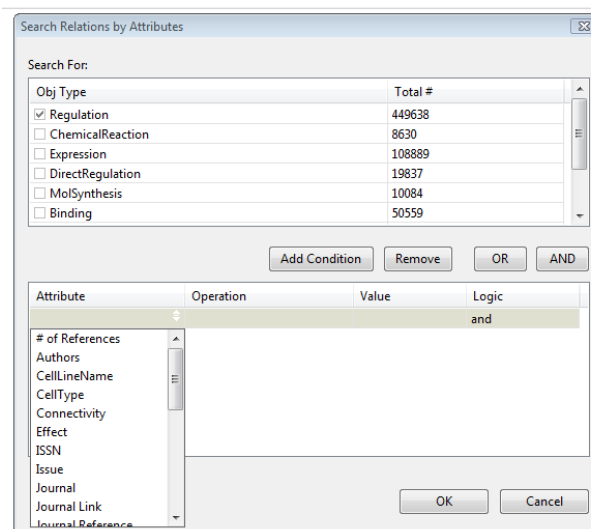
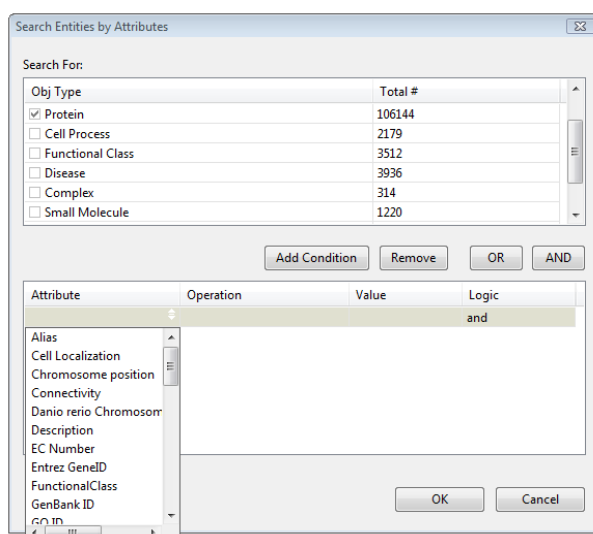
1. Select the Search Database by Keyword option in the Information Pane and type in the protein name: MYBL2



The results of your search are found in the List Pane at the bottom of the screen.

Name	Description	Entrez GeneID	Connectivity	#
MYBL2	v-myb myeloblastosis viral oncogene homolog (avian)-like 2	4605, 17865, 296344	147	1

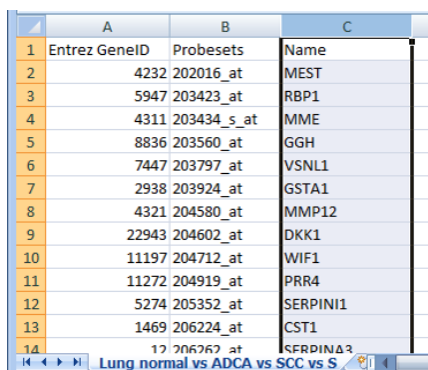
2. Examine the more specific search options:
 - a. Search Entities by Keyword – can specify specific entity types for your search
 - b. Search Entities by Attribute – use this option to search specific annotation fields for a selected entity type (left figure below)
 - c. Search Relations by Attribute – use this option to search specific annotation fields for a selected relationship type (right figure below)



Leave the tab in the List Pane containing the search results for MYBL2 open for now.

Now let's import of list of entity IDs. Use the MS Excel file provided with this training manual.

3. Open the MS Excel file that accompanies this manual (SCLC genes) and copy all the column contents with the header "name" to the clipboard.



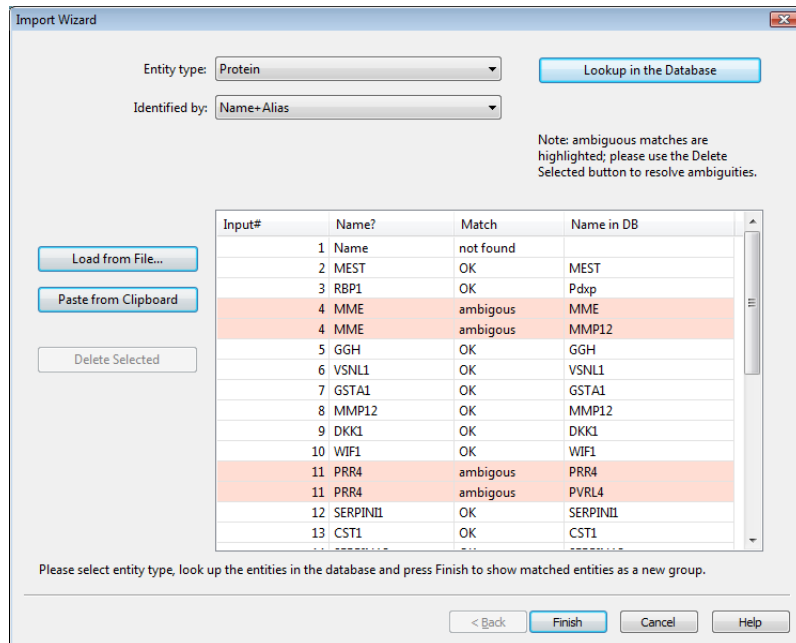
	A	B	C
1	Entrez GeneID	Probesets	Name
2	4232	202016_at	MEST
3	5947	203423_at	RBP1
4	4311	203434_s_at	MME
5	8836	203560_at	GGH
6	7447	203797_at	VSNL1
7	2938	203924_at	GSTA1
8	4321	204580_at	MMP12
9	22943	204602_at	DKK1
10	11197	204712_at	WIF1
11	11272	204919_at	PRR4
12	5274	205352_at	SERPINI1
13	1469	206224_at	CST1
14	17	206767_at	KERDINAR

4. Go to Import > Gene List. The Import Wizard is displayed.

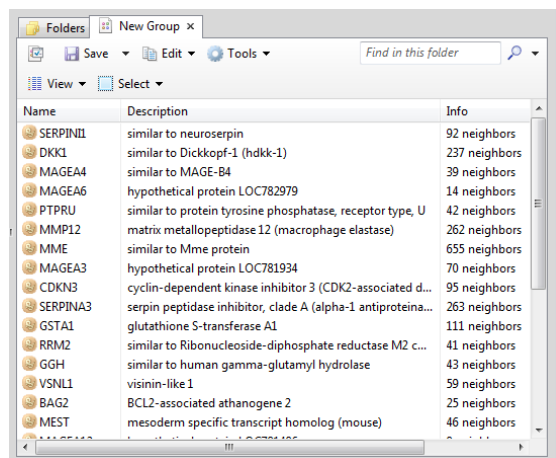
Note: Use Import > Gene List to import any list of entities: proteins (genes), small molecules, cell objects, cell processes, complexes, diseases, functional classes and treatments. The "Identified by" ID list will reflect the selected entity type.

Note: Recall that in Pathway Studio the gene and the product of the gene (protein) are merged into one entity.

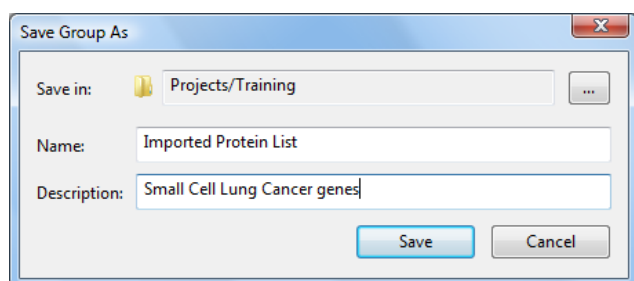
5. Select "Paste from Clipboard" to copy the list of protein names into the Wizard.
6. For "Entity type" select: Protein. For "Identified by" select: "Name+alias".
7. Select "Lookup in the Database" to identify the imported entities in the ResNet database.
8. "Not found" in the Match column indicates the entity was not found in (note that "Name" from the column header in the MS Excel sheet is indicated as "not found").



9. Items with ambiguous mapping are highlighted.
10. Remove undesired mapping by highlighting the rows and click “Delete Selected” button to remove.
11. When all undesired mapping is removed, select “Finish” to import the list. The imported list will appear in a New Group window.



12. Choose Save to save the group (provide a destination folder and name for the group – create a new folder if desired). Description information can be appended to the saved group if desired.



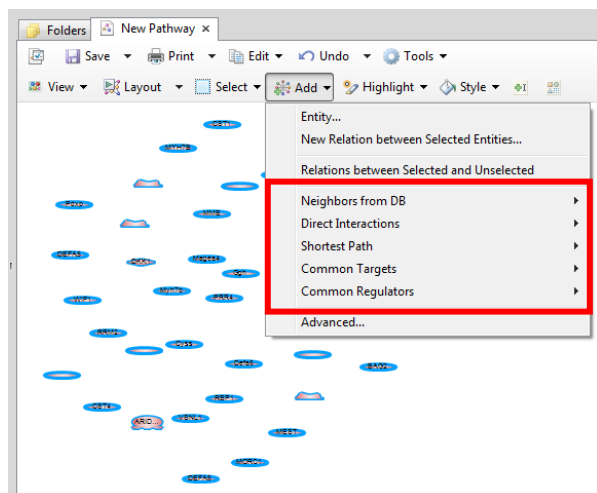
End Exercise: Search ResNet for Entities and Relationships and Import a Protein List

Build Pathway Tools in Pathway Studio

Pathway Studio provides both quick short cut menus for pathway building as well as an advanced menu that allows you to define more specific network options.

Build Pathway Tool – Quick Menus

With the entities in a new Pathway View, select the Add tool to see the short cut menu for building pathways.



Each of these short cut menu options has easy-to-interpret submenus. Select the appropriate option for the desired pathway.

Note: If an option is grayed out, it is not an appropriate option for the number of entities selected.

The Relationship types and Entity types included in the Shortcut menu options:

		Type(s) of Relationships	Type(s) of Entities
Neighbors from DB	Expression Targets	Expression, PromoterBinding	Proteins, Complexes, Functional Classes
	Expression Regulators	Expression, PromoterBinding	Proteins, Complexes, Functional Classes
	Physical Interactions	Direct Regulation, Binding	Proteins, Complexes, Functional Classes
	Protein Modification Targets	Prot Modification	Proteins, Complexes, Functional Classes
	Protein Modification	Prot Modification	Proteins, Complexes, Functional Classes
	Enzymes		
Direct Interactions	All	All	All
	Physical Interactions	Direct Regulation, Binding	Proteins, Complexes, Functional Classes
	Expression Regulation	Expression, PromoterBinding	Proteins, Complexes, Functional Classes
	Protein Modification	Prot Modification	Proteins, Complexes, Functional Classes
Shortest Path	Physical Interactions	Direct Regulation, Binding	Proteins, Complexes, Functional Classes
	Expression Regulation	Expression, PromoterBinding	Proteins, Complexes, Functional Classes
	Protein Modification	Prot Modification	Proteins, Complexes, Functional Classes
Common Targets	Expression Targets	Expression, PromoterBinding	Proteins, Complexes, Functional Classes
	Physical Interactions	Direct Regulation, Binding	Proteins, Complexes, Functional Classes
	Protein Modification Targets	Prot Modification	Proteins, Complexes, Functional Classes
	All	All	All
Common Regulators	Expression Regulators	Expression, PromoterBinding	Proteins, Complexes, Functional Classes
	Physical Interactions	Direct Regulation, Binding	Proteins, Complexes, Functional Classes
	Protein Modification	Prot Modification	Proteins, Complexes, Functional Classes
	Enzymes		

Build Pathway Tool – Advanced Menu Desktop

The Advanced Menu option (Add > Advanced) opens a wizard that provides you to options for more specifically defining the pathway building filter options. After selecting the entities to including in your build pathway algorithm, the four wizard windows allow you to specify: a.) algorithm type for pathway building, b.) select the directionality of the resultant relationships, c.) the entities to be included in the pathway and d.) the relationships to be included in the new pathway.

Select Algorithm Type

☒ Add neighbors
of expansion:

☐ Add direct interactions

☐ Add shortest path

☐ Add common targets

☐ Add common regulators

☐ Add neighbors from group
Group: ...

< Back Next > Cancel Help

a.

Select Directions for Each Entity

☐ Upstream
☐ Downstream
☒ Both

Name	Type	Description	Direction
SERPINE1	Protein	similar to neuroserpin	<-->
DIK1	Protein	similar to Dickkopf-1 (hdkk-1)	<-->
MAGEA4	Protein	similar to MAGE-B4	<-->
MAGEA6	Protein	hypothetical protein LOC782979	<-->
PTPRU	Protein	similar to protein tyrosine phosphatase, receptor type, U	<-->
MMP12	Protein	matrix metalloproteinase 12 (macrophage elastase)	<-->
MME	Protein	similar to Mme protein	<-->
MAGEA3	Protein	hypothetical protein LOC781934	<-->
CDKN3	Protein	cyclin-dependent kinase inhibitor 3 (CDK2-associated dual...	<-->
SERPINA3	Protein	serpin peptidase inhibitor, clade A (alpha-1 antiproteinas...	<-->
GSTA1	Protein	glutathione S-transferase A1	<-->
PRDM7	Protein	similar to Ribonucleoside-diphosphate reductase M2 chain	<-->

< Back Next > Cancel Help

b.

Set Filter Parameters

Entity Type	Use Filter	Operation	Value
<input checked="" type="checkbox"/> Protein	ALL	⊕	
<input checked="" type="checkbox"/> Cell Process	ALL	⊕	
<input checked="" type="checkbox"/> Functional Class	ALL	⊕	
<input checked="" type="checkbox"/> Disease	ALL	⊕	
<input checked="" type="checkbox"/> Complex	ALL	⊕	
<input checked="" type="checkbox"/> Small Molecule	ALL	⊕	
<input checked="" type="checkbox"/> Treatment	ALL	⊕	

Check all Clear all

< Back Next > Cancel Help

c.

Set Filter Parameters

Relation Type	Use Filter	Operation	Value
<input checked="" type="checkbox"/> Regulation	ALL	⊕	
<input checked="" type="checkbox"/> ChemicalReaction	ALL	⊕	
<input checked="" type="checkbox"/> Expression	ALL	⊕	
<input checked="" type="checkbox"/> DirectRegulation	ALL	⊕	
<input checked="" type="checkbox"/> MolSynthesis	ALL	⊕	
<input checked="" type="checkbox"/> Binding	ALL	⊕	
<input checked="" type="checkbox"/> PromoterBinding	ALL	⊕	
<input checked="" type="checkbox"/> ProtModification	ALL	⊕	
<input checked="" type="checkbox"/> MolTransport	ALL	⊕	

Check all Clear all

< Back Next > Cancel Help

d.

Build Pathway Tool – Advanced Menu Enterprise

The Advanced Menu option (Add > Advanced) opens a wizard that provides you to options for more specifically defining the pathway building filter options. After selecting the entities to including in your build pathway algorithm, the four wizard windows allow you to specify: a.) algorithm type for pathway building, b.) select the directionality of the resultant relationships, c.) the entities to be included in the pathway and d.) the relationships to be included in the new pathway.

Points unique to the Enterprise Advanced Build Pathway Wizard:

1. The “Add Common Targets” and “Add Common Regulators” allows for user-defined thresholds of number of entities in the target group that have a shared regulator or shared target.
2. Relationship types available may differ depending on what additional information is in your database.

a.

Name	Type	Description	Direction
SERPINE1	Protein	similar to neuroserpin	<-->
DKK1	Protein	similar to Dickkopf-1 (hdkk-1)	<-->
MAGEA4	Protein	similar to MAGE-B4	<-->
MAGEA6	Protein	hypothetical protein LOC782979	<-->
PTPRU	Protein	similar to protein tyrosine phosphatase, receptor type, U	<-->
MMP12	Protein	matrix metalloproteinase 12 (macrophage elastase)	<-->
MME	Protein	similar to Mme protein	<-->
MAGEA3	Protein	hypothetical protein LOC781934	<-->
CDKN3	Protein	cyclin-dependent kinase inhibitor 3 (CDK2-associated dual...	<-->
SERPINA3	Protein	serpin peptidase inhibitor, clade A (alpha-1 antiproteinas...	<-->
GSTA1	Protein	glutathione S-transferase A1	<-->
BDM2	Protein	similar to Dihydropyridine-dihydrogenase reductase M2 chain	<-->

b.

Entity Type	Use Filter	Operation	Value
<input checked="" type="checkbox"/> Protein	ALL		
<input checked="" type="checkbox"/> Cell Process	ALL		
<input checked="" type="checkbox"/> Functional Class	ALL		
<input checked="" type="checkbox"/> Disease	ALL		
<input checked="" type="checkbox"/> Complex	ALL		
<input checked="" type="checkbox"/> Small Molecule	ALL		
<input checked="" type="checkbox"/> Treatment	ALL		

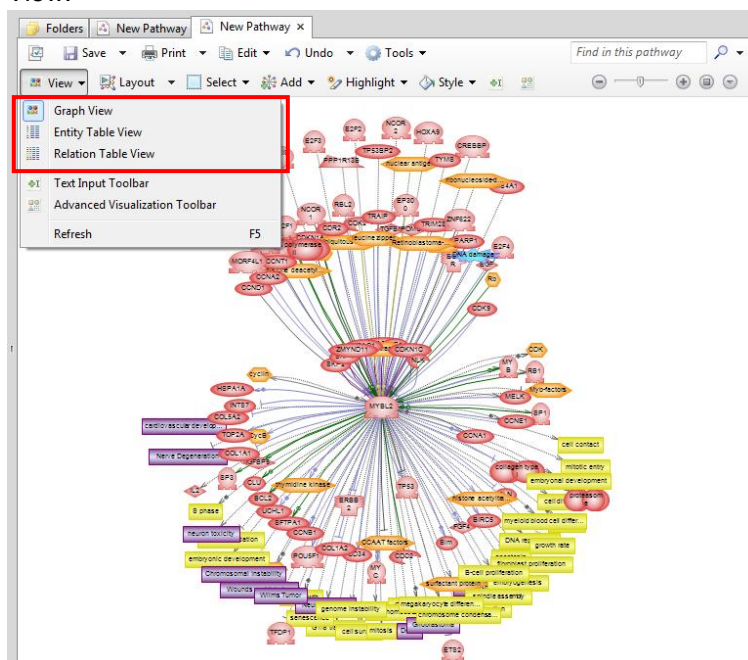
c.

Relation Type	Use Filter	Operation	Value
<input checked="" type="checkbox"/> Regulation	ALL		
<input checked="" type="checkbox"/> ChemicalReaction	ALL		
<input checked="" type="checkbox"/> Expression	ALL		
<input checked="" type="checkbox"/> DirectRegulation	ALL		
<input checked="" type="checkbox"/> MolSynthesis	ALL		
<input checked="" type="checkbox"/> Binding	ALL		
<input checked="" type="checkbox"/> PromoterBinding	ALL		
<input checked="" type="checkbox"/> ProtModification	ALL		
<input checked="" type="checkbox"/> MolTransport	ALL		

d.

Graph View/ Entity Table view / Relation Table view

There are three options for viewing resultant pathways: Graph View, Entity Table View and Relation Table View.



Customizing tables, filtering by significance

The Entity Table view and Relation Table view provide customizable tables of summary information of the pathway. Any information deleted from either of these tables is also deleted from the Graph View. Customize the columns of information by selecting Tools > Customize Columns.

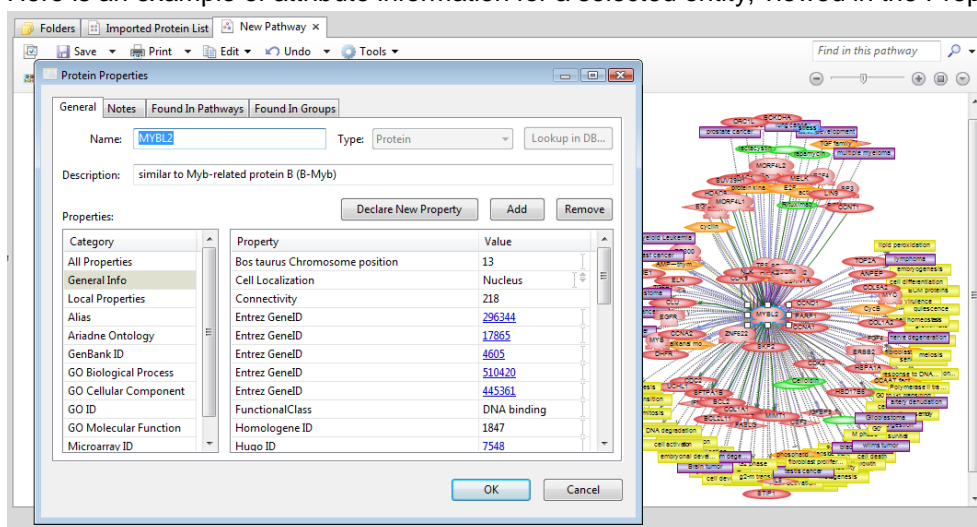
Relation	Type	Sentence	MedLine Reference	Connect...	# of References	Owner	Journal Link	Journal Reference
MYBL2 ---> a...	Regulation	The MYBL2 gene is strongly induced at th...	11264176.10173, 15548681.10116	2	30	Public	http://www.bloodjournal.org/content/119/10/p1483	v97 i7 p2091, v64 i22 p8167, v23...
MYBL2 ---> a...	Regulation	Choose the Columns to be Displayed			21	Public	http://jcs.biologists.org/content/119/10/p1483	v119 i8 p1483, v19 i6 p719, v280...
MYBL2 ---> c...	Regulation				66	Public	http://cancerres.aacrjournals.org/content/65/12/p439	v65 i2 p439, v65 i21 p9751, v7 i4...
MYBL2 ---> c...	Regulation				2	Public	http://www.bloodjournal.org/content/105/10/p3855	v105 i10 p3855, v280 i16 p15628...
MYBL2 ---> c...	Regulation				29	Public	http://www.current-biology.org/content/7/14/p26	v7 i4 p26, v274 i51 p36741, v105...
MYBL2 ---> c...	Regulation				57	Public	http://jcs.biologists.org/content/119/8/p1483	v119 i8 p1483, v119 i2 p6, v18 i2...
MYBL2 ---> l...	Regulation				30	Public	http://www.cancerres.aacrjournals.org/content/65/12/p439	v1 i4 p7, v59 i14 p3365, v62 i8 p2...
MYBL2 ---> v...	Regulation				1	Public	http://www.bloodjournal.org/content/105/10/p3855	v95 i12 p3900
MYBL2 ---> v...	Regulation				5	Public	http://www.cancerres.aacrjournals.org/content/65/12/p439	v1 i4 p7, v64 i7 p2561, v65 i7 p28...
MYBL2 ---> v...	Regulation				8	Public	http://cancerres.aacrjournals.org/content/65/12/p439	v59 i14 p3365, v275 i28 p21055, ...
MYBL2 ---> v...	Regulation				1	Public	http://www.bloodjournal.org/content/105/10/p3855	v96 i3 p1013
MYBL2 ---> v...	Regulation				5	Public	http://www.jbc.org/content/275/14/p10692	v275 i14 p10692, v97 i7 p2091, v...
MYBL2 ---> v...	Regulation				6	Public	http://jcs.biologists.org/content/119/8/p1483	v119 i8 p1483, v18 i23 p2837, v2...
MYBL2 ---> v...	Regulation				2	Public	http://cancerres.aacrjournals.org/content/65/12/p439	v60 i16 p4519
MYBL2 ---> v...	Regulation				13	Public	http://cancerres.aacrjournals.org/content/65/12/p439	v59 i14 p3365, v119 i8 p1483, v2...
MYBL2 ---> v...	Regulation				36	Public	http://www.genesdev.org/content/19/16/p719	v19 i6 p719, v278 i11 p9655, v11...
MYBL2 ---> v...	Regulation				8	Public	http://cancerres.aacrjournals.org/content/65/12/p439	v62 i15 p4499, v275 i28 p21055, ...
MYBL2 ---> v...	Regulation				2	Public	http://cancerres.aacrjournals.org/content/65/12/p439	v62 i15 p4499, v62 i23 p6803
MYBL2 ---> n...	Regulation	Raschella Expression of Insulin-like Grow...	9408744.12121, 11134182.10669	2	10	Public	http://edrv.endojournals.org/content/18/6/p801	v18 i6 p801, v107 i1 p73, v275 i2...
MYBL2 ---> n...	Regulation	Several genes novel to testicular tumorig...	11956097.10013	2	1	Public	http://cancerres.aacrjournals.org/content/65/12/p439	v62 i8 p2359
MYBL2 ---> n...	Regulation	The defects caused by reduced B-Myb le...	16551698.10038	2	1	Public	http://jcs.biologists.org/content/119/8/p1483	v119 i8 p1483

Note: In the Relationship View table you can see the number of references in the database that support an individual relationship. High numbers of references can be used as a measure of confidence for a relationship. Lower numbers of references can indicate newly identified relationships or potential false positives from MedScan. The reference sentence is available in order for you to determine the accuracy of the interpretation by MedScan. Undesired (false positive) relationships can be manually deleted from the database. See **Appendix D: Deleting Entities and Relations from a Local Database**.

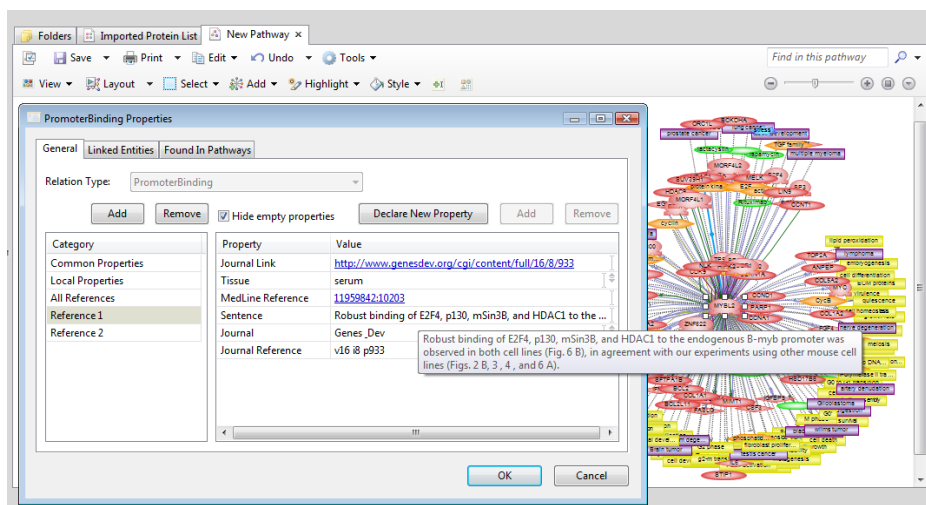
Viewing Details about Entities/Relationships

Detailed information about entities and relationships displayed in the Graph View can be seen by a.) mouse over the object or b.) double click to open the properties dialog.

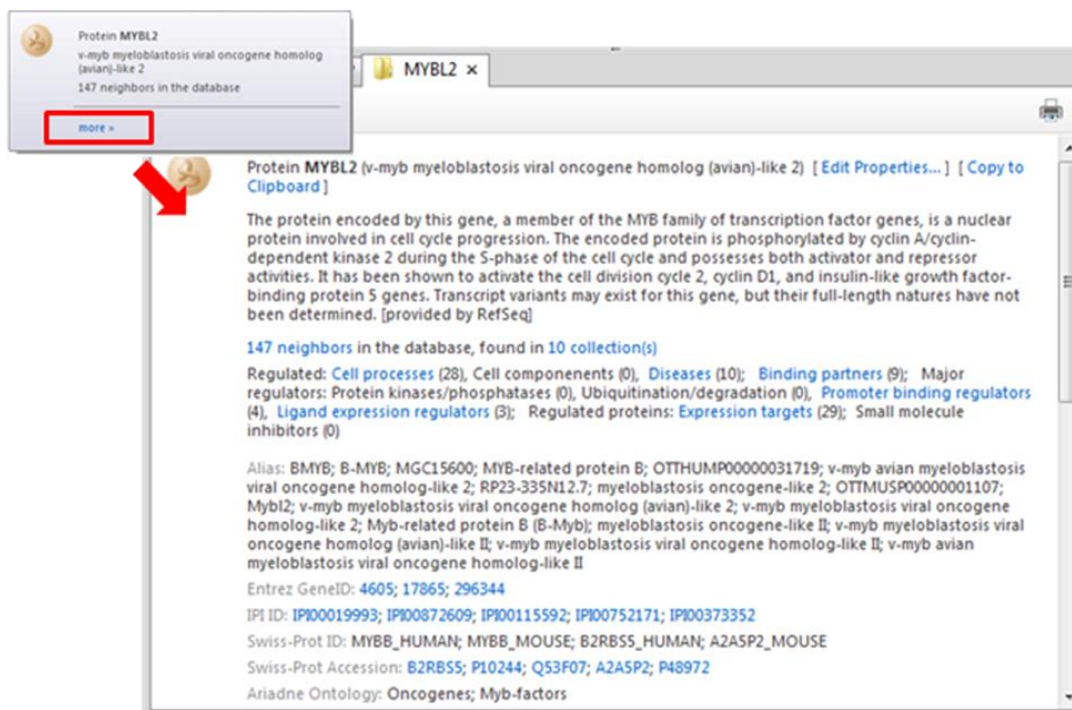
Here is an example of attribute information for a selected entity, viewed in the Properties dialog:



The properties dialog for the relationship provides reference information that supports the relationship. You can view the sentences identified by MedScan that support the relationship.



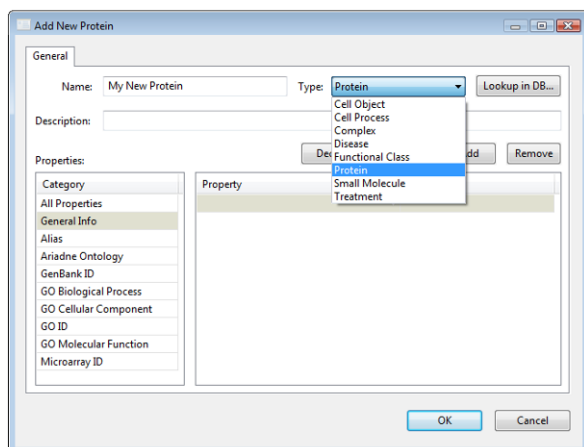
In addition annotation for entities, relations, pathways and groups can be viewed in a Preview Pane, an browser format for displaying annotation. To open the Preview Pane select "more" from the mouse over pop-up window, or select "Preview" from the right-click menu.



Creating new Entities and Relationships

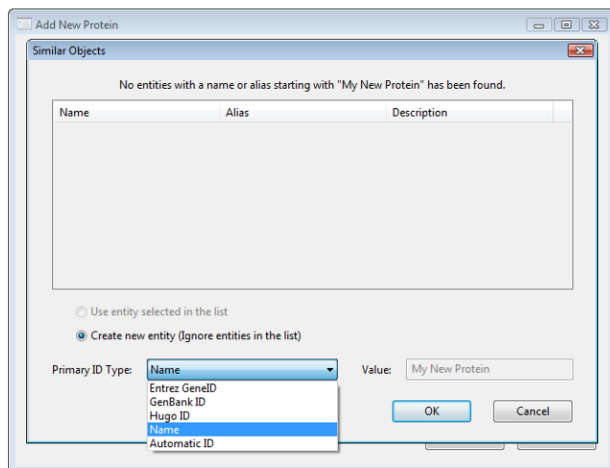
You can create a new entity or a new relationship between entities by using the Add > Entity or Add > New Relation between selected entities options.

To add a new Entity, provide appropriate information in the Add New Protein dialog:



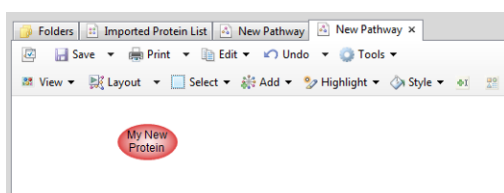
Note: When you attempt to add a new entity to ResNet, Pathway Studio will first check to see if an entity with the given name already exists in the database. It is advisable not to create a new entity with the same name or identifier as an existing entity.

Use the Add New Protein dialog for all entities you choose to add to ResNet, not just protein entities.



If your identifier is unique, complete the dialog by providing the type of primary identifier (in the example shown “name”) and select OK. This will create this new entity in ResNet.

Newly created entity:



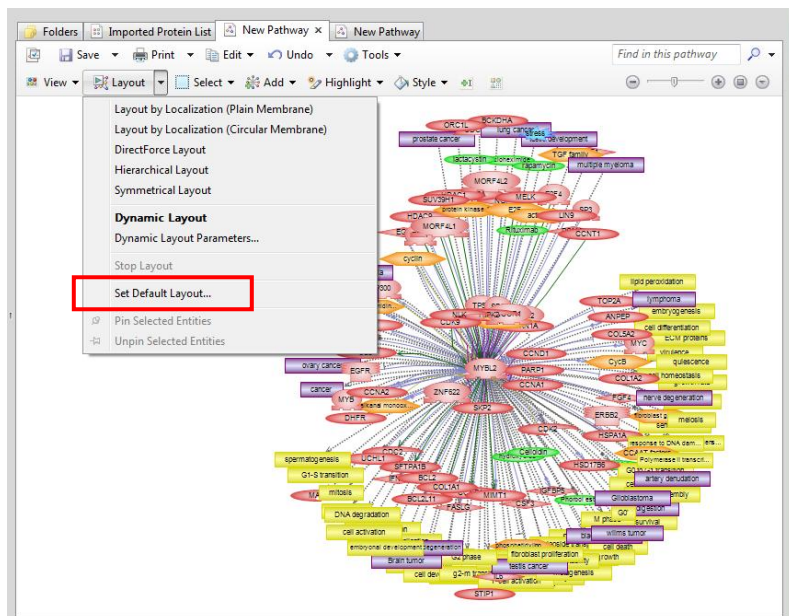
You can permanently delete an entity or relationship from ResNet by selecting it in a Graph View, then when the mouse is in the white space of the graph, right-click and select “Delete Selected Entities/Relations from the Database.”

Customization of Pathway Layouts

There are multiple layout options for displaying a pathway. (Please see the User’s Manual for a description of how each layout is calculated.)

The Layout by Localization options utilizes Gene Ontology cellular localization assignments when placing the entity with respect to cellular objects. Be aware that entities can have more than one legitimate localization assignment within Gene Ontology (an example would be a protein that shuttles between the nucleus and cytoplasm). In this case the layout displays only one of the localization assignments. You can change the assigned localization of an entity by selecting the entity, right-click, choose “Localization in Pathway” and choose the desired cellular localization.

You can change the default layout view by choosing Layout > Set Default Layout.



● Exercise Four: Building Pathways

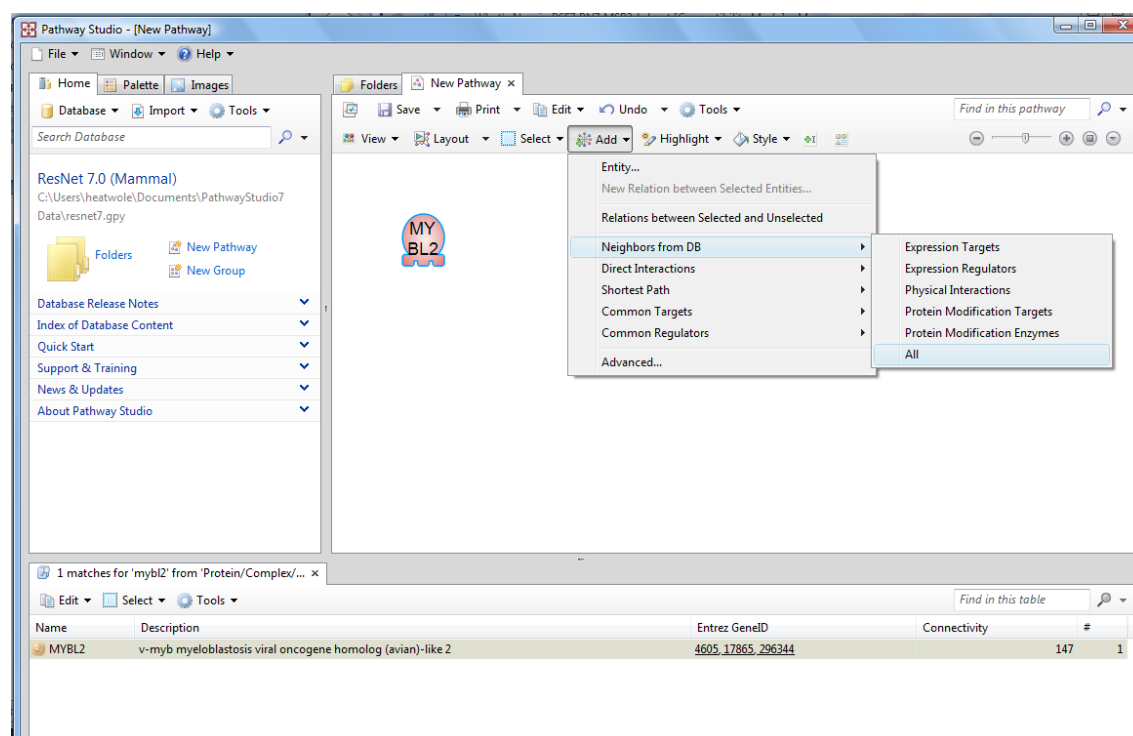
Begin Exercise:

Objective: To build pathways using both the Quick Menu and the Advanced Menu option and to become familiar with basic functionality used for viewing and modifying pathways.

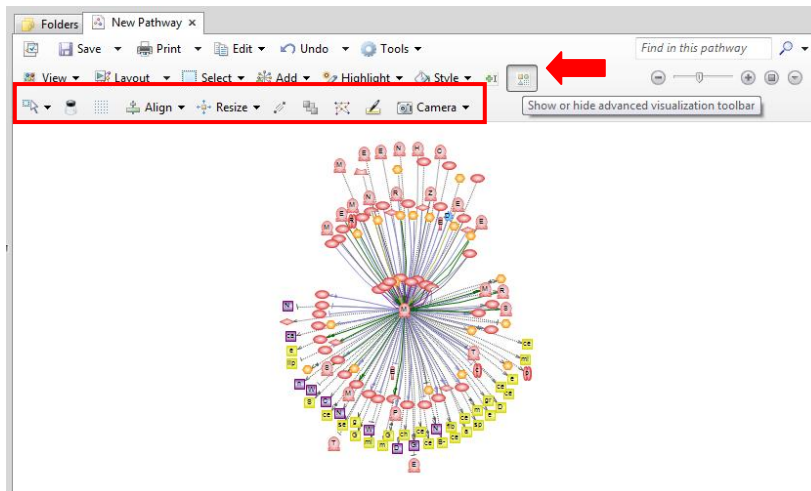
Build Pathway Tool – Quick Menu

Let's build a simple pathway displaying all connections in the ResNet 7.0 Mammalian database to our protein MYBL2.

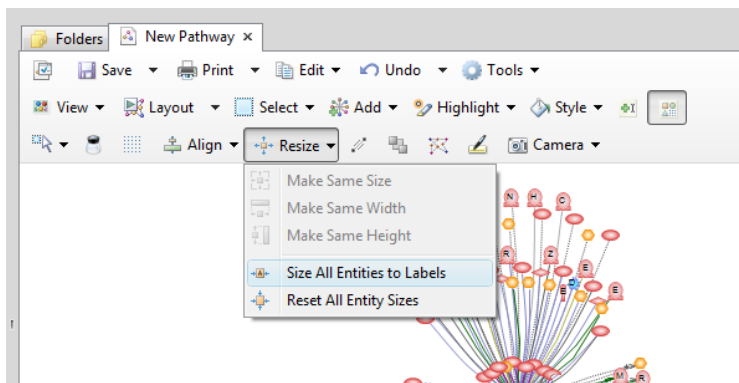
1. Select "New Pathway" from the Information Pane. A new pathway view will open to the right.
2. Click and drag the MYBL2 protein icon from the List Pane below and drag it to the new pathway window.
3. With the MYBL2 protein entity selected (highlighted in blue) go to the Add menu and select > Neighbors from DB > All. This will identify all entities connected to MYBL2 in the database.



4. The Results Preview window displays a summary of information about the resultant pathway. You can manually review the results here and manually delete any undesired entities or relationships at this time.



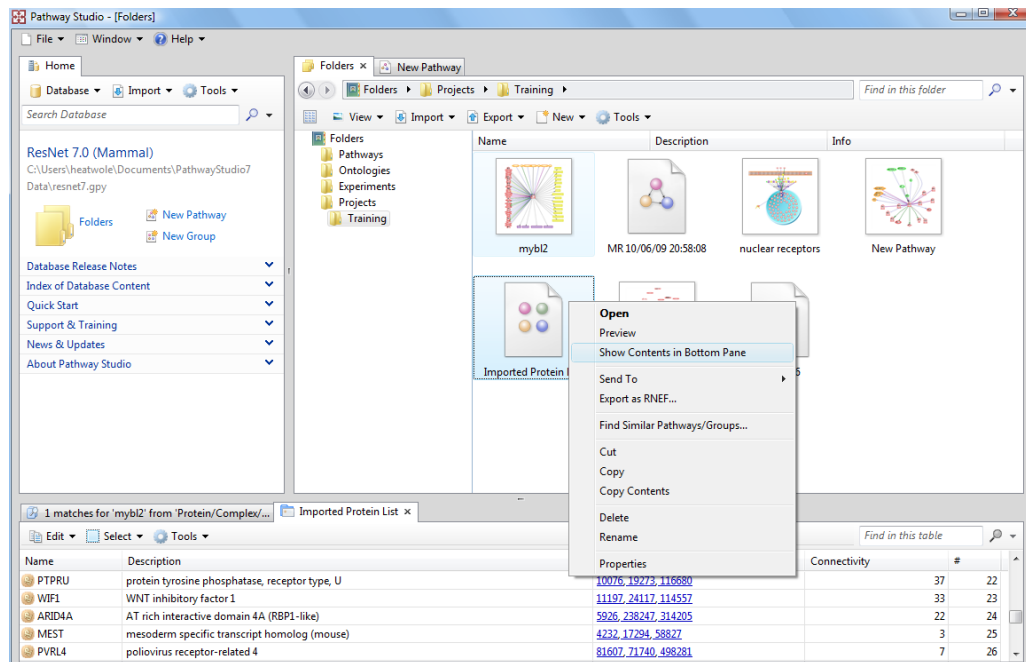
8. Select “Size All Entities to Labels” in the Resize menu to better visualize the names of the entities.



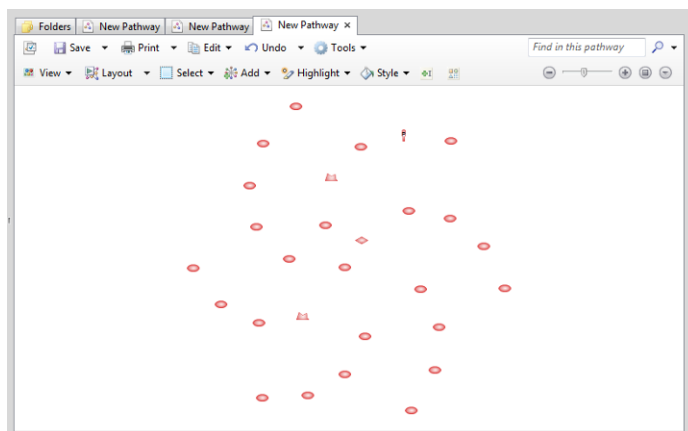
Build Pathway Tool – Advanced Menu

Now let’s use a different Build Pathway algorithm to build a pathway from the list of imported proteins.

1. Right-click the group icon from the saved list of imported proteins. Select “Show contents in Bottom Pane.” This will open the group list in the List Pane.



2. Select “New Pathway” in the Information Pane and select all the entities in the List Pane. Click and drag the entire list to the new pathway window. Select “Fit all Entities to Window” in the upper right to visualize all the entity icons.



3. Choose Select > All to select all entities. Then choose Advanced from the Add menu. The Advanced Build Pathway wizard will appear.

Let's find common transcriptional regulators shared by at least two members of our group.

4. In the Select Algorithm window, select “Add common regulators”. This will identify upstream regulators that shared by two or more members of the group. Recall that some relations have directionality, which allows us to define direction of influence.
5. Next select the type of entity or entities for the regulators. In this example, select proteins.

Select Algorithm Type

☐ Add neighbors
 # of expansion: 1

☐ Add direct interactions

☐ Add shortest path

☐ Add common targets

☒ Add common regulators

☐ Add neighbors from group
 Group:

< Back Next > Cancel Help

Set Filter Parameters

Entity Type	Use Filter	Operation	Value
<input checked="" type="checkbox"/> Protein	ALL		
<input type="checkbox"/> Cell Process			
<input type="checkbox"/> Functional Class			
<input type="checkbox"/> Disease			
<input type="checkbox"/> Complex			
<input type="checkbox"/> Small Molecule			
<input type="checkbox"/> Treatment			

Check all Clear all

< Back Next > Cancel Help

- Finally, select the types of relation the transcriptional regulators will have with the targets. In this example, we will select Expression and PromoterBinding. Although regulation could also be utilized, the category of Regulation contains by far the largest number of relationships and is less specific, so we will leave it out for now.

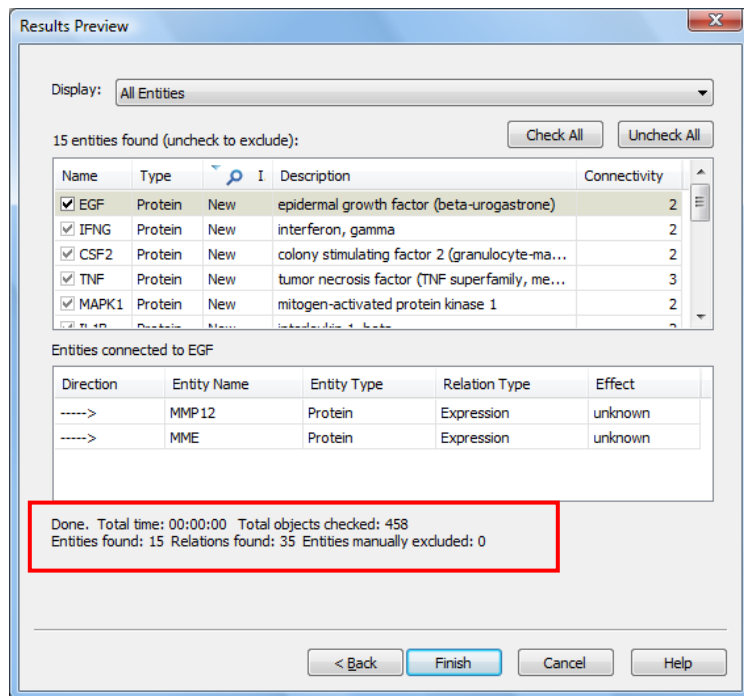
Set Filter Parameters

Relation Type	Use Filter	Operation	Value
<input type="checkbox"/> Regulation			
<input type="checkbox"/> ChemicalReaction			
<input checked="" type="checkbox"/> Expression	ALL		
<input type="checkbox"/> DirectRegulation			
<input type="checkbox"/> MolSynthesis			
<input type="checkbox"/> Binding			
<input checked="" type="checkbox"/> PromoterBinding	ALL		
<input type="checkbox"/> ProtModification			
<input type="checkbox"/> MolTransport			

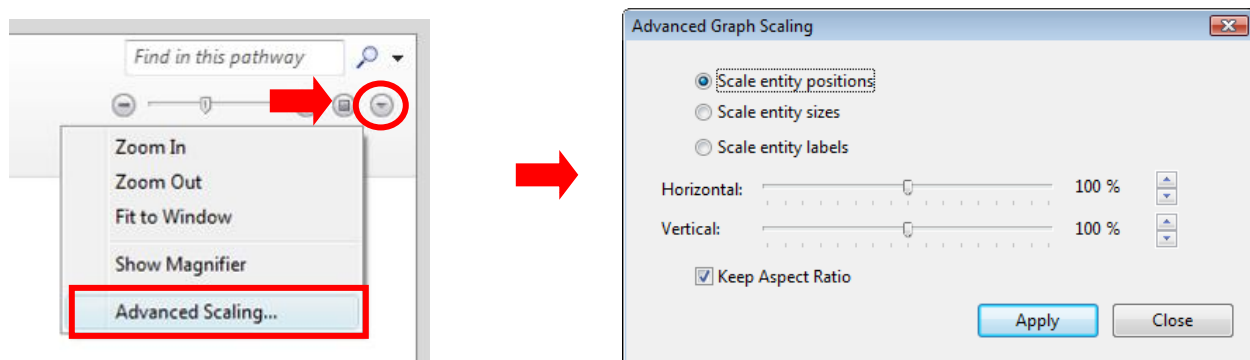
Check all Clear all

< Back Next > Cancel Help

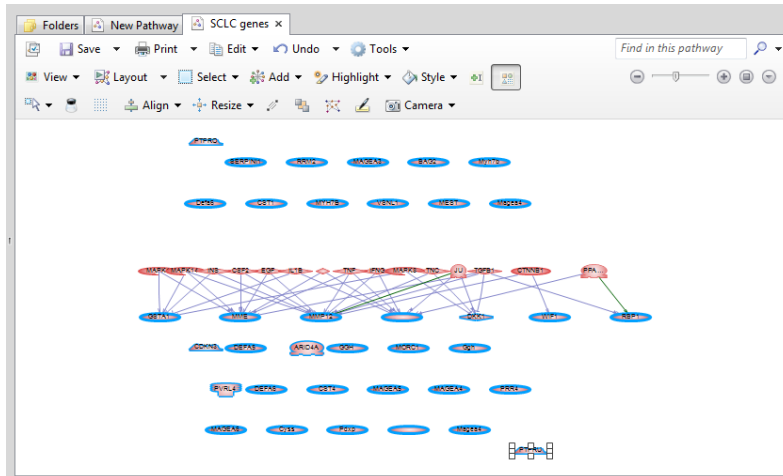
- When the algorithm is finished, select Finish to view the pathway. Note: at the bottom of the Results Preview window there is a summary of the resultant pathway. In the example below, 15 shared potential transcriptional regulators were identified and 35 relations between those regulators and the starting group of entities will be added to the pathway view.



8. Use the "Fit All Entities to Window" button to maximize the pathway view.
9. Select "Fit Text to Entities" in the Advanced Visualization tool bar.
10. As this pathway depicts upstream transcriptional regulators, select "Hierarchical Layout" from the layout menu. This will display the transcriptional regulators above the targets they regulate.
11. If needed, adjust the x and y-axis scale of the image by selecting Advanced Scaling from the menu in the far upper right. (Uncheck "Keep Aspect Ratio" before changing horizontal and vertical proportions.)



12. The new pathway contains the original group list and 15 new transcriptional regulators. Let's select the members of the original group. From the List Pane, choose Select > All, then Select > Mirror Selection to active pathway. The original members will be selected (blue line around selected entities). Alternatively you can Select > All, Edit > copy the list, then in the pathway view Select > Entities on Clipboard.

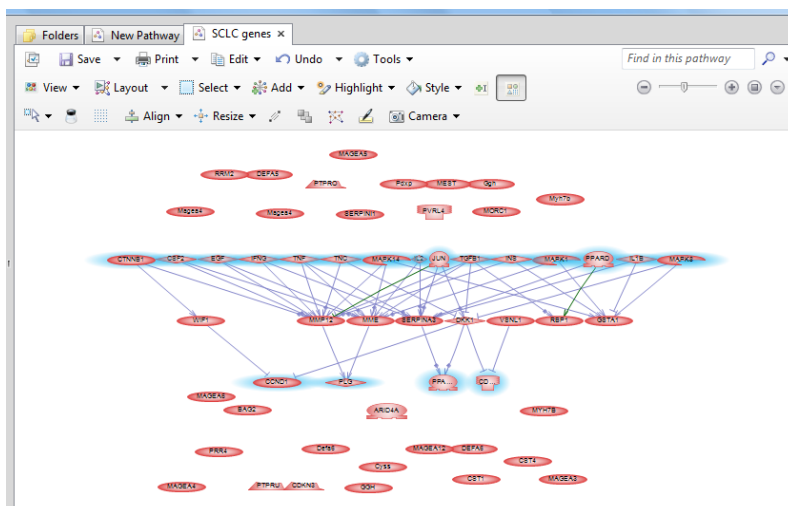


Now let's find shared transcriptional targets for our original list of proteins.

13. With the original group members selected, go to the Add menu and select Advanced.
14. Select algorithm type > Find common targets. This will find targets shared by two or more from our selected list.
15. Select Entity type Proteins and Relation type Expression and PromoterBinding.
16. Select Finish to view the pathway. Adjust the visualization by first selecting Layout > Hierarchical Layout, fit the pathway to the window.
17. Select the members of the original group by Select > All and then Select > Mirror Selection to Active Pathway

Let's add highlighting (a color halo) to the upstream regulators and the downstream targets.

18. With the original list selected, go to Select > Invert Selection. Now the added upstream regulators and downstream targets are selected. (Note: relations will also be selected which is ok as highlight won't apply to relationships.)
19. Choose Highlight > blue to highlight the upstream regulators and downstream targets. None of the members of your original list will be highlighted.
20. Save the pathway with the name "Upstream Regulators and Downstream Targets."



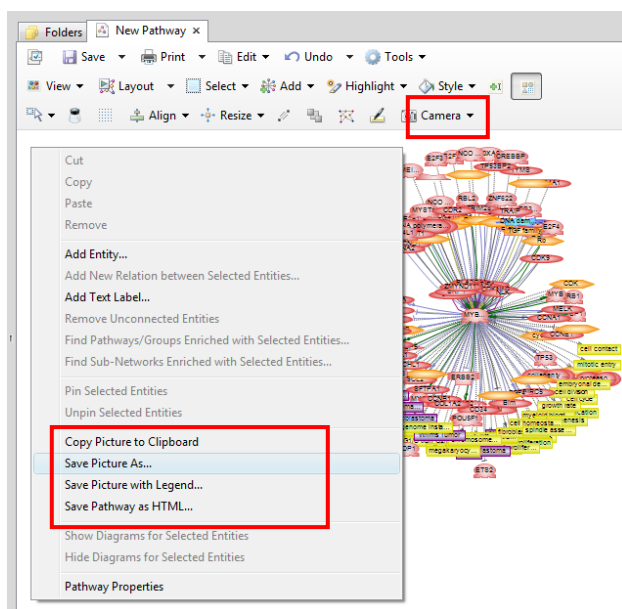
End Exercise: Building Pathways

Saving Pathway Images

Right-click on the pathway to access the four options for saving an image:

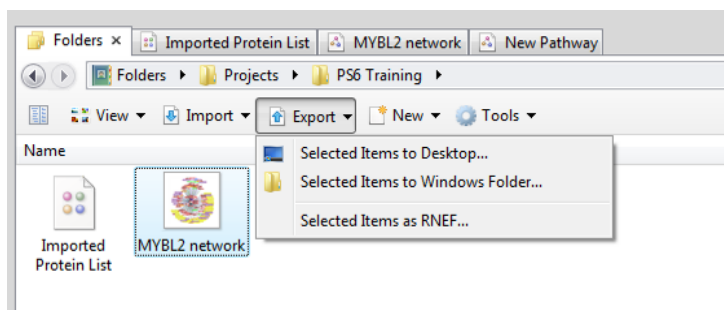
- Copy Picture to clipboard (available to paste into many programs),
- Save Picture As (saves the pathway image as a *.gif – default or *.jpeg, *.png, *.tif, and *.bmp),
- Save Picture with Legend (save an image of a pathway with a legend included. There are also options to scale the image size –height/width and a choice of image resolutions from screen resolution up to 1200 dpi).
- Save Pathway as HTML (export and save the pathway to an HTML file which enables the visualization of a pathway outside of Pathway Studio such as Internet Explorer. The HTML figure has hyperlinks for entities.)

Options for saving a pathway image are also found in the Advanced Visualization Tool bar menu when selecting the camera icon.

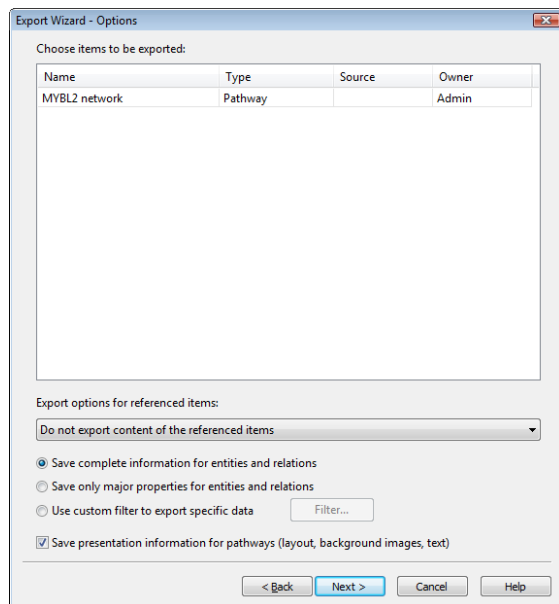


Exporting Pathways

From the folder view select the pathway file for export and select Export. A menu box appears with options for destinations to export the file.

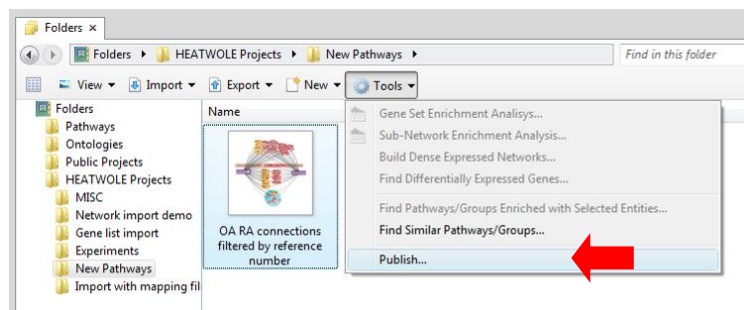


The third option, “Selected Items as RNEF” opens an export wizard that allows for more flexibility in selecting data about the pathway for export.



Publishing Pathways (PC-Client, Enterprise only)

Enterprise PC-client users have the option of submitting their pathway for publication to their institution's server database, thus making the pathway available for other users to access. To request the publication of a pathway to the public area, select the pathway file in the folders view, and then select Publish in the Tools



Individuals at your institution have been designated as publishers and granted access rights to facilitate process your publication request. They work through the web-client to review and accept the pathways for publication. Consult with your system administrator to determine who has publishing rights at your institution.

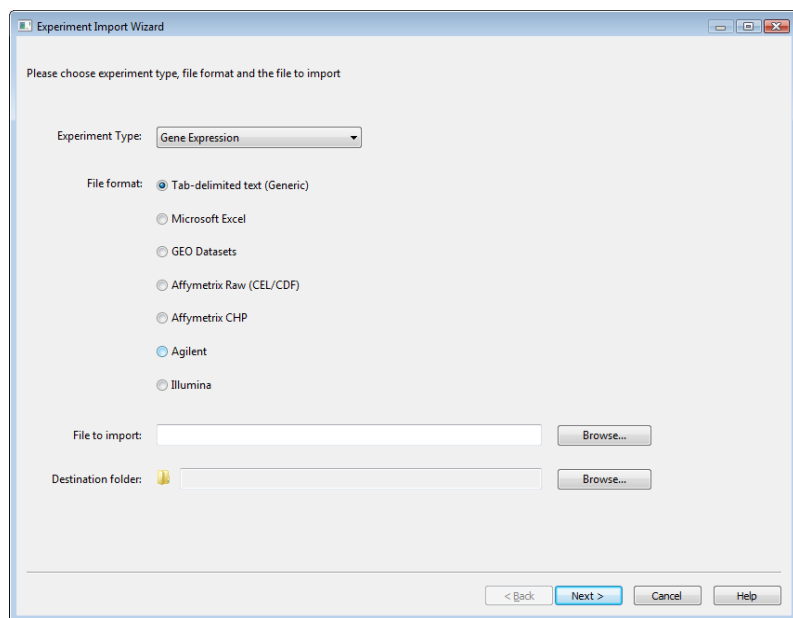
Section 4: Importing Experimental Data

Data formats that can be imported into Pathway Studio

The Import Wizard can handle the following data formats: tab-delimited files, MS Excel files, Affymetrix .cel files or .chp files and GEO dataset SOFT files using the importer. Gene expression data, metabolomics and proteomics data can all be analyzed in Pathway Studio, as long as their IDs can be mapped to identifiers in Pathway Studio (i.e. Entrez GeneIDs, CAS ID, etc.) Pathway Studio provides a broad range of IDs, built into the system to facilitate this mapping. Additionally, you can provide your own map files that map from your assay's probe id to an appropriate ID in Pathway Studio (see: Mapping Files for Microarrays below.)

When importing tab-delimited or MS Excel files the data can be experiment data such as intensity values or other measured values, or it can be data already analyzed externally such as fold-change and p-values. All experimental data must include a primary identifier such as a probe ID or a gene symbol. If you upload individual experimental values, Pathway Studio can provide the ability to 1) view the sample data through a Correlation Viewer to identify potential outliers, 2) add annotation to individual samples, 3) set up multiple user-defined comparisons between sample groups (based on user provided annotation) and 4) compute ratio/fold-change values and p-values for the comparisons. If the imported data already contains computed results of the analysis, the user can simply skip through steps 1-4 listed above and import the tables of results for viewing and further pathway analysis in Pathway Studio.

The Experiment Import Wizard walks you through the steps to import your data.



The screenshot shows the 'Experiment Import Wizard' dialog box. It has a title bar with standard window controls. The main area contains the following elements:

- A message: 'Please choose experiment type, file format and the file to import'.
- 'Experiment Type': A dropdown menu currently set to 'Gene Expression'.
- 'File format': A group of radio buttons with the following options:
 - ☒ Tab-delimited text (Generic)
 - ☐ Microsoft Excel
 - ☐ GEO Datasets
 - ☐ Affymetrix Raw (CEL/CDF)
 - ☐ Affymetrix CHP
 - ☐ Agilent
 - ☐ Illumina
- 'File to import': A text input field followed by a 'Browse...' button.
- 'Destination folder': A text input field preceded by a folder icon, followed by a 'Browse...' button.
- At the bottom, four buttons: '< Back', 'Next >', 'Cancel', and 'Help'.

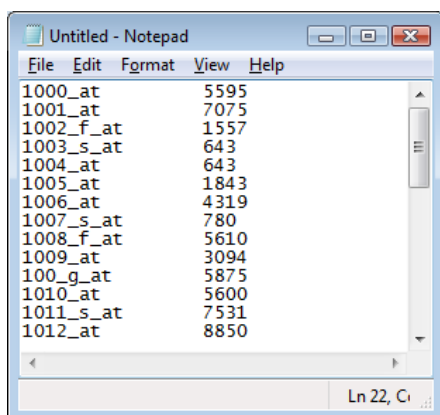
Note: The options for Agilent and Illumina data accept tab delimited data, output from the vendor's analysis software, such as GenomeStudio. Other than Affymetrix, Pathway Studio cannot accommodate direct import of raw instrument data in proprietary format. This data must be normalized in the vendor's platform software prior to import in Pathway Studio.

Mapping Files for Microarrays

Mapping files for many common microarrays, such as most Affymetrix, Agilent, and Illumina are provided within Pathways Studio for use when importing array data. If you are importing data from another type of chip or from a custom designed chip then you will need to generate your own mapping file.

The map file for microarray IDs is a two-column file where the first column contains the microarray IDs used in your chip and the second column contains corresponding Entrez Gene IDs.

For Pathway Studio Desktop, when utilizing a custom-made mapping file, save the file in the following location: C:\ProgramData\AriadneGenomics\PathwayStudio 6\MapFiles\Custom. This will then allow the mapping file to be selected when importing data. For Enterprise users, mapping files must be loaded by your system administrator.



File	Edit	Format	View	Help
1000_at	5595			
1001_at	7075			
1002_f_at	1557			
1003_s_at	643			
1004_at	643			
1005_at	1843			
1006_at	4319			
1007_s_at	780			
1008_f_at	5610			
1009_at	3094			
100_g_at	5875			
1010_at	5600			
1011_s_at	7531			
1012_at	8850			

For Pathway Studio Desktop user, when utilizing a Vista computer and a custom-made mapping file, save the file in the following location: C:\ProgramData\AriadneGenomics\PathwayStudio 7\MapFiles\Custom. This will then allow the mapping file to be selected when importing data.

For Enterprise users, mapping files must be loaded by your system administrator.

Using the Import Wizard to import data

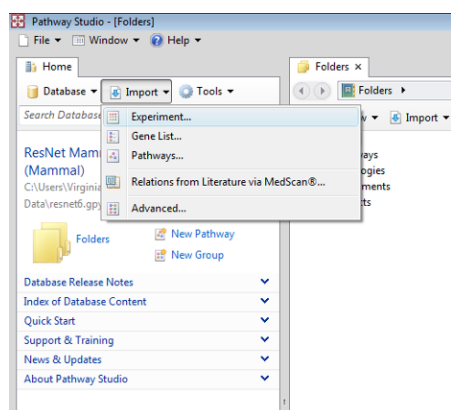
The Experiment Import Wizard is an easy-to-use interface to import experiment data into Pathway Studio. Follow the steps in the exercise below to import a GEO dataset. The step will be slightly different for different data types (ex. .CEL files, MS Excel etc.)

● Exercise Five: Import data

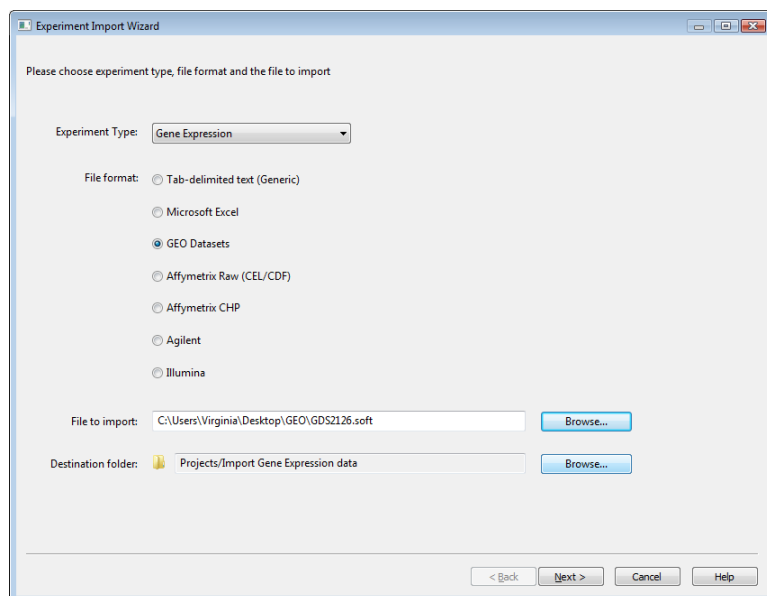
Begin Exercise: Import data

Objective: To launch the data Import Wizard and become familiar with the steps for importing data, including data formats, annotation, utilizing mapping files, sample correlation (outlier analysis), find differentially expressed genes (applicable for gene expression data) and viewing imported data in the Experiment Pane.

1. Launch the Experiment Import Wizard from the Import menu.



2. Select the Experiment Type (Gene Expression), the file format (GEO Dataset), the location of the file to import and the Destination folder where to import the file. This example will demonstrate the import of a publicly accessible GEO data set.



You can upload tab-delimited files, MS Excel files, Affymetrix .CEL files or .CHP files and GEO dataset SOFT files using the importer.

3. In the next three steps you will use the wizard to identify the header row, the first row that contains data, and the column with the probe identity. The wizard will attempt to identify this data for you. If the wizard selects the incorrect row or column, simply can change the selection to the right column.

Experiment Import Wizard

Step 3. Select column with probe identity

ID_REF	IDENTIFIER	GSM34379	GSM34383	GSM34385	GSM34388	GSM34391	GSM34393	GSM34394
1000_at	MAPK3	468.800	374.900	385.800	375.600	354.700	476.800	368.300
1001_at	TIE1	56.800	48.200	52.800	85.900	67.100	43.200	52.700
1002_f_at	CYP2C19	12.700	4.100	2.300	2.700	6.900	2.500	1.500
1003_s_at	CXCR5	15.800	15.400	22.700	20.400	3.400	22.000	20.400
1004_at	CXCR5	39.200	27.900	21.500	8.000	20.700	8.000	27.100
1005_at	DUSP1	2781.500	1764.600	3510.800	2535.700	3492.300	2363.200	1531.300
1006_at	MMP10	7.500	2.000	1.800	2.500	1.300	3.000	3.100
1007_s_at	DDR1	104.600	90.800	155.600	113.000	107.500	98.200	145.600
1008_f_at	EIF2AK2	258.600	184.000	134.400	108.200	90.600	213.800	428.000
1009_at	HINT1	1088.300	841.900	762.800	681.200	744.000	878.700	660.100
100_g_at	RABGGTA	122.300	114.100	97.700	121.300	134.100	104.700	92.400
1010_at	MAPK11	4.600	15.000	8.900	3.500	6.800	6.500	5.900
1011_s_at	YVHAE	131.100	133.000	99.000	79.200	96.300	129.500	134.200
1012_at	PCAF	12.700	6.900	13.800	3.900	4.100	12.400	19.000
1013_at	SMAD5	12.100	17.300	12.600	6.700	6.500	11.800	11.200
1014_at	POLG	147.100	128.600	125.700	156.700	162.900	83.600	104.800
1015_s_at	LIMK1	16.600	19.100	17.700	14.800	8.000	7.500	15.800
1016_s_at	IL13RA2	16.800	15.900	15.700	13.700	1.500	11.200	11.100
1017_at	MSH46	13.700	18.500	10.500	27.300	16.700	27.300	26.100

Legend: Header row First data row Column with probe identity

< Back Next > Cancel Help

Note: If p-values are included in the dataset, arrange the input columns such that the p-value is directly adjacent to each sample value. Then put “2” in the Sample width box and identify which position in the column pair is the location for the p-value and the experiment (expression) value (1 or 2).

4. Next the wizard selects the sample data, and the p-value (if applicable).

Experiment Import Wizard

Step 4. Select first sample column and define sample layout:

Sample width: 1 Expression Value column position: 1

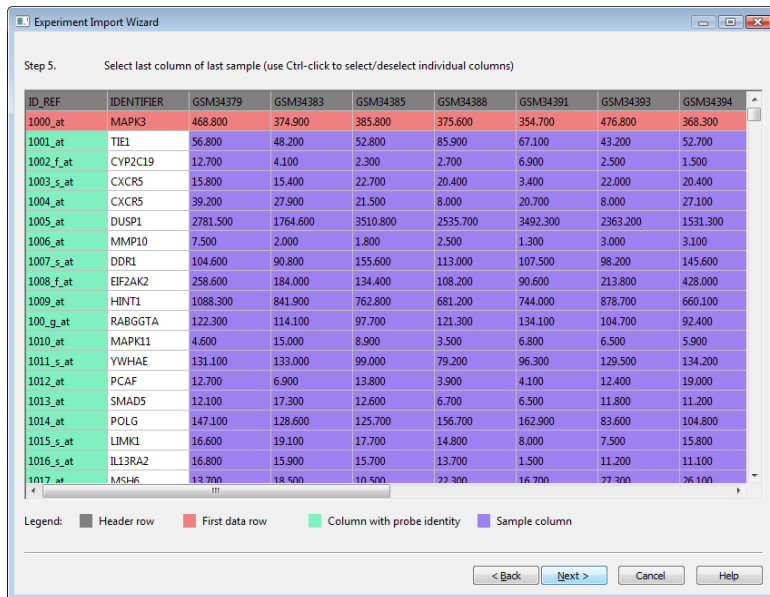
☐ p-value column position: 0

ID_REF	IDENTIFIER	GSM34379	GSM34383	GSM34385	GSM34388	GSM34391	GSM34393	GSM34394
1000_at	MAPK3	468.800	374.900	385.800	375.600	354.700	476.800	368.300
1001_at	TIE1	56.800	48.200	52.800	85.900	67.100	43.200	52.700
1002_f_at	CYP2C19	12.700	4.100	2.300	2.700	6.900	2.500	1.500
1003_s_at	CXCR5	15.800	15.400	22.700	20.400	3.400	22.000	20.400
1004_at	CXCR5	39.200	27.900	21.500	8.000	20.700	8.000	27.100
1005_at	DUSP1	2781.500	1764.600	3510.800	2535.700	3492.300	2363.200	1531.300
1006_at	MMP10	7.500	2.000	1.800	2.500	1.300	3.000	3.100
1007_s_at	DDR1	104.600	90.800	155.600	113.000	107.500	98.200	145.600
1008_f_at	EIF2AK2	258.600	184.000	134.400	108.200	90.600	213.800	428.000
1009_at	HINT1	1088.300	841.900	762.800	681.200	744.000	878.700	660.100
100_g_at	RABGGTA	122.300	114.100	97.700	121.300	134.100	104.700	92.400
1010_at	MAPK11	4.600	15.000	8.900	3.500	6.800	6.500	5.900
1011_s_at	YVHAE	131.100	133.000	99.000	79.200	96.300	129.500	134.200
1012_at	PCAF	12.700	6.900	13.800	3.900	4.100	12.400	19.000
1013_at	SMAD5	12.100	17.300	12.600	6.700	6.500	11.800	11.200

Legend: Header row First data row Column with probe identity Sample column

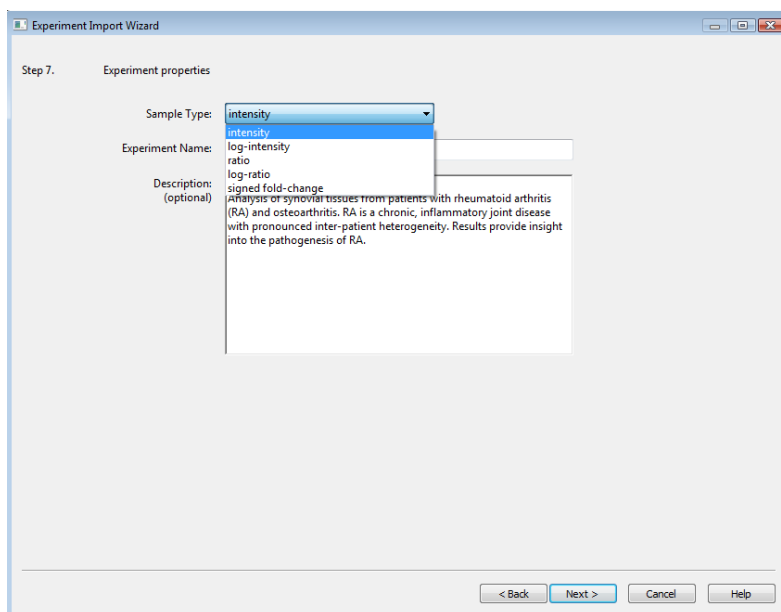
< Back Next > Cancel Help

5. When all the data (rows and columns) have been correctly identified, select Next.

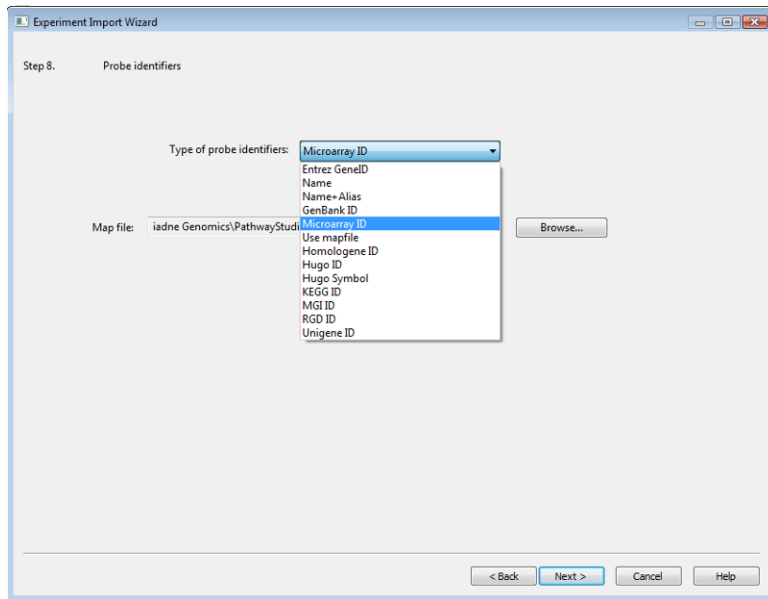


6. An option to include extra, unused columns will be provided. For this exercise, skip the additional identifier column and simply select Next again.

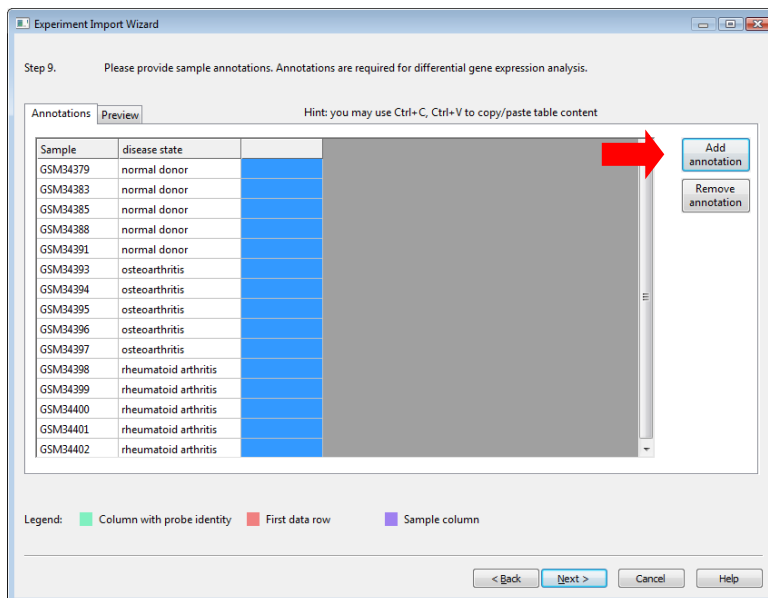
7. Select the appropriate data type to match the data you are importing. In this example you are importing intensity values from an Affymetrix chip study. Give your experiment an appropriate name.



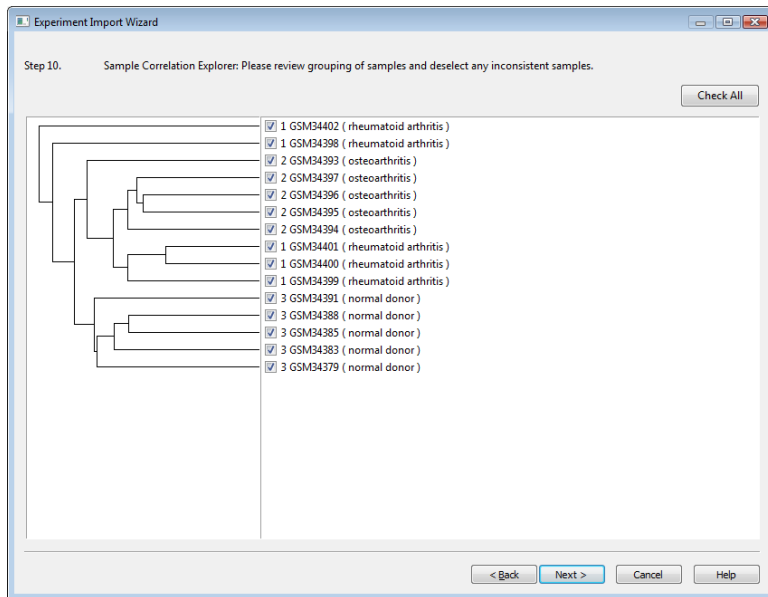
8. Use the Browse button to select the appropriate mapping file for your data.



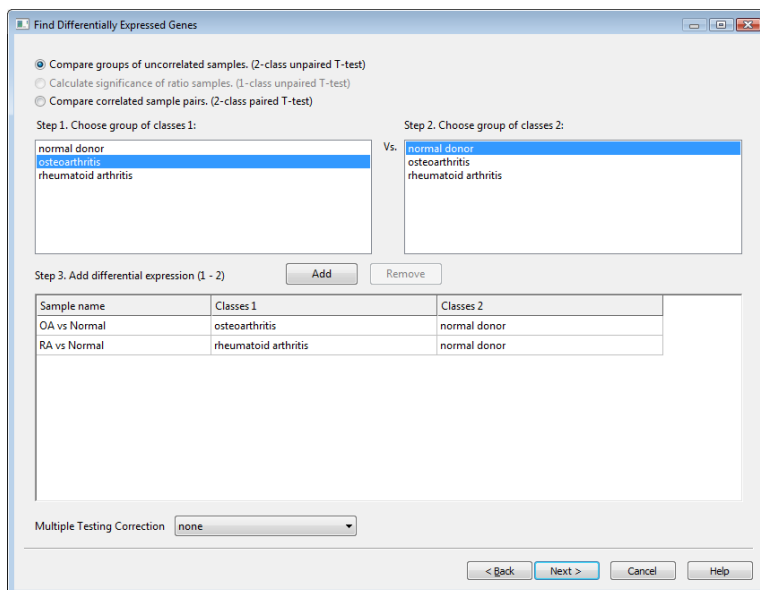
9. Next, annotate the samples with information into appropriate groups for comparisons. Note: The GEO soft file has annotation for the samples included. Additional annotation information can be added by selecting “Add Annotation” button and filling in the newly created fields. At least one column of annotation in addition to Sample is required.



10. The Sample Correlation Explorer allows you to see if there are any potential outliers in your analysis. (Outliers can be the result of errors in annotation, biological variations in the samples or technical variations in the experiment.) To exclude a sample from further analysis, simply uncheck the box and it will be excluded from the data set.



11. Next define the comparisons for computing differentially expressed genes by selecting one group from each of the Step 1 and Step 2 box and then select Add. This adds the comparison in the Step 3 box. You can rename the comparison by typing a new name in the Sample Name box in Step 3. [Optional: Select a preferred testing correction algorithm.] When done, select Next.



12. The Final Preview provides a summary of the imported and mapped data. If the summary is correct, select Finish to import the data.

Final Preview

Total Rows: 12651 Mapped Rows: 11926 Unmapped Rows: 725

Name	Entrez GeneID	GSM34379	GSM34383	GSM34385	GSM34388	GSM34391	GSM34393	GSM34394
disease state		normal donor	normal donor	normal donor	normal donor	normal donor	osteoarthritis	osteoarthritis
MAPK3	5595	468.799988	374.899994	385.799988	375.600006	354.700012	476.799988	368.299988
TIE1	7075	56.799999	48.200001	52.799999	85.900002	67.099998	43.200001	52.700001
CYP2C19	1557	12.700000	4.100000	2.300000	2.700000	6.900000	2.500000	1.500000
CXCR5	643	15.800000	15.400000	22.700001	20.400000	3.400000	22.000000	20.400000
CXCR5	643	39.200001	27.900000	21.500000	8.000000	20.700001	8.000000	27.100000
DUSP1	1843	2781.500000	1764.599976	3510.800049	2535.699951	3492.300049	2363.199951	1531.300049
MMP10	4319	7.500000	2.000000	1.800000	2.500000	1.300000	3.000000	3.100000
DDR1	780	104.599998	90.800003	155.600006	113.000000	107.500000	98.199997	145.600006
EIF2AK2	5610	258.600006	184.000000	134.399994	108.199997	90.599998	213.800003	428.000000
HINT1	3094	1088.300049	841.900024	762.799988	681.200012	744.000000	878.700012	660.099976
RABGGTA	5875	122.300003	114.099998	97.699997	121.300003	134.100006	104.699997	92.400002
MAPK11	5600	4.600000	15.000000	8.900000	3.500000	6.800000	6.500000	5.900000
YWHAE	7531	131.100006	133.000000	99.000000	79.199997	96.300003	129.500000	134.199997
PCAF	8850	12.700000	6.900000	13.800000	3.900000	4.100000	12.400000	19.000000

Legend: ■ Column with probe identity ■ First data row ■ Sample column After finish: ☒ Open Experiment ☐ Run GSEA

< Back Finish Cancel Help

Note: If gene symbols do not appear when Name is selected in this menu, this is an indication that they were not mapped to Entrez GeneIDs.

Some probes may measure controls, for which there is no corresponding gene. Other probes may measure ESTs, which are not contained in the ResNet database.

If many probes are not mapped, you should identify and understand the reason before proceeding with further analysis.

- The imported experiment data will be displayed in a table in the Experiment Pane.
- You can adjust the table view in the Colors menu. You can view the values by color, by text only or both. In addition the color contrast can be adjusted by selecting a desired contrast level.

Rheumatoid and osteoarthritis synovial tissue x

Link Edit Select Tools

View Colors

Name	Contrast	Highest	p-values for RA ...
disease state			
MST1R			7.866e-001
RBBP4			1.094e-001
DDR2			7.989e-004
TAF11			3.175e-002
PTPN21	0.4828	1.418e-001	-0.4746
EMP1	-0.3117	1.627e-001	0.1548
E2F1	-0.4677	4.409e-001	-0.1237
UBB	-0.0437	7.655e-001	0.0063
RAD9A	0.8136	1.043e-001	1.0308
SMAD1	-0.5737	1.073e-001	-0.1850
CASP10	0.2037	1.968e-001	0.8633

Advanced Options...

- The data can be viewed as log-change, fold-change, signed fold change or absolute fold-change. To access these options right click on the header.

Rheumatoid and osteoarthritis synovial tissue x

Link Edit Select Tools

View Colors

Name	OA vs Normal	p-values for OA...	RA vs Normal	p-values for RA ...
disease state				
MST1R	1.1971			7.866e-001
RBBP4	1.4719			1.094e-001
DDR2	-1.2756			7.989e-004
TAF11	-1.6522			3.175e-002
PTPN21	1.3975			4.649e-001
EMP1	-1.2412	1.627e-001	1.1132	7.039e-001
E2F1	-1.3829	4.409e-001	-1.0895	7.897e-001
UBB	-1.0307	7.655e-001	1.0044	9.525e-001
RAD9A	1.7576	1.043e-001	2.0431	1.393e-001
SMAD1	-1.4884	1.073e-001	-1.1368	6.431e-001
CASP10	1.1516	1.968e-001	1.8192	3.305e-004

Log-change
Fold-change
Signed Fold-change
Absolute Fold-change
Delete Sample...

End Exercise: Import data

Section 5: Experimental data analysis tools

The experimental data analysis tools in Pathway Studio provide two different enrichment analysis algorithms, the Fisher's Exact Test and Gene Set Enrichment Analysis. Identification of enrichment of defined gene sets (pathways and groups) as well as user-defined sub-network identified in the ResNet database are both possible.

gene sets algorithms	Known Gene Sets (ontologies, curated pathways)	Sub-Networks (user defined from ResNet)
Fisher's Exact Test (experimental values <u>not</u> utilized)	Find Pathways/Groups Enriched with Selected Entities	Find Sub-Networks Enriched with Selected Entities
Gene Set Enrichment Analysis (GSEA) (experimental values <u>are</u> considered in the analysis)	Gene Set Enrichment Analysis	Sub-Network Enrichment Analysis

Names of the tools as they appear in Pathway Studio

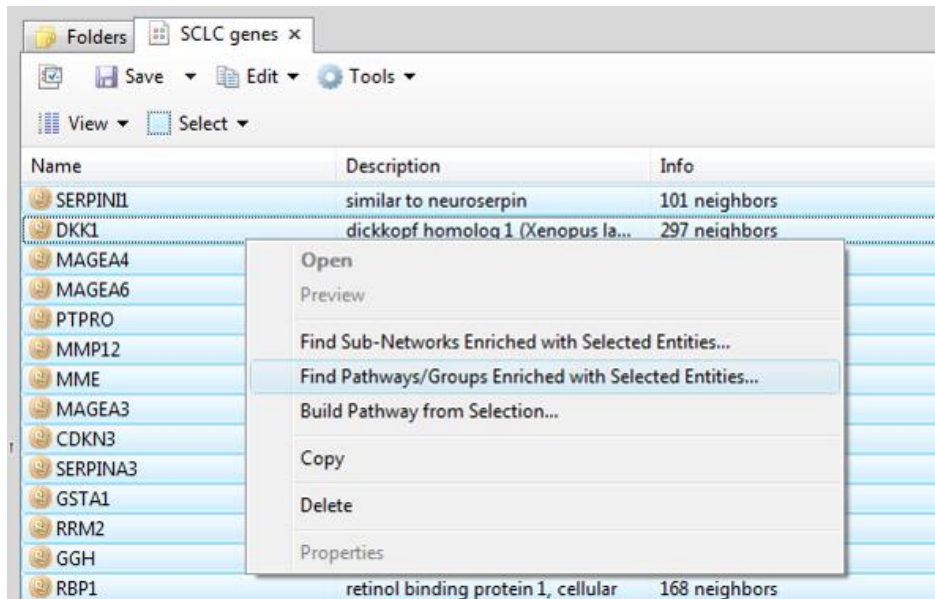
Fisher's Exact Test

Fisher's Exact test is a statistical test used to determine if there are nonrandom associations between two categorical variables. You can use the Fisher's Exact test to see if there are gene groups (such as ontology groups) or pathways that are statistically enriched in your list of genes. In addition you can identify if sub-networks are enriched.

For more information about the Fisher's Exact test see: <http://mathworld.wolfram.com/FishersExactTest.html>

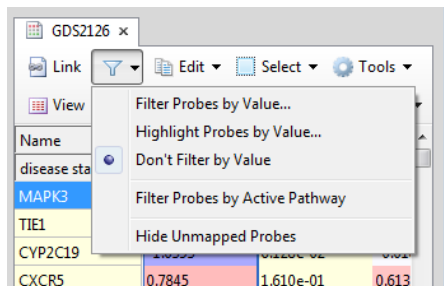
In gene expression analysis, the Fisher's Exact test is typically run on the list of genes that have been determined (ex. by fold change / p-value) to be statistically significantly differently expressed between experimental conditions. **The experimentally derived values are not utilized in the Fisher's Exact test when calculating enrichment in the gene list.**

You can launch the Fisher's Exact test from the Group view by selecting the proteins in the group, then right-click and select "Find Pathways/Groups Enriched with Selected Entities".

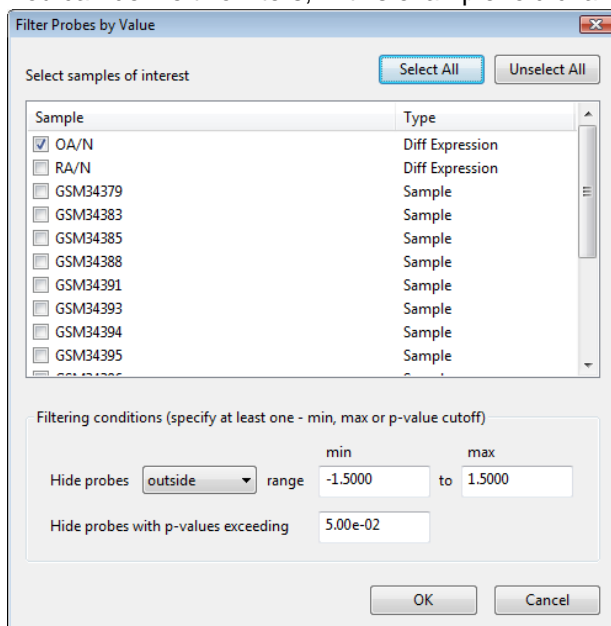


You can also launch the Fisher's Exact Test from the Experiment Pane. First filter the view of your results table by fold change / p-value to identify the list of statistically significant genes.

Filter a table by selecting the filter icon and choosing: "Filter Probes by Value"



You can define two filters, in this example fold change range and p-value range.



Note: Applying this filter does not remove the rows that are filtered out. The rows containing values that do not meet the filter criteria are displayed in gray.

To run an analysis on only the probes that did meet the filter criteria, you must first choose "Select Unfiltered Probes."

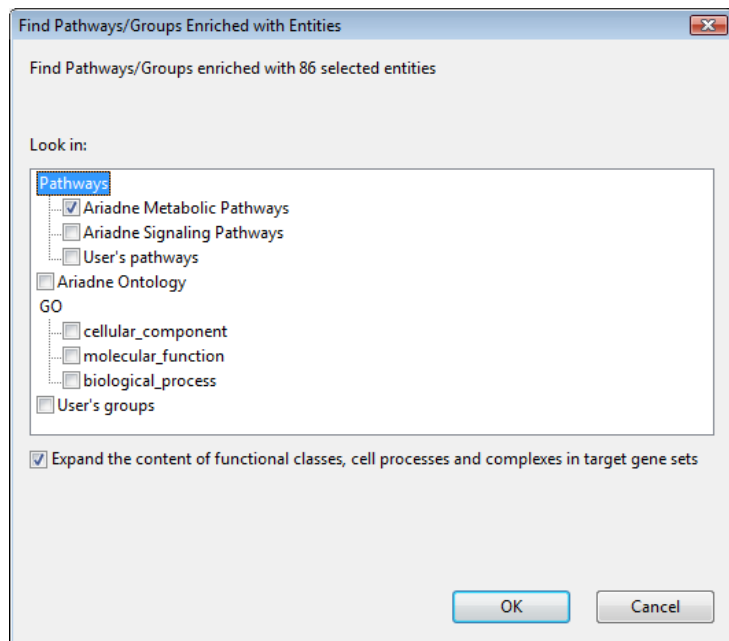
In addition, you can determine this list of genes in an analysis outside of Pathway Studio, and import the resultant list into Pathway Studio for Fisher's Exact test analysis.

To run the Fisher's Exact test from the Experiment Pane, select the genes to be included in the analysis by choosing "Select Unfiltered Probes", then right-click on the gene name column and select "Find Pathways/Groups Enriched with Selected Entities."

The screenshots show the Pathway Studio interface with a table of gene expression data. The table has columns for Name, OA/N, p-values for..., RA/N, and p-values for R... The data is color-coded by p-value. The right-click context menu is open, showing options like 'Build Pathway from Selection...', 'Save Selection as Group...', 'Find Pathways/Groups Enriched with Selected Entities...', 'Find Sub-Networks Enriched with Selected Entities...', 'Copy', 'Select Unfiltered Probes', and 'Bring Selected Probes Together'. The 'Find Pathways/Groups Enriched with Selected Entities...' option is highlighted in blue in the right screenshot.

Name	OA/N	p-values for...	RA/N	p-values for R...
disease state				
DDIT4	-3.2274	1.108e-07	-2.1787	6.397e-04
HIFX	-2.2321	3.932e-07	-2.1971	7.210e-04
HIFX	-1.1628	7.150e-07	-1.1660	1.276e-03
MMP3	3.8320	3.986e-06	4.3299	5.974e-05
RAB8A	0.6196	5.130e-06	0.5693	9.755e-03
ITPKC	-3.1132	6.559e-06	-2.0506	5.059e-06
TSC22D3	-1.4353	7.789e-06	-1.1027	1.186e-03
MAOA	-2.1788	8.079e-06	-3.3359	6.010e-06
FAS				
HNF1A				
VEGFA				
GADD45A				
GNE				
TCTA				
SLC36A1				
MAFF				
TNFRSF11A	2.0172	2.275e-05	1.4092	2.796e-02
PPAP2B	-1.1413	2.750e-05	-1.7817	4.347e-05
TRIM14	1.9815	2.888e-05	1.7270	5.693e-04
HSPA1A	-2.0447	3.001e-05	-2.1024	1.782e-04
FADS1	-1.8955	3.342e-05	-1.7986	3.151e-05
RGS19	1.0190	3.780e-05	1.0598	5.787e-05

The Find Pathways/Groups Enriched with Entities dialog opens. Here you can select by check box the groups and pathways to include in the analysis. In Pathway Studio you have available the classic Gene Ontology as well as Ariadne Ontology and the Ariadne reference pathway collection to use in the analysis. In addition you can include any pathways and groups you have created in the analysis.



Note: Some types of entities, such as functional classes in the Cellular Process pathways contain proteins that have been mapped to them. To include these proteins in the analysis make sure the check box labeled, “Expand the content of functional classes and complexes in target gene sets” is selected.

The results of the enrichment analysis are displayed in a table in the List Pane, ranked by p-value.

Enriched Pathways/Groups x										
Save Edit Select Tools Find in this table										
Name	Type	Total Entities	Expanded # of Entities	Overlap	Percent Overlap	Overlapping Entities	p-v...	Data Source	#	
Estrogen-like	Group	9	9	2	22	PGR, ESRRB	0.00120421	Ariadne Ontology	1	
Melanogenesis	Pathway	51	682	13	1	ITPKC, ADM, TGFA, VEGFA, CXCL12, PRKCB, CSF1R, ED...	0.00286401	Ariadne Signaling Pathways	2	
ECM degradation	Group	14	14	2	14	MMP3, MMP14	0.00298662	Ariadne Ontology	3	
Adipocytokine Signaling	Pathway	52	780	14	1	ADIPOQ, IL6R, FASN, ADFP, TNFRSF11A, FABP4, ACSL1...	0.00334572	Ariadne Signaling Pathways	4	
Oncogenes	Group	273	273	6	2	RABL3, CSF1R, ETS2, RET, MYBL1, MAFF	0.00531976	Ariadne Ontology	5	
Gonadotrope Cell Activat...	Pathway	71	698	12	1	ITPKC, ADM, TGFA, VEGFA, MMP3, PRKCB, MMP14, CS...	0.00976601	Ariadne Signaling Pathways	6	
Adiponectin	Group	2	2	1	50	ADIPOQ	0.0117851	Ariadne Ontology	7	
GADD45	Group	3	3	1	33	GADD45A	0.0176263	Ariadne Ontology	8	

The “# of Entities” indicates the number of entities shown in the group or pathway. The “Expanded # of Entities” includes the total number of entities shown on the pathway and contained in complexes and functional classes. The “Overlap” is the number of proteins shared in common between your input group and the resultant group or pathways. Data Source indicates the source of the group or pathway.

You can save the results of the analysis by exporting to MS Excel: Select > All, Tools > Send Data to Excel. In addition, you can save the results in Pathway Studio by selecting “Save” from the tool bar just above the table.

● Exercise Six: Experimental Data Analysis – Fisher’s Exact Test for Enriched Groups and Pathways

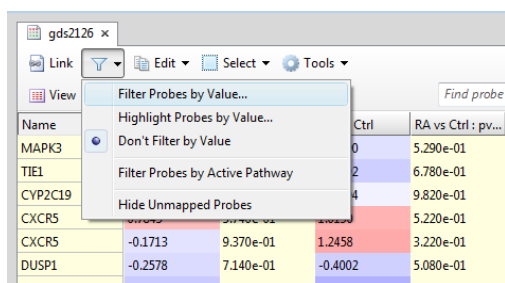
Begin Exercise:

Objective: To run a Fisher’s Exact test analysis on a filtered list from gene expression data to identify ontology groups and pathways enriched in the significantly differentially expressed gene set.

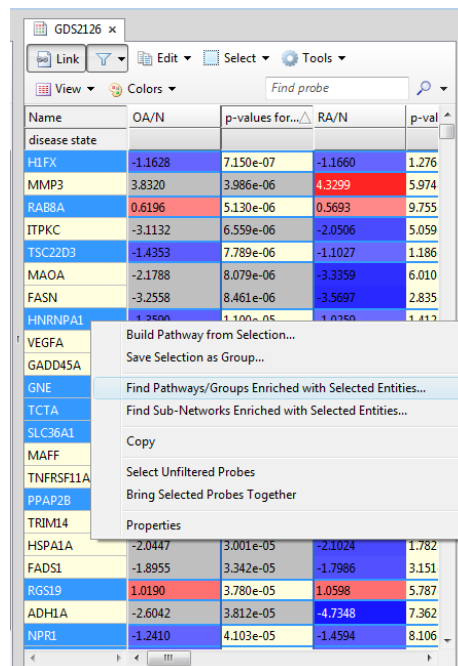
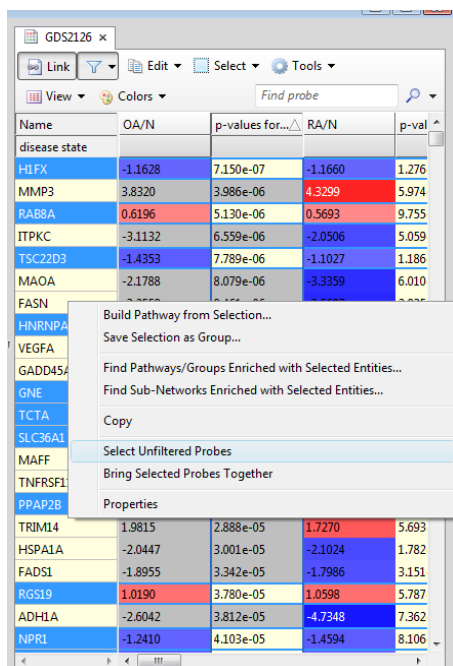
Let’s use the osteoarthritis and rheumatoid arthritis synovial tissue data set to run the Fisher’s Exact test.

Fisher’s Exact Test

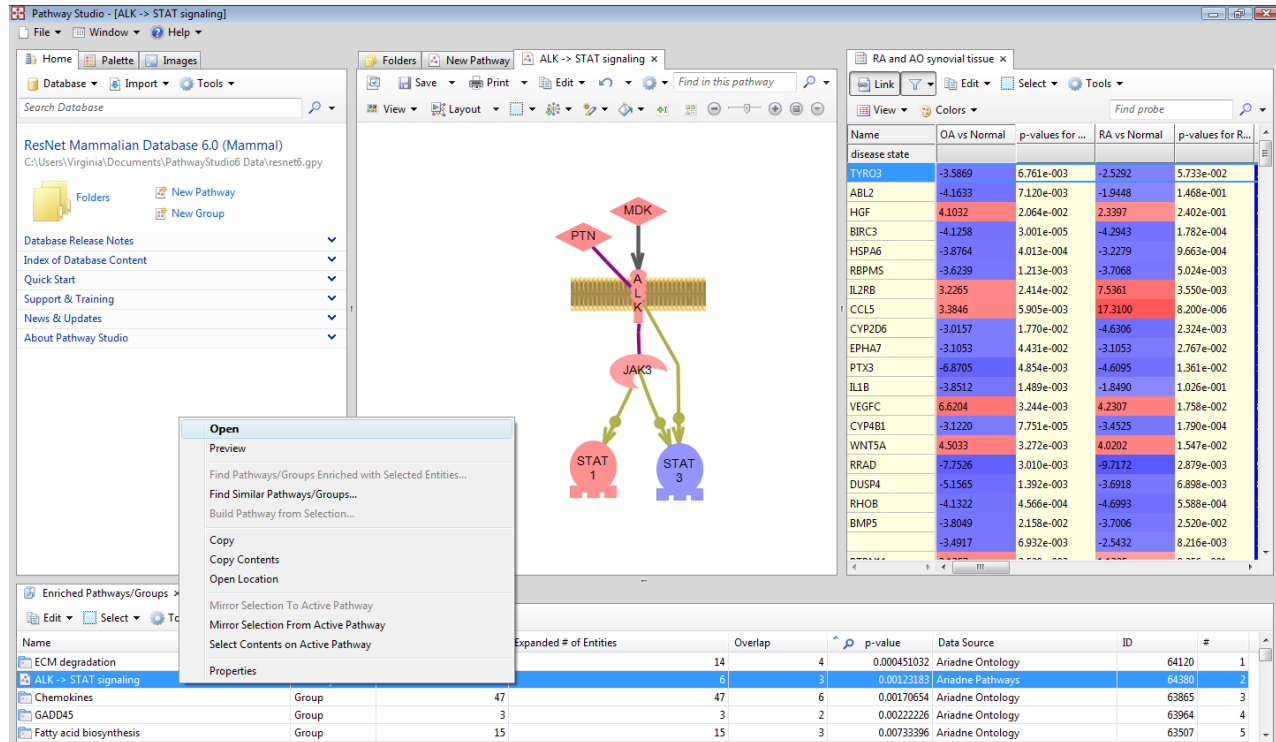
1. Filter the table by selecting the filter icon and selecting “Filter Probe by Value”.



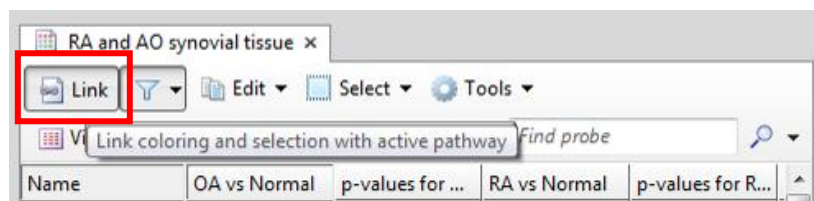
2. Set the filter for the Osteoarthritis data set (check that box) to include only genes with a log change range greater than 1.5 (i.e. set the range to be outside of -1.5 and 1.5), and set the p-value cut off at 0.05.
3. Select “Set Filter.” All genes that don’t meet the filter criteria and displayed in gray.
4. Right click on the table and choose “Select Unfiltered Probes.” Then right click and select “Find pathways/groups Enriched with Selected Entities”. This runs the Fisher’s Exact test.



5. Select “Ariadne Pathways” and “Ariadne Ontology” for the analysis and then select “OK” to run the analysis. (Make sure the check box for “Expand the content of functional classes and complexes in target gene sets” is selected).
6. See the results in the List Pane.
7. Find the top pathway in the list and right-click and select “Open” to see it in a pathway view.

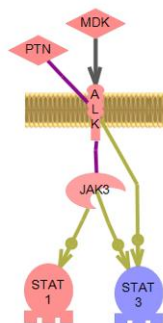


8. If the proteins in the pathway are not colored by the experimental values (here red and blue), then select the “Link” button above the Experiment Pane. This will apply the colored values of the experiment to the nodes on the pathway.

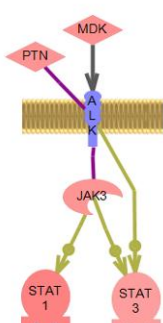


9. With the “Link” function active (the pathway nodes colored by the experiment values) select the second dataset, rheumatoid arthritis, by clicking that column header. Do you see any change in the expression levels of any of the proteins in the pathway?

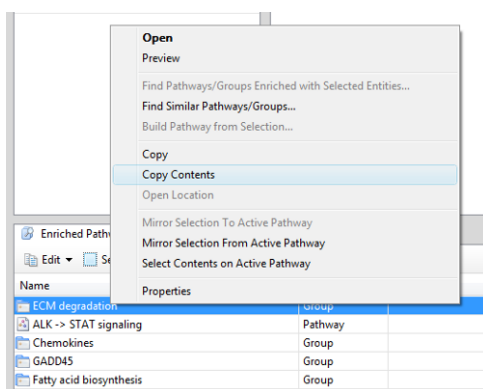
Osteoarthritis



Rheumatoid arthritis



- Find the members of an ontology group that overlap with your gene list, right click on the ontology group and select "Copy Contents." This will copy the proteins in the ontology to the clipboard. Note: if the ontology contains a child ontology group, the contents of that lower level will not be copied.



- In the Experiment Pane choose: Select > Entities on Clipboard. The proteins of the ontology group will be highlighted in the Experiment table.

Name	OA vs Normal	p-values for...	RA vs Normal	p-values for R...
disease state				
JMJD6	-4.0087	1.989e-003	-1.3712	1.553e-001
MMP1	10.2186	2.008e-003	25.3763	4.960e-003
PDE4A	-3.6573	2.028e-003	-3.2914	1.160e-002
STXBPL	-3.3703	2.075e-003	-1.0146	9.752e-001
STMN2	4.6250	2.323e-003	2.1641	1.698e-001
MT2A	-3.0628	2.353e-003	-2.4260	5.068e-002
NID2	3.8366	2.357e-003	3.2449	4.084e-003
3-8-1	4.0255	2.387e-003	2.7571	3.152e-002
CD27	10.2953	2.406e-003	32.1093	2.663e-005
MMP7	-4.0085	2.596e-003	-3.8596	1.181e-002
CNR1	-3.4803	2.792e-003	-4.8477	1.177e-002
LAMA2	-3.4851	2.812e-003	-10.5789	1.398e-003

- Results of the analysis can be saved by exporting to MS Excel: Select > All, Tools > Send Data to Excel. Results can also be saved in Pathway Studio by selecting Save from the Tool bar above the table.
- Leave the tab open for use in the following exercise.

End Exercise: Experimental Data Analysis – Fisher's Exact Test

Gene Set Enrichment Analysis

Gene Set Enrichment Analysis (GSEA) is similar to the Fisher's Exact test in that it identifies statistical enrichment in experimental data of known groups (such as ontologies) and curated pathways. Gene Set Enrichment analysis differs from Fisher's Exact test in that the **rank** (based on experimental data values) of the genes in the experimental dataset is taken into consideration when identifying enrichment and the entire dataset can be used (no statistical threshold needs to be initially defined).

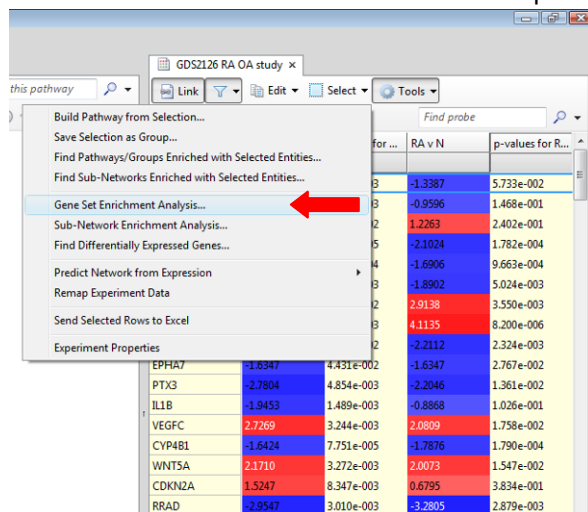
Potential Advantages of Gene Set Enrichment over Fisher's Exact Test:

- Threshold relevance – In Fisher's Exact test only the subset of genes determined by a relevance threshold is considered. This list can be variable depending on the user defined chosen threshold (ex. fold change or p-value).
- In Fisher's Exact Test the rank position of the gene in the experimental results is not considered (it is either on the list or its not). In GSEA rank is considered.
- GSEA is able to identify when many members of a pathway are changed even if none are changed above the threshold used in the Fisher's Exact test (correlation).

It is unnecessary to filter your data set by fold change/p-value before running the Gene Set Enrichment Analysis; however you can filter to remove genes with high p-values if desired. (Recall the filters applied in Pathway Studio simply filter the view, but not the underlying data, and have no effect when used in the GSEA analysis.)

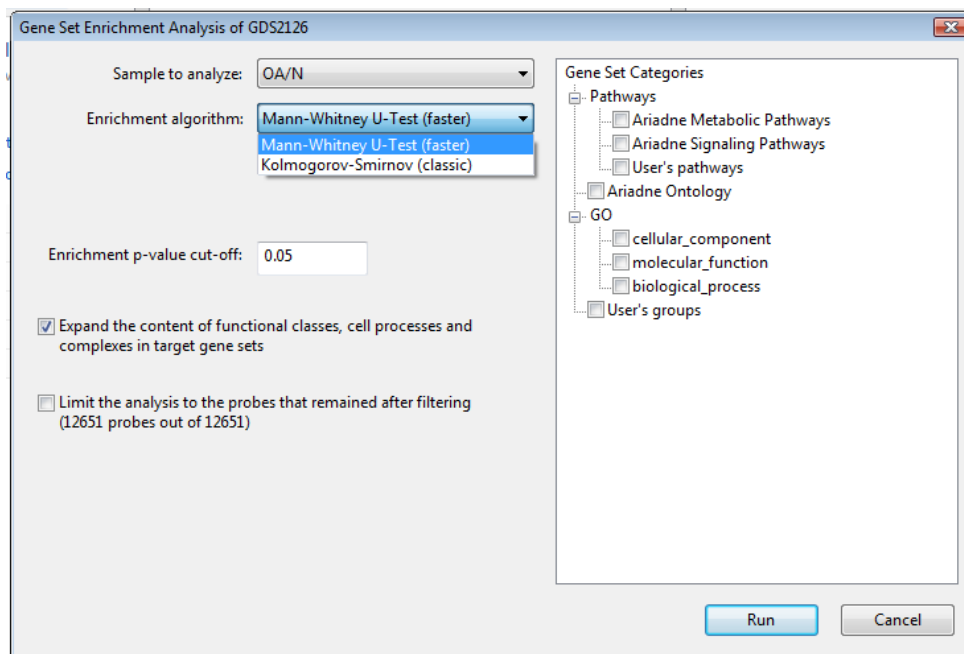
Note: The Gene Set Enrichment Analysis algorithm was developed at the Broad Institute, (<http://www.broad.mit.edu/gsea/>). The Broad Institute has curated a large number of gene sets that can be used in this analysis. Gene sets downloaded from the Broad Institute can then be imported into Pathway Studio by selecting Advanced in the Import menu in the Information Pane. Select Gene Sets/Gene Sets in Broad Institute Format to download these gene sets.

Launch GSEA from the Tools menu in the Experiment Pane.



Select the dataset for the analysis, the gene sets (ontologies/pathways,) and the preferred algorithm. You will get very similar results from both the Mann-Whitney U-Test or the Kolmogorov-Smirnov algorithms however the

Mann-Whitney U-Test will run much faster. As with the Fisher's Exact test, check the box “Expand the contents of functional classes and complexes in target gene sets, to include those genes in your analysis.



Select “Run” to compute the results and view the results in the bottom List Pane. Enriched gene sets are ranked by p-value.

Significant Gene Sets for GDS2126: OA/N									
Name	Type	Total Entities	Expanded # of En...	# of Measured En...	Measured Entities	Median change	p-value	Gene Set Category	#
T-cell receptor -> NFATC sig...	Pathway	37	172	49	PIK3C2G, TRBV19, TRAV20, CD86, CD80, CD8...	1.48681	2.18275e-06	Ariadne Signaling Path...	4
Melanogenesis	Pathway	51	682	515	ALDH3B2, AOC2, CALML3, GPR183, GPR18, ...	1.02095	2.88109e-06	Ariadne Signaling Path...	5
Focal Adhesion Regulation	Pathway	41	308	253	PIK3C2G, ITGBL1, TAOK3, EGF, PLG, HGF, FG...	1.08994	1.10461e-05	Ariadne Signaling Path...	6
Gap Junction Regulation	Pathway	51	639	469	GPR183, GPR18, GUCY1A2, SGSM3, TAOK3, ...	1.00207	1.23365e-05	Ariadne Signaling Path...	7
Mast Cell Activation	Pathway	64	529	383	CALML3, AP2S1, CLTB, IFNA16, PIK3C2G, ST...	1.0827	1.37767e-05	Ariadne Signaling Path...	8
T-cell receptor -> ATF/CREB...	Pathway	49	189	66	PIK3C2G, TRBV19, TRAV20, RAC1, CD86, CD...	1.43807	1.45816e-05	Ariadne Signaling Path...	9
TGF family	Group	25	25	16	TGFB1, BMP4, BMP7, TGFB2, TGFB3, BMP2, I...	-1.20278	9.44121e-05	Ariadne Ontology	10
T-cell receptor -> CREBBP si...	Pathway	36	176	47	CALML3, PIK3C2G, TRBV19, TRAV20, CD86, ...	1.48681	0.000134352	Ariadne Signaling Path...	11

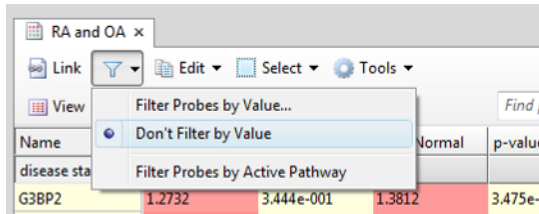
● Exercise Seven: Experimental Data Analysis – Gene Set Enrichment Analysis

Begin Exercise:

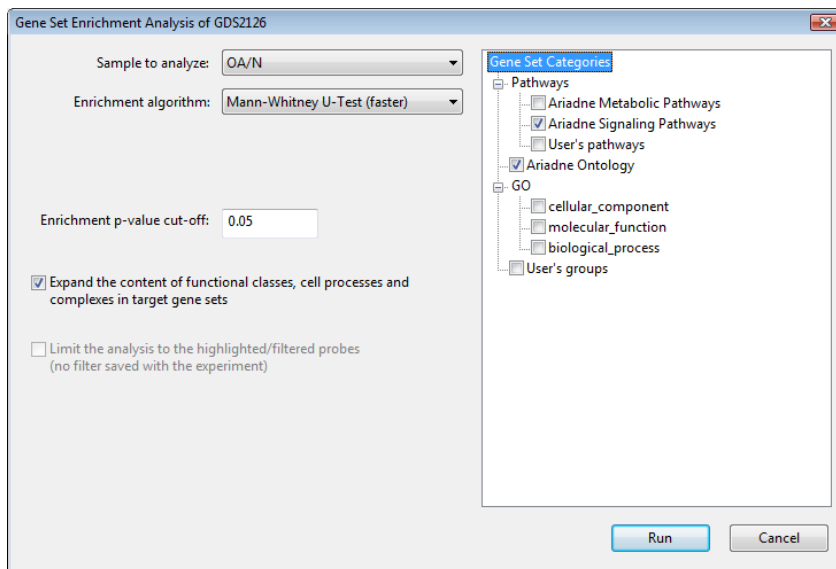
Objective: To run the Gene Set Enrichment Analysis (GSEA) on a gene expression data set to find ontology groups and pathways that are enriched in the dataset and to compare the results of GSEA to the results obtained by the Fisher's Exact test.

Let's use the osteoarthritis and rheumatoid arthritis synovial tissue data set to run the Gene Set Enrichment Analysis and find Ariadne Pathways and Ariadne Ontology groups enriched in the experimental results.

1. If a filter has been applied to the experimental dataset, remove the filter before continuing.



2. Select Tools > Gene Set Enrichment Analysis.



3. Select: Sample to analysis: (osteoarthritis), Gene Set Categories: Ariadne Pathways and Ariadne Ontology, Enrichment algorithm, Mann-Whitney U-Test, p-value cut-off: 0.05, check "Expand the contents of functional classes and complexes in target gene sets." When finished, select "Run" to compute the results. The results appear in the List Pane, sorted by p-value.
4. Right-click on the top pathway and select open to view the pathway. See that this pathway contains many functional classes.

The screenshot shows the Pathway Studio 6.0 interface. On the left is the 'Database' pane with 'ResNet Mammalian Database 6.0 (Mammal)' selected. The main window displays a network diagram of 'Gonadotrope Cell Activation'. On the right, a table titled 'RA and AO synovial tissue' shows gene expression data. Below the network diagram, a table lists various pathways and their associated gene sets.

Name	OA vs Normal	p-values for...	RA vs Normal	p-values for R...
disease state				
MAPK3	1.1631	1.127e-001	-1.2725	2.110e-001
KCNK3	-1.2837	1.127e-001	-1.5072	7.673e-002
ITPR1	1.3804	1.127e-001	1.2301	3.017e-001
PTBP1	-1.2893	1.128e-001	-1.2791	1.375e-001
SLC30A1	1.3831	1.130e-001	1.7075	4.674e-002
IDS	1.5174	1.130e-001	1.3348	1.849e-001
RTCD1	1.3715	1.131e-001	1.6947	8.698e-002
TRIM58	1.5898	1.131e-001	1.2038	4.519e-001
HIST2H4A	2.3114	1.131e-001	2.2767	6.032e-002
FEN1	1.2821	1.132e-001	1.4117	1.970e-002
APOBEC3C	1.3504	1.132e-001	1.5702	2.331e-003
LAIR2	1.8972	1.132e-001	2.8316	1.977e-002
LCP2	1.4899	1.133e-001	2.8430	2.609e-003
BMI1	1.3925	1.133e-001	1.2304	2.828e-001
PKFB3	-1.4728	1.133e-001	-1.2934	3.460e-001
CL4orf109	1.3069	1.133e-001	-1.0533	7.637e-001

Expanded # of Entities	# of Measured Entities	p-value	Gene Set Category	ID	#	
71	702	514	9.79568e-010	Ariadne Pathways	64224	1
70	558	421	6.92444e-007	Ariadne Pathways	64225	2
59	525	373	9.91523e-007	Ariadne Pathways	64220	3
51	687	510	1.53027e-006	Ariadne Pathways	64218	4
37	54	44	5.01762e-006	Ariadne Pathways	64222	5
64	531	380	9.35419e-006	Ariadne Pathways	64227	6
51	644	467	1.01549e-005	Ariadne Pathways	64216	7
41	309	252	2.57149e-005	Ariadne Pathways	64231	8
49	71	61	3.59648e-005	Ariadne Pathways	64427	9

5. Select a functional class and right-click to open the properties view. Select the members tab to see the members of the functional class. Select "Open list in bottom pane."

The screenshot shows the 'Functional Class Properties' dialog box with the 'Members' tab selected. It lists various proteins associated with the functional class.

Members
AXL
CSF1R
EGFR
ERBB2
ERBB3
ERBB4
FGFR1
FGFR2
FGFR3
FGFR4
FGFR6
FGFR11
FLT1
FLT3
FLT4
KDR
KIT
MERTK
MET
MST1R
PDGFRA
PDGFRB
TEK
TIE1
TYRO3

Buttons: Open list in bottom pane, OK, Cancel

6. With the Functional Class open in the list pane choose Select > All, Edit > Copy. This copies the list of proteins to the clipboard.
7. In the Experiment Pane, choose Select > Entities on Clipboard. Next, right-click and choose "Bring Selected Probes Together." This will group all the selected proteins in the table together. If no proteins are selected, there is no overlap between the selected functional class and your dataset.

RA and AO synovial tissue x

Link View Colors Find probe

Name	OA vs Normal	p-values for...	RA vs Normal	p-values for R...
disease state				
FGFR1	1.0289	7.336e-001	-1.2432	1.260e-001
ERBB3	-1.0905	7.508e-001	-1.4807	2.204e-001
FGFR2	1.1789	7.516e-001	-1.0751	8.261e-001
FLT3	1.0754	7.570e-001	1.1702	4.539e-001
FGFR1	1.0515	7.627e-001	-1.1694	3.224e-001
FGFR3				
FGFR1				
ERBB3				
ERBB2				
CSF1R				
FLT3				
KDR				
KIT				
FLT1				
ERBB2				
PDGFRA				
FLT3	1.0185	0.715e-001	1.2770	5.037e-001

Build Pathway from Selection...
Save Selection as Group...
Gene Set Enrichment Analysis...
Sub-Network Enrichment Analysis...
Find Differentially Expressed Genes...
Find Pathways/Groups Enriched with Selected Entities...
Copy
Bring Selected Probes Together
Properties

8. To find the overlap of pathways and groups between the results of the Fisher's Exact test and the Gene Set Enrichment Analysis for a dataset, first choose the tab containing the Fisher's Exact test result (labeled "Enriched Pathways/Groups"). This will allow you to examine what pathways/groups in your results that are unique to one of the algorithms and which are common to both.
9. Choose Select > All, Edit > Copy.
10. Next, choose the GSEA results tab "Significant Gene Sets". Choose Select > Entities on Clipboard. The groups and pathways that are in both analysis results will be selected.

Enriched Pathways/Groups Significant Gene Sets x

Edit Select Tools

Name	Type	# of Entities	Expanded # of Entities	# of Measured Entities	p-value	Gene Set Category	ID	#
Neuromodulator	Group	62	62	47	0.0194734	Ariadne Ontology	63873	57
Adipocytokine Signaling	Pathway	52	783	566	0.0201127	Ariadne Pathways	64230	58
AngiotensinR -> STAT signaling	Pathway	7	7	7	0.0208396	Ariadne Pathways	64338	59
VEGFR -> ATF/CREB/ELK-SRF signaling	Pathway	43	43	37	0.0222992	Ariadne Pathways	64412	60
EndothelinRb -> AP-1/CREB/ELK-SRF signaling	Pathway	51	103	86	0.0227455	Ariadne Pathways	64336	61
NGFR -> AP-1/CEBPB/CREB/ELK-SRF/TP53 signaling	Pathway	35	38	37	0.0235623	Ariadne Pathways	64439	62
CholinergicRn -> CREB signaling	Pathway	23	47	40	0.0244919	Ariadne Pathways	64237	63
FSHR -> CREB/ELK-SRF/GATA4 signaling	Pathway	29	65	55	0.0260229	Ariadne Pathways	64348	64
TGFBR -> AP-1 signaling	Pathway	16	16	16	0.0262815	Ariadne Pathways	64430	65
Aromatic aminoacid metabolism	Group	125	125	91	0.029165	Ariadne Ontology	63500	66
EphrinR -> STAT signaling	Pathway	15	15	14	0.0305249	Ariadne Pathways	64386	67

11. Choose Tools > Send Data to Excel to save the selected groups and pathways to an Excel spreadsheet.

End Exercise: Experimental Data Analysis – Gene Set Enrichment Analysis

Sub-Network Enrichment Analysis

For every entity in the ResNet database, the Sub-Network Enrichment Analysis (SNEA) algorithm uses the relationships in the ResNet database to build “sub-networks” based on user specified criterion. It then uses these sub-networks and the Fisher’s Exact test or GSEA algorithm to identify the networks that are significantly enriched.

The user-defined sub-networks consist of a single “regulator” gene and its targets. The significance of the target expression levels in every built network is evaluated. The result is the identification of individual “regulators” which most likely affect the differentially expressed genes, thus providing the one plausible explanation for the observed expression changes in the experiment.

Defining the Sub-Networks

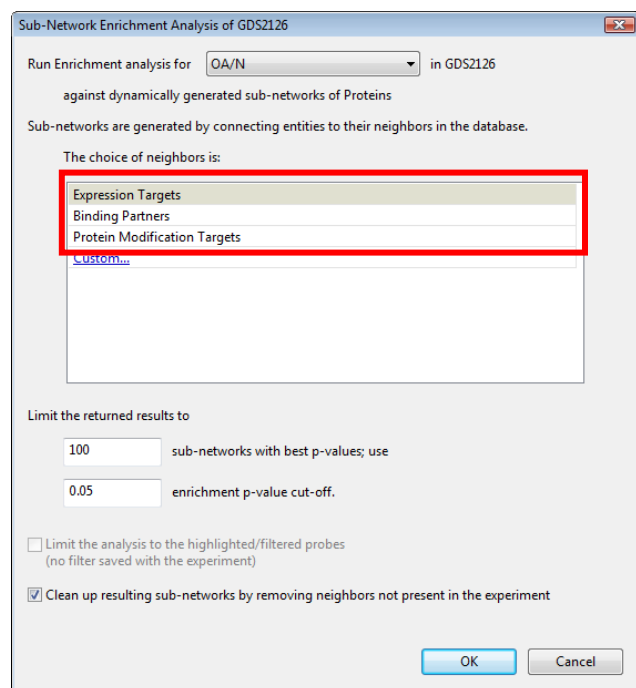
Setting user-defined criteria for building the sub-networks involves first defining the “regulator” (also referred to as the “seed”) and then the “neighbors” (the “targets” defined by selecting specific relationship types).

For example, defining a protein seed and promoter binding relationships will give these types of networks:



The network consists of the target proteins with which the seed protein has a promoter binding relationship.

The Sub-Network Enrichment Analysis menu provides some short-cut menu options (see table below) or you can use the “custom” options for more flexibility in sub-network definitions.



See definitions for these short-cut menu options in table below.

(Desktop Menu)

Note: Ariadne recommends that you check the box “Clean up resulting sub-networks by removing neighbors not present in the experiment.” This way, only measured entities for which experimental data exists (for example proteins/genes in a microarray experiment) are included in the network. (A small molecule or a disease would not have experimental data on a gene expression microarray and would have no expression data associated with it.)

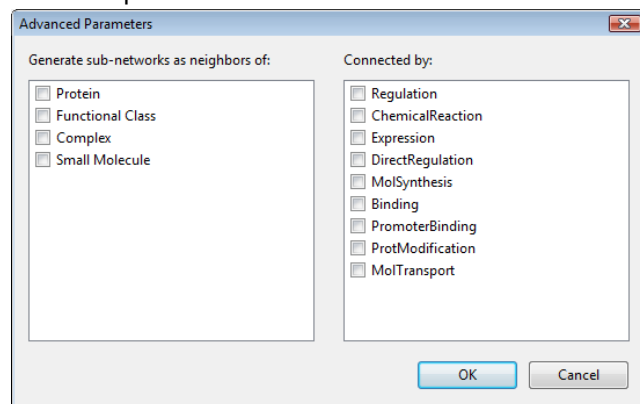
Sub-network preset conditions for short-cut menu:

Sub-network Type	Seed	Neighbors	Relationship	Direction
Expression Targets	Gene	Gene	Promoter Binding, Expression	Outbound from seed
Protein Binding Partners	Protein	Protein	Binding	No direction
Protein Modification Targets	Protein	Protein / Complex	Protein Modification	Outbound from seed
¹ Proteins / Chemicals Regulating Disease	Disease	Protein / Molecule	Regulation	Inbound towards seed
¹ Proteins / Chemicals Regulating Cellular Processes	Cell Process	Protein / Chemical	Regulation	Inbound towards seed

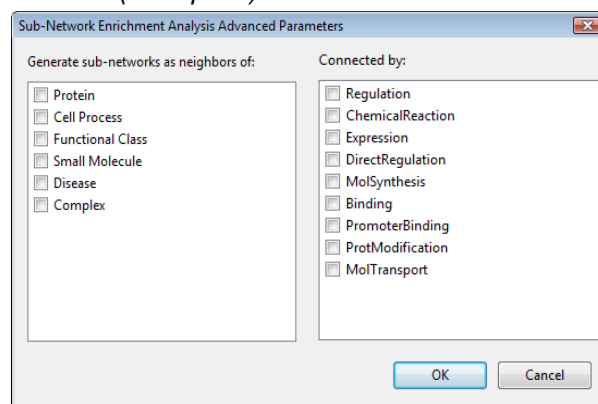
¹Only available in Pathway Studio Enterprise Edition

Use the custom menu to select specific seeds and relationships.

Desktop Custom Menu



PC-Client (Enterprise) Menu

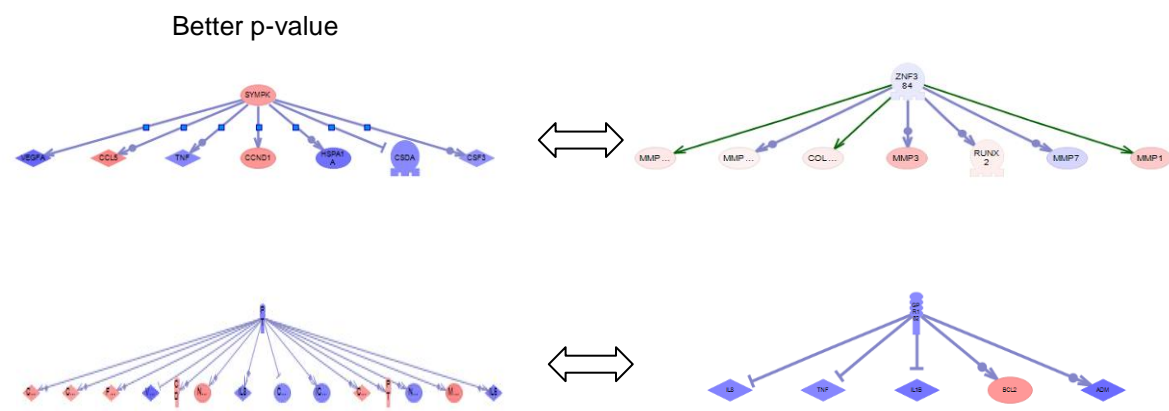


Note: all relationships are outbound from seed to regulators except when the seed is “disease” or “cell process”.

The results of the SNEA are viewed in a table sorted by p-value. Each network is named by the regulator of the network.

Enriched Sub-networks for GDS2126: OA/N									
Name	Type	Total # of Neighbors	# of Measured Neighbors	Gene Set Seed	Measured Neighbors	Median change	p-value	#	
Expression Targets of MAP2K7	Pathway	21	20	MAP2K7	FOS, CSTA, MYOG, IL6, CDH2, CDKN2A, IL2, CCL5, T...	1.14409	0.000269647	1	
Expression Targets of SPARC	Pathway	23	22	SPARC	MMP2, MMP3, MMP9, CDH3, CCNA2, SERPINE1, IGF...	1.56278	0.000390552	2	
Expression Targets of PYCARD	Pathway	9	9	PYCARD	IL1B, IL6, TNF, NFKBIA, IL8, LPL, MYC, FABP4, CD44	-2.36761	0.000478811	3	
Expression Targets of STN	Pathway	67	59	STN	CD80, SCARB1, MAPK14, F3, NOS3, LDLR, PON1, CRP...	1.11846	0.000640743	4	
Expression Targets of neurohorm...	Pathway	35	33	neurohormone	TAC1, ICAM1, FOS, PRL, CALCA, TNF, IL6, IL18, MMP...	-1.20613	0.000687735	5	
Expression Targets of SAMHD1	Pathway	6	5	SAMHD1	VCAM1, IL8, IL6, CXCL10, ICAM1	-2.19564	0.0012632	6	
Expression Targets of HSPB1	Pathway	22	20	HSPB1	TNF, SOD2, FGF2, IL10, MCAM, MMP9, HSPA1A, PTG...	1.41719	0.00143174	7	
Expression Targets of ATF4	Pathway	37	35	ATF4	IL8, IL6, CCL2, CCNA2, HRAS, SCG2, PENK, ATF5, VEG...	-1.0273	0.00182389	8	
Expression Targets of DCN	Pathway	22	21	DCN	CDKN1B, CDKN1A, DUSP1, MMP14, CTNNB1, ACTG2...	-1.35642	0.00183731	9	
Expression Targets of CCR1	Pathway	6	5	CCR1	TNF, CCR5, MMP9, CCR6, CCR2	3.26935	0.00204477	10	

When networks are the same size, those with greater differential gene expression will have better p-values.
 When differential expression is equal, larger networks will have better p-values than smaller networks.



Both enrichment algorithms, Fisher's Exact Test and Gene Set Enrichment Analysis, can be used to identify enriched Sub-Networks.

	Known Gene Sets (ontologies, curated pathways)	Sub-Networks (user defined from ResNet)
Fisher's Exact Test (experimental values <u>not</u> utilized)	Find Pathways/Groups Enriched with Selected Entities	Find Sub-Networks Enriched with Selected Entities
Gene Set Enrichment Analysis (GSEA) (experimental values <u>are</u> considered in the analysis)	Gene Set Enrichment Analysis	Sub-Network Enrichment Analysis

Names of the tools as they appear in Pathway Studio

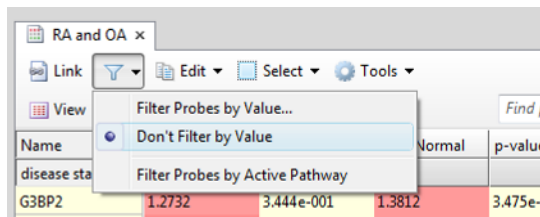
● Exercise Eight: Experimental Data Analysis – Sub-Network Enrichment Analysis (with GSEA)

Begin Exercise:

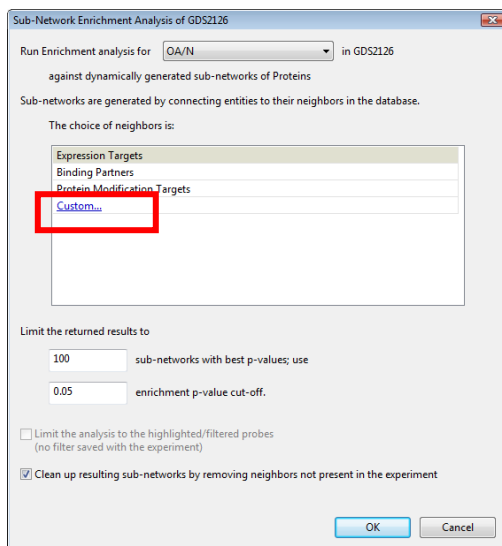
Objective: To run the Sub-Network Enrichment Analysis on a gene expression dataset to identify dynamically defined transcriptional regulatory networks enriched in the dataset and to visualize multiple networks in one view to allow for identification of overlap between networks.

Let's use the osteoarthritis and rheumatoid arthritis synovial tissue data set to run the Sub-Network Enrichment Analysis and find transcriptional regulatory networks enriched in the experimental dataset.

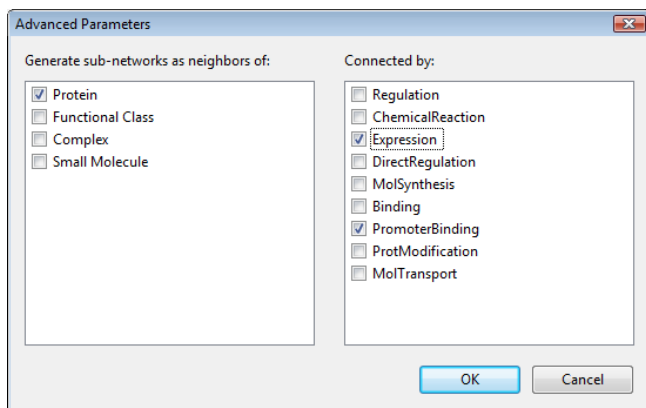
1. If a filter has been applied to the experimental dataset, remove the filter before continuing.



2. Select Tools > Sub-Network Enrichment Analysis
3. In the Run Enrichment analysis menu, select the osteoarthritis study.
4. In the choice of neighbors menu select "Custom"



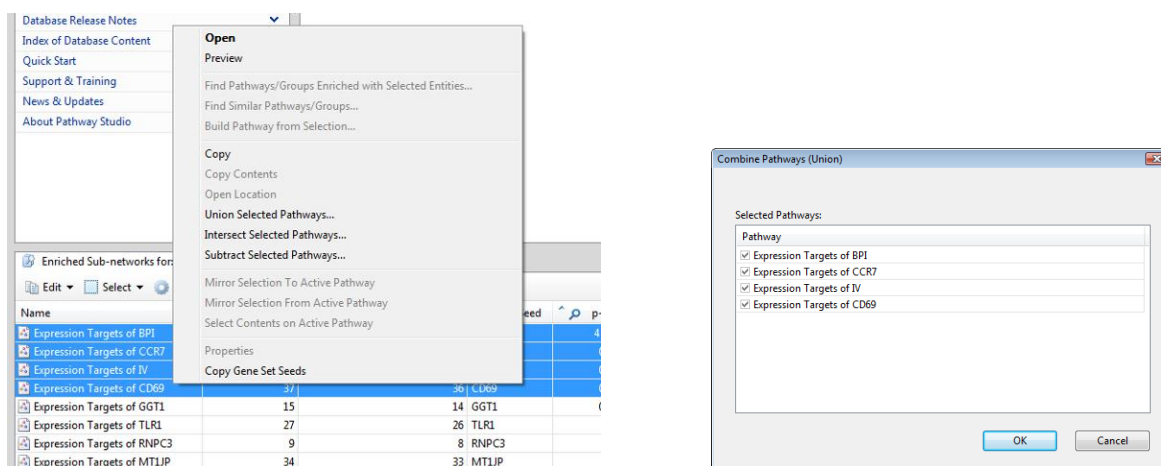
5. Here we will define transcriptionally regulated networks: Select Protein for the (transcriptional) regulator "seed" and select Expression and PromoterBinding to define the transcriptionally regulated "target" network. Note: We will not utilize Regulation in this example, as this is the largest relation category in ResNet.



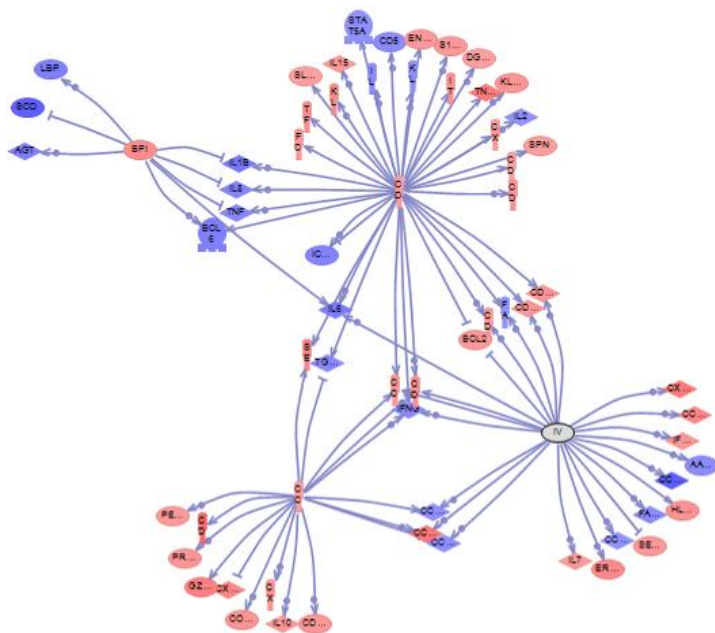
6. Select OK.
7. In the Sub-Network Analysis tool ensure that the box “Clean up resulting sub-networks by removing neighbors not present in the experiment”, then select OK to run the analysis.
8. View the results in the List Pane. Each network is named by the “seed” (the regulator).

Enriched Sub-networks for: OA vs Normal							
Edit Select Tools							
Name	# of Entities	# of Measured Neighbors	Gene Set Seed	p-value	ID	#	
Expression Targets of BPI	9	8	BPI	4.44579e-005	900000001	1	
Expression Targets of CCR7	18	17	CCR7	0.000193931	900000002	2	
Expression Targets of IV	24	23	IV	0.000641278	900000003	3	
Expression Targets of CD69	37	36	CD69	0.000666078	900000004	4	
Expression Targets of GGT1	15	14	GGT1	0.000833474	900000005	5	
Expression Targets of TLR1	27	26	TLR1	0.00136019	900000006	6	
Expression Targets of RNPC3	9	8	RNPC3	0.00164947	900000007	7	
Expression Targets of MT1JP	34	33	MT1JP	0.00179706	900000008	8	
Expression Targets of DCN	30	29	DCN	0.00220177	900000009	9	
Expression Targets of jararhagin	7	6	jararhagin	0.00232483	900000010	10	

9. Let's view the top four networks in one pathway. Select the top four networks, right-click and choose “Union Selected Pathways.” In the “Combine Pathway (Union)” window, select OK.



The resultant network should look something like this (using the dynamic layout). Note: although your network may look similar, the dynamic layout may generate a slightly different layout than the one see here.

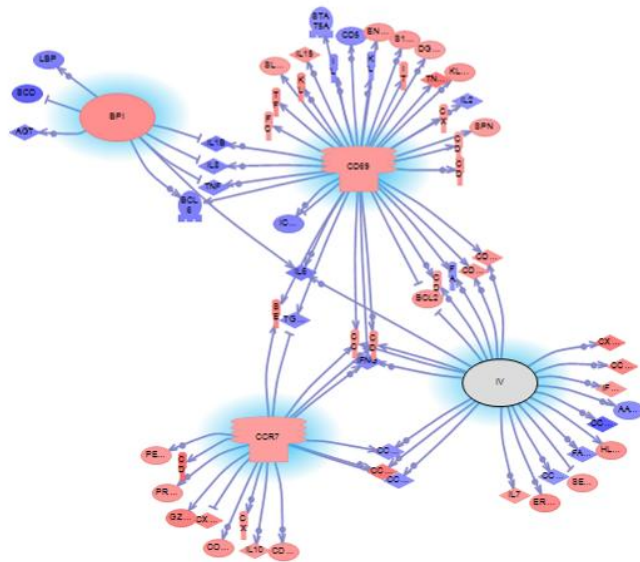


Now let's modify this network to better view the entity names and emphasize the four regulators in this network by applying highlighting.

10. Select "Fit all entities to window" to maximize the size of the network in the window.
11. Open the Advanced Visualization tool bar and select "Fit text to entities."
12. Next, using the control key, select the four regulators, BP1, CCR7, IV and CD69. Note that selected entities have a blue halo. One of the selected entities will also have small white boxes around it. Click on the corner white box and expand the size of this entity.



13. With the other three entities (regulators) still selected, go to the Advanced Visualization tool bar and select "Set the same size for selected entities." This will expand all the regulators to the same larger size.
14. While the four regulators remain selected, go to the highlight menu and select the blue highlight. This will put a blue halo around the regulators. Your network should look similar to this:



Note that one of the four regulators is gray. This is because this regulator was not measured in the gene expression experiment. The Sub-Network Enrichment Analysis is able to utilize all the relationship information in ResNet to identify potentially important regulators for a gene expression experiment, even if that specific regulator was not included on the microarray.

End Exercise: Experimental Data Analysis – Sub-Network Enrichment Analysis (with GSEA)

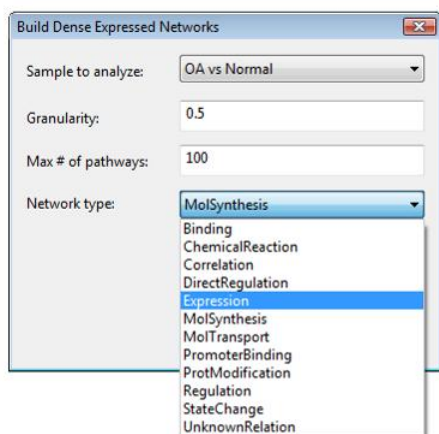
Build Dense Expressed Networks (PC-Client, Enterprise only)

The Build Dense Expressed Networks (BDEN) algorithm searches the ResNet network of interactions for dense expressed clusters. A network cluster consists of genes or metabolites that are more densely interconnected with each other through the ResNet relations than with the rest of the network. At the same time, the clusters also contain significantly differentially expressed genes.

The BDEN algorithm can be run only on ratio (log-ratio) expression data. Absolute values of expression values are used in calculations. The algorithm does not take into account direction and effect of the relations. Also note that self-regulatory loops are discarded from the network.

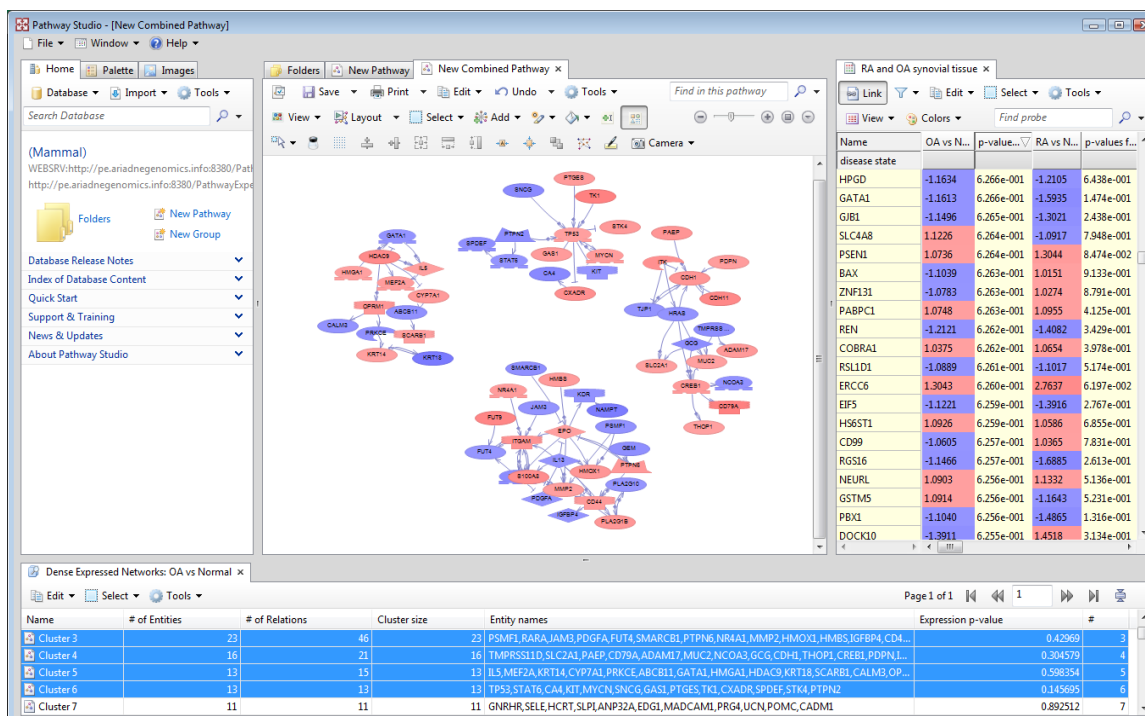
The BDEN algorithm requires several parameters that should be specified in the Build Dense Expression Networks dialog:

- **Sample to analyze** should be set to one of the ratio columns of the Expression Experiment;
- **Granularity**. The optimal value for this parameter depends on expression data, network type and desired cluster sizes. We recommend starting from larger values (greater than one). Higher granularity values give smaller clusters, lower granularity values give larger clusters.
- **Max # of pathways** is the maximum number of clusters in the output;
- **Network Types**. BDEN runs on network constructed from all relations of a specified type. Allowed relation types are: PromoterBinding, Binding, DirectRegulation, Expression, ProtModification, MolSynthesis, and MolTransport.



Results

The algorithm outputs several sub-networks (clusters) sorted by number of edges, the top clusters are denser and generally larger. The resulting clusters are annotated by expression *p*-values and edge densities. The expression *p*-values are calculated using Mann-Whitney U-test in same way as in the Find Differentially Expressed Genes algorithm. The edge density of a cluster is the fraction of all possible edges included in the cluster.



Applications

Depending on the selected relation type, the reported clusters may have different biological interpretations. For example, if Binding is selected, then reported clusters could contain expressed genes corresponding to protein complexes.

Expression and PromoterBinding network options can be used for analysis of the gene expression data to find major regulators, regulatory cascades.

MolSynthesis and MolTransport network options can be used for analysis of the metabolomics data to find major metabolism regulators, regulatory metabolism cascades.

● Exercise Nine: Experimental Data Analysis – Build Dense Expressed Networks (PC-Client, Enterprise only)

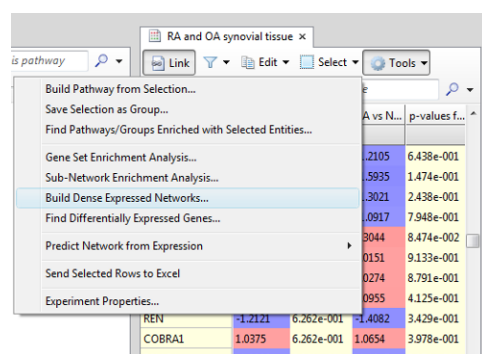
Begin Exercise:

Objective: To run the Build Dense Expressed Networks analysis on a gene expression dataset to identify dense expression networks in the results dataset.

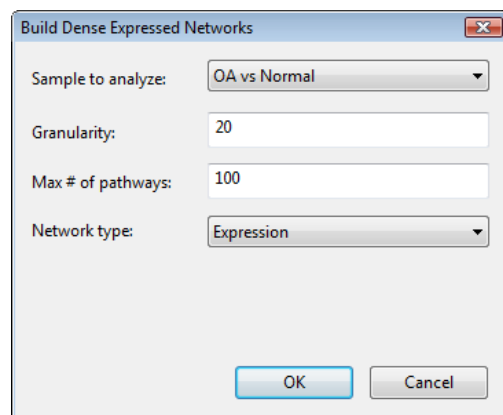
Build Dense Expressed Networks (BDEN) is available only through the PC-Client (Enterprise Edition).

Let's use the osteoarthritis and rheumatoid arthritis synovial tissue data set to run the Build Dense Expressed Networks analysis and find regulatory networks enriched in the experimental dataset.

1. Launch the Build Dense Expressed Networks tool from the Tools menu.



2. Select: Sample to Analyze: the osteoarthritis data set; Granularity of 20 and Network type: Expression.



3. When the results appear in the List Pane, examine the p-values for the clusters.
4. Next try increasing the granularity. What does that do to the resultant cluster size? What if you decrease the granularity?

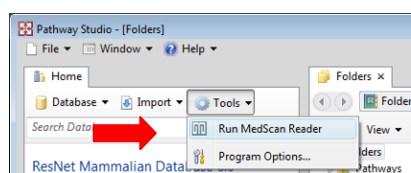
End Exercise: Experimental Data Analysis – Build Dense Expressed Networks

Section 6: MedScan Reader

MedScan technology is a Natural Language Processing Technology used to extract relationship information from biomedical literature. The MedScan technology is used by Ariadne to create and update the ResNet Mammalian database by extracting information from all PubMed abstracts and 61 full-text journals. (See the introduction in this manual for more information about fact extraction from the literature by MedScan.)

Pathway Studio users have access the MedScan technology through the MedScan Reader. Thus you can add additional information to your ResNet database from sources that you have access to (ex. by subscription to a journal, or your own internal documents or additional searching through web-based information sources.)

The MedScan Reader can be launched from the Tools menu in the Information Pane.



Using MedScan to search internet resources:

You can use MedScan Reader to extract information from literature source accessible from an internet browser. This includes commonly utilized sites such as PubMed, Google Scholar, HighWire Press etc. This is easily done by accessing these resources through the MedScan Reader browser view, which we will examine below in the PubMed example.

Selecting Reading Settings: dictionaries and relation extraction

Prior to utilizing MedScan Reader to extraction information, review the reading settings. In the Project Home tab, select “Change Reading Settings” to view the selected dictionary and extraction algorithm.



There are multiple different cartridges (dictionaries and grammar extraction rules) that are available to use Pathway Studio 6.2/MedScan Reader 2.2. For our examples we will use the mammalian dictionary with the ResNet 6 Mammalian database.

Project Home

Active MedScan Cartridge Configuration

This option defines what MedScan configuration is used for information extraction.

Select **Standard Cartridge (Human, Mouse, Rat, ...) v3.1.0**

Protein name dictionaries: Mammals (Human, Mouse, Rat)

Relation extraction: **Extract relations**

Save Changes

Select Plants Cartridge v3.1.0

There are two relevant relation extraction types available:

- **Extract Relations:** This is the most commonly used option. This extracts relations identified between two entities in a given sentence.
- **Extract Co-occurrence:** This setting will identify when two entities exist in the same sentence even if no relation between them is identified. In this instance the “relation” between them will be indicated as “Unknown Relation.” (This is a much less stringent extraction filter and will result in more “noise” in your extracted results.)
- (Extract Basic Relations gives fewer results for small molecule relations than does Extract Relations.)

Project Home **Import**

Active MedScan Cartridge Configuration

This option defines what MedScan configuration is used for information extraction.

Select **Standard Cartridge (Human, Mouse, Rat, ...) v3.1.0**

Protein name dictionaries: Mammals (Human, Mouse, Rat)

Relation extraction: **Extract relations**

Extract relations

Extract cooccurrence facts

Save Changes

Select Save to return to the Project Home page.

Select Reading Interests

Select “Set My Reading Interests” on the Project Home page to view options about applying inclusive and exclusive filters and flagging specific entities of interest in your results table.

MedScan Reader

- [Change Reading Settings](#)
- **[Set My Reading Interests](#)**
- [Digest Pane Legend](#)

The Track Entities via Highlighting option will flag the entities you designate here in the MedScan search results table. In this example a gene, BRCA1, has been selected to be flagged in the results table.

Project Home

Track Entities via Highlighting

Enter one or more terms separated by commas, e.g. p53, p300

Add

BRCA1


Project Home

Track Entities via Highlighting

::

BRCA1

Delete




 Protein **BRCA1** (urn:agi-llid:672, msid:672, BRCA1)

















































1 entities recognized

Add

Tracked entities are highlighted by red flags in the Relations summary results table for easy identification.

Current Page Digest

  View 

	BRCA1		 Breast Neoplasms
	BRCA1		 PALB2
	BRCA1		 Neoplasms
	BRCA1		 DNA repair
	BRCA1		 Neoplasms
	BRCA1		 Ovarian Neoplasms
	TP53		 cell growth
	BRCA2		 PALB2
	BRCA2		 Ovarian Neoplasms
	BRCA1		 cell cycle
	anti-oncogene		 Breast Neoplasms
	BARD1		 BRCA1
	BRCA1		 DNA recombination
	anti-oncogene		 DNA repair
	anti-oncogene		 cell cycle
	TP53		 Breast Neoplasms

338 Relations

787 Entities

Here two entities, brca1 (a gene/protein) and breast cancer (a disease), are designated, as well as three relationship types: regulation, promoter binding and expression.

All the results of this MedScan query will highlight these entities. In addition, only regulation, promoter binding and expression relations including at least one of these entities will be extracted.


See that each relationship identified has *brca1* or breast cancer as one of the entity types, and all relations are either regulation, promoter binding or expression.

Relations

- ≡ Binding
- ⇐ Regulation (negative)
- Regulation
- Regulation (positive)
- ⇐ Control of Expression (negative)
- ⇐ Control of Expression
- Control of Expression (positive)
- ⇐ Promoter Binding
- Molecular Transport
- ⇐ Molecular Synthesis
- Protein Modification
- ≡ Direct Regulation (negative)
- Direct Regulation (unknown or positive)
- Other relations (chemical reaction, cell object control, ...)

Here is the key to the relationship type symbols as displayed in MedScan.

The Disregard Entities and Relations filter is used to exclude specific entities or relationship types from the MedScan results.

Disregard Entities and Relations 

Enter one or more terms separated by commas, e.g. p53, p300

Relation Types	<input type="checkbox"/> Binding	<input type="checkbox"/> Promoter Binding	<input type="checkbox"/> Control of Expression
<input type="checkbox"/> Cell Object Control	<input type="checkbox"/> Mol. Transport	<input type="checkbox"/> Mol. Synthesis	
<input type="checkbox"/> Protein Modification	<input type="checkbox"/> Chemical Reaction	<input type="checkbox"/> Direct Regulation	
<input type="checkbox"/> Regulation	<input type="checkbox"/> Possible Relation		

Search PubMed through MedScan Reader browser

Once you have selected your reading settings you are ready to run MedScan Reader. MedScan Reader can be utilized to scan internet sites and extract information. Links to commonly used biological resources are provided for your convenience in the Popular Destinations section.

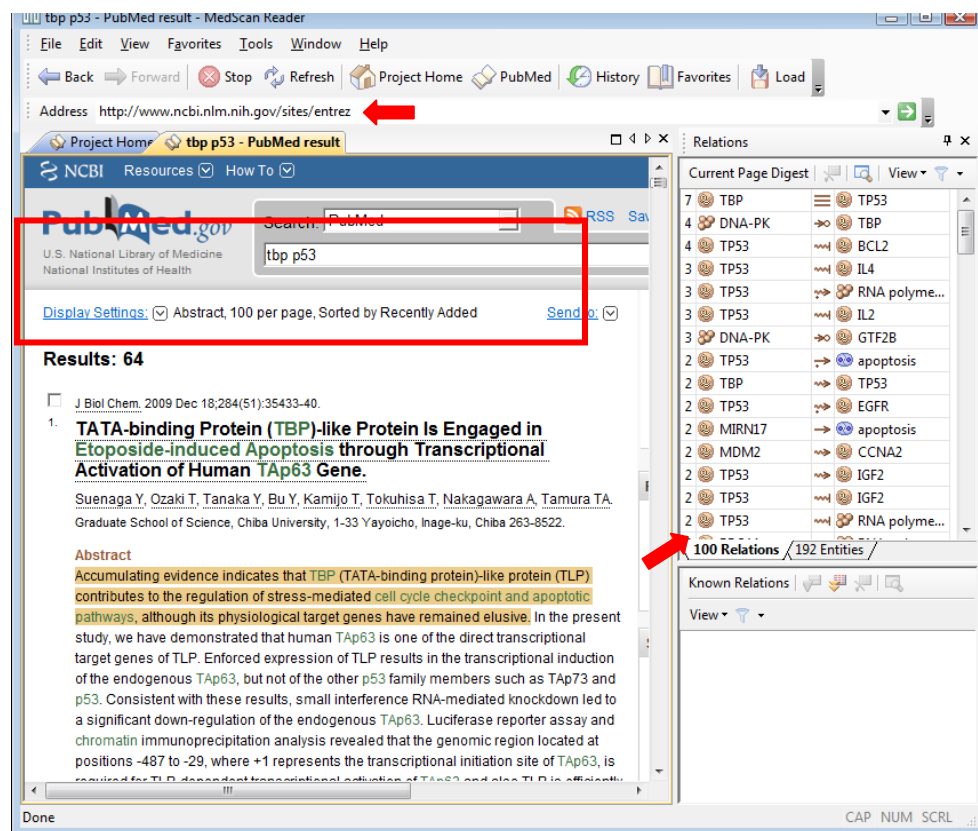
Popular Destinations

- [Search PubMed](#)
- [Search BioMed Central](#)
- [Search HighWire Press](#)
- [Search Google Scholar \(beta\)](#)
- [Search Google](#)

When doing a search with MedScan through an internet site sometimes there are size limitations on the amount of information you can access at one time. For example, PubMed can only display 500 abstracts at one time. This is a browser limitation of PubMed and not a MedScan limitation.

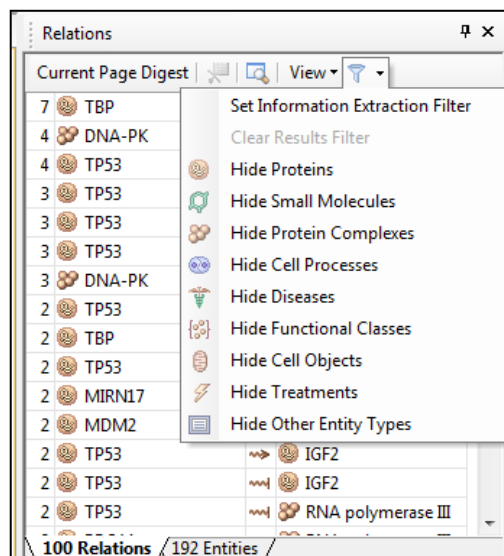
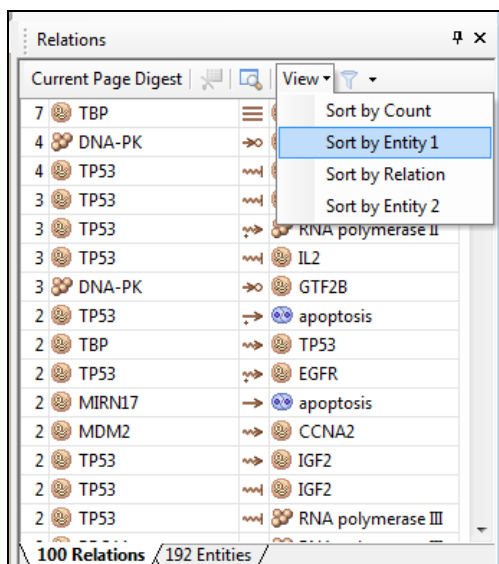
When the Search PubMed link is selected in the Popular Destinations section, the PubMed page is opened within the MedScan browser. Any information displayed in the browser will be read by MedScan and the results are summarized in the Relations box in the upper right. Select to display the abstracts to maximize the number of facts extracted. (If only the Summaries are displayed in the browser, only the summaries are read by MedScan.) In addition, select the number in “show” to be high enough to display all the identified abstracts on one web page. Recall that MedScan will only read what is displayed in the browser page.

Each time the PubMed screen redraws, a pop-up window appears:

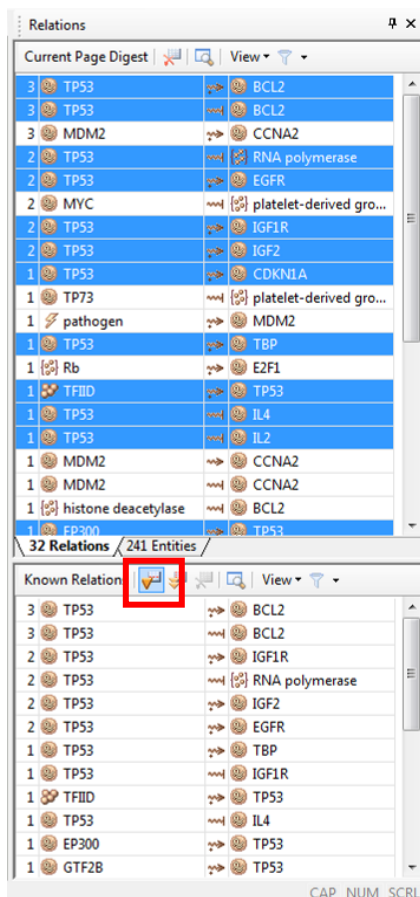


Here you can see in the browser window displaying the PubMed search results that entities recognized by MedScan are highlighted in green. Sentences that contain extracted relationships are highlighted in gold. A summary of all the extracted information for these abstracts (shown above) can be seen in the Relationship table in the upper right.

The summary Relationship table can be sorted by entity, count and relation columns. In addition filters can be applied to this view (see below).

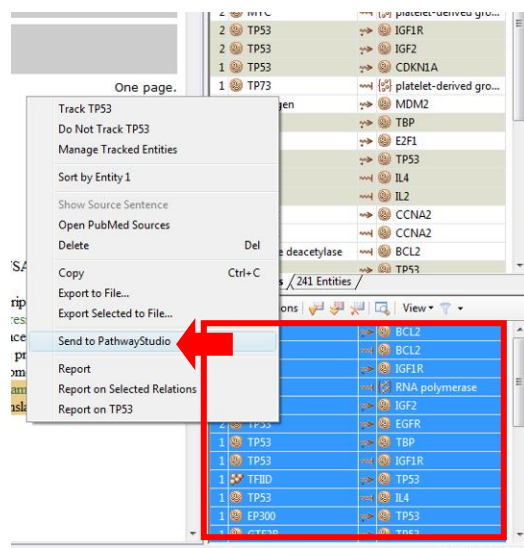


The summary Relation table is a transient summary and will change each time the information displayed in the browser view changes. To save facts listed in the transient summary Relation table, select the desired relationships and add them to the bottom persistent Known Relations table.



Select the desired relations in the transient summary table above and select "Add Selected" to move them to the persistent Known Relations table.

Once you have completed all your MedScan fact extraction activities and collected all the desired relationships in the lower persistent table, you can send this list of relationship information to Pathway Studio and build a network.



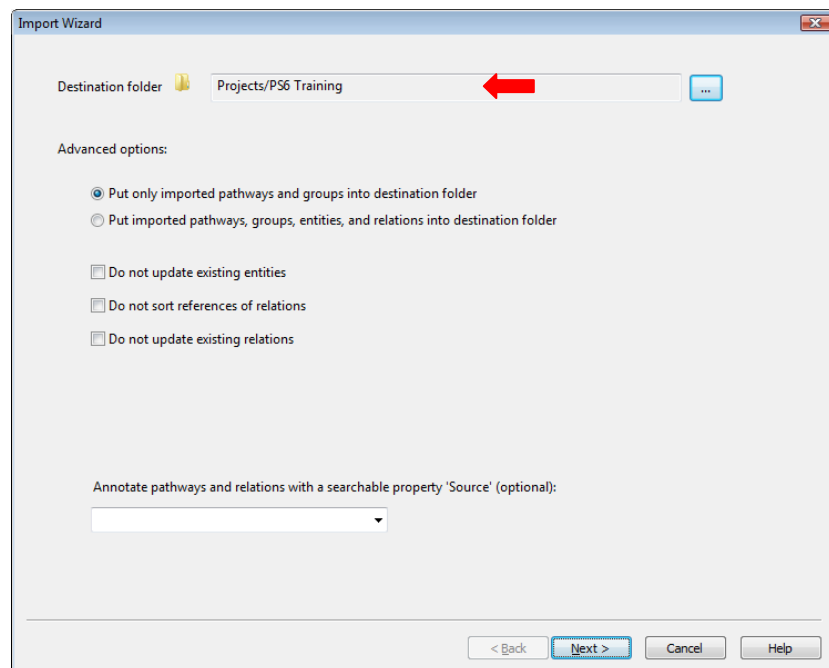
Select the relationships in the persistent box, right-click and select “Send to Pathway Studio.”

Note: BEFORE you do this, make sure that you do not have an active pathway view open in Pathway Studio, because these relationships will be added to the active pathway.

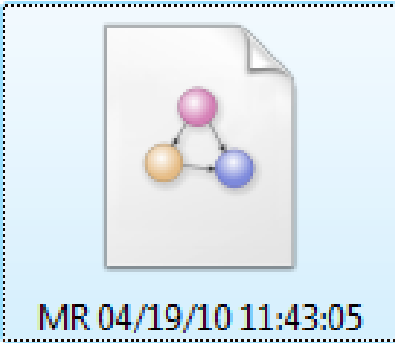
Therefore, before you send this information to Pathway Studio, it is best to open a new pathway window.

Note: For Desktop users, when you send these selected relationships to Pathway Studio, you are adding them also to your local ResNet database. For Enterprise users, you are adding this information to your private space.

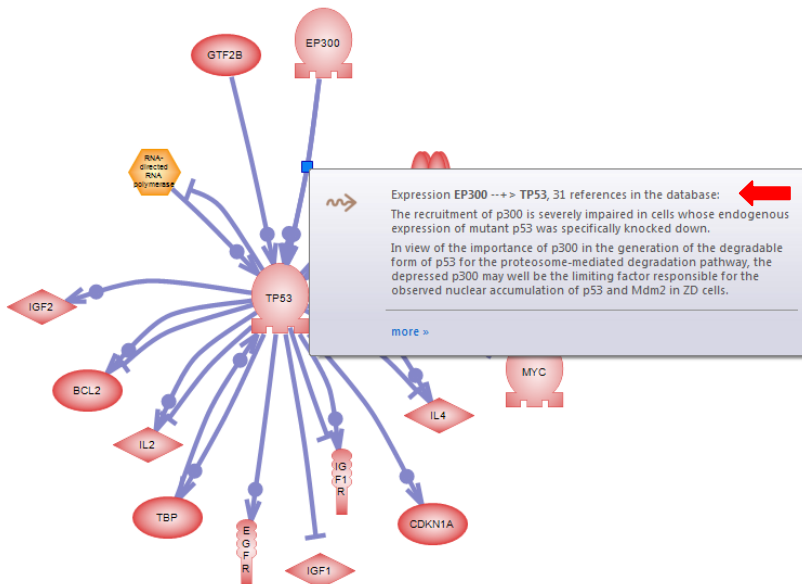
The import wizard will appear. Provide a destination folder for the network you are about to import.



The thumbnail for the imported network will not show a pathway image until it is opened and then saved. The file will be named “MR” and appended with a date and time it was generated. You can change the name of the file at any time. Simply right-click on the icon and select “rename.”



Double-click on the network icon to open it and examine the network.



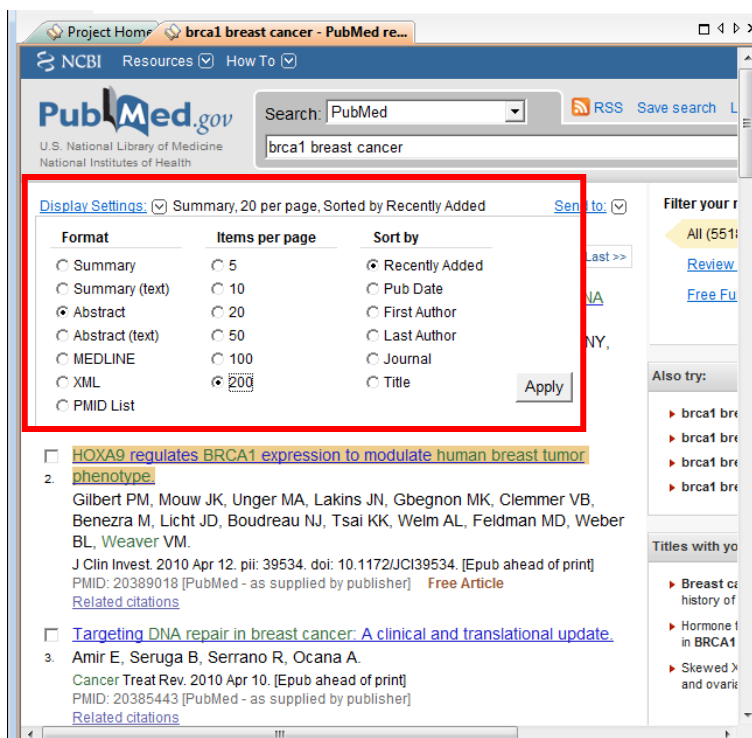
When examining relationship reference information in this network, not only will the newly identified relationship reference be listed, but if there existed in ResNet already other references that support the relationship, those pre-existing references will also be available for review.

● Exercise Ten: Use MedScan to extract facts from the results of a PubMed query – Popular Destinations

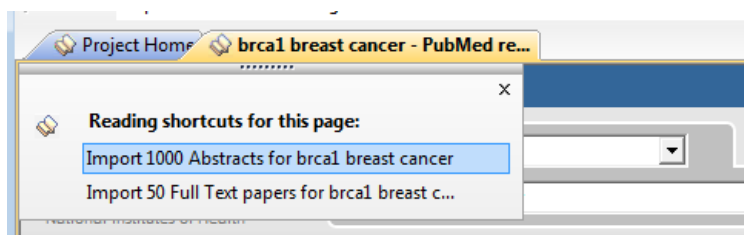
Begin Exercise:

Objective: To become familiar with using MedScan to extract information from web-base literature sources, using PubMed as an example

1. Select: Tools > MedScan Reader. The MedScan application will launch.
2. On the Project Home page, select “Change Reading Settings”.
3. Select the mammalian protein name dictionary and the “Extract Relations” extraction setting.
4. Select “Save” to close the reading settings.
5. From the Popular Destinations folder, select “Search PubMed”. The browser window will display the PubMed launch page.
6. In the PubMed query field input: “brca1 breast cancer” and select “Go”.
7. When the query has completed, change the Display from Summary to Abstracts and the Show from 20 to 200.



The pop-up window offers the option of extracting relations from 1000 abstracts.



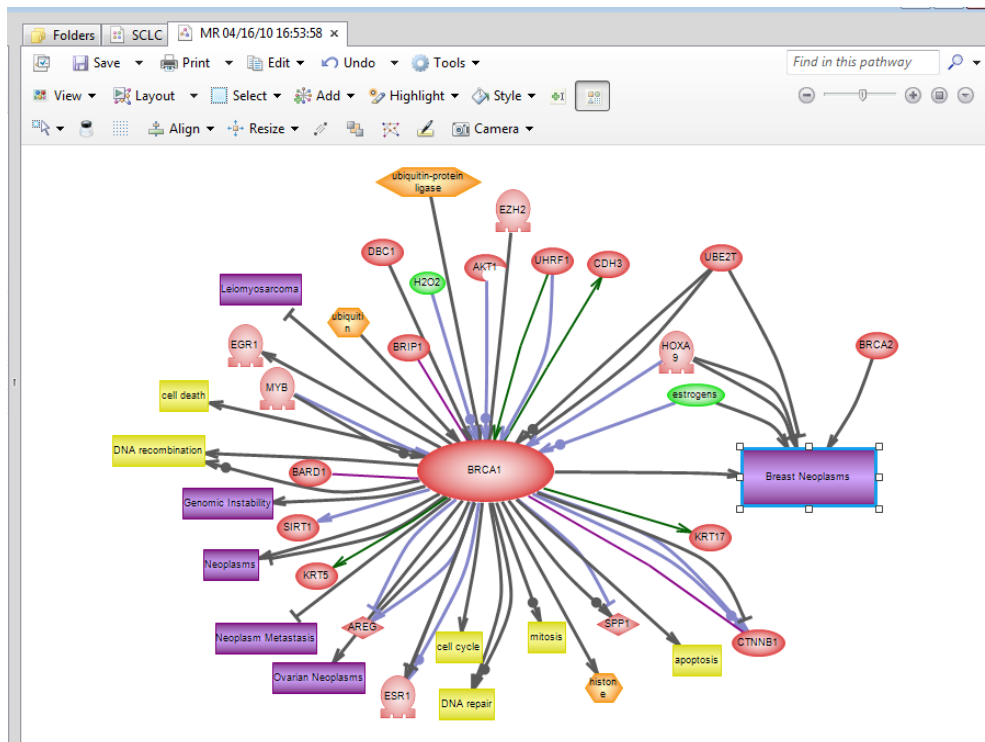
8. View the summary of extracted relationships in the Relations table (upper right). Note the number of relations identified (for the following exercise). In this example 100 relations were identified (and this number will increase as more publications appear in PubMed over time).

The screenshot shows a PubMed article titled "HOXA9 regulates BRCA1 expression to modulate human breast tumor phenotype" by Gilbert PM, Mouw JK, Unger MA, Lakins JN, Gbegnon MK, Clemmer VB, Benezra M, Licht JD, Boudreau NJ, Tsai KK, Weim AL, Feldman MD, Weber BL, Weaver VM. The abstract discusses the role of HOXA9 in breast cancer. To the right, a 'Relations' table displays 100 extracted relationships. A red box highlights this table, which shows pairs of entities connected by arrows, representing the extracted relationships from the text.

9. Sort the results by the first entity and select all relations that contain "brca1" and "breast cancer" in the first column. Add these selected relationships to the persistent Known Relations table at the bottom using the "Add Selected" arrow.
10. Sort the results by the second entity and select all relations that contain "brca1" and "breast cancer" in the second column. Add these selected relationships to the persistent Known Relations table at the bottom using the "Add Selected" arrow.
11. Select all the relations in the Known Relations table, right-click and select "Send to Pathway Studio". (Recall, if you currently have an active pathway open in Pathway Studio, this information will be added to that pathway so it is advisable to open a new pathway window before performing this step.) The Import Wizard window will appear as this data is being imported into your local ResNet database (Desktop users).

The screenshot shows the 'Relations' table with a context menu open over the 'Known Relations' section. The menu includes options for tracking and managing BRCA1-related entities, as well as options for exporting and sending data to Pathway Studio. The 'Send to PathwayStudio' option is highlighted, indicating the next step in the workflow.

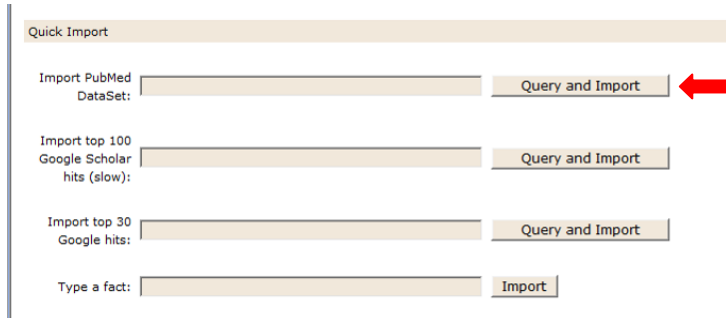
12. View your resultant network in Pathway Studio.



End Exercise: Use MedScan to extract facts from the results of a PubMed query – Popular Destinations

Quick Import of PubMed Search

The Quick Import options allow you to by-pass displaying information in a browser format. This is a faster method for fact extraction.



Quick Import

Import PubMed DataSet:

Import top 100 Google Scholar hits (slow):

Import top 30 Google hits:

Type a fact:

The Quick import allows for extraction from 1000 PubMed abstracts at one time, twice as many as the PubMed browser window.

Note that in this instance only abstracts are read, but titles are not.

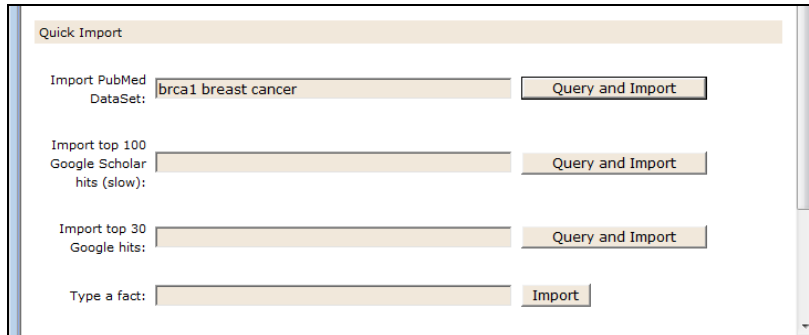
When your PubMed search produces very large numbers of abstracts (tens of thousands) you can still use MedScan to import that information into Pathway Studio to provide you a summary of this information. Keep in mind that most of these relationships will already exist in ResNet (expect possibly the most recently published facts), however, it still may be useful to see a summary of the literature for a particular topic defined in your query.

● Exercise Eleven: Use MedScan to extract facts from the results of a PubMed query – Quick Import

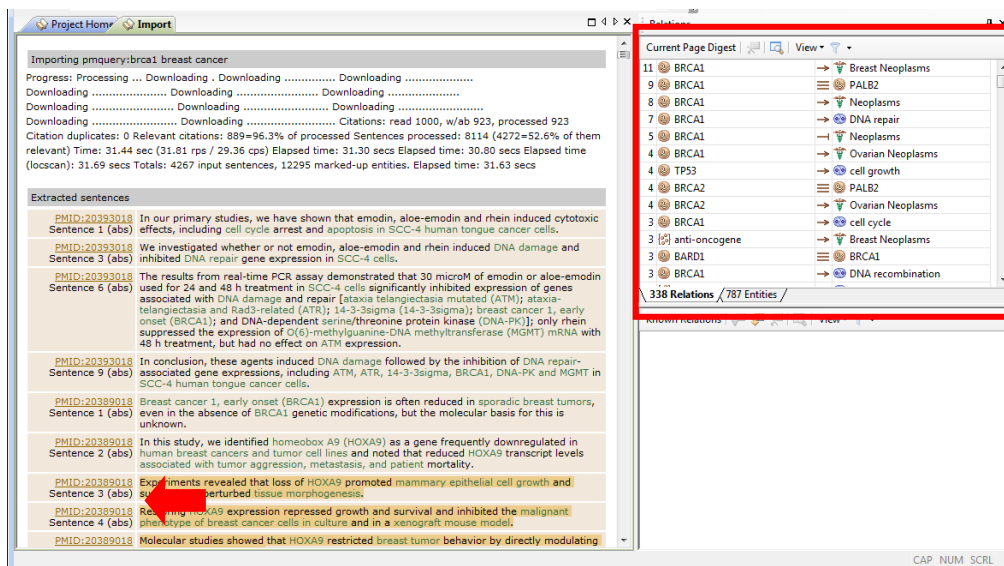
Begin Exercise:

Objective: To use the Quick Import to perform a MedScan extraction of a PubMed query.

1. In the Quick Import section on the Project Home page, in the “Import PubMed Dataset” input the same query as was used above: brca1 breast cancer.



2. When the results are provided, examine the number of relations identified. This number obtained here can be smaller than the browser query for less than 200 abstracts because the Quick Import does not extract information provided in titles as happens in the browser window, but instead only extracts information from the body of the abstract.



3. Find a sentence that has a relation identified (highlighted in gold). Select the link to the PubMed ID (PMID) and click to view that reference.
4. When the PubMed reference page opens, notice that the number of relations in the upper right Relation table has changed. Why?

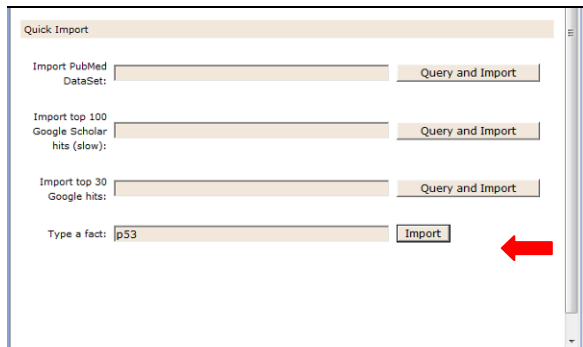
End Exercise: Use MedScan to extract facts from the results of a PubMed query – Quick Import

● Exercise Twelve: Query the MedScan Dictionary for an Entity

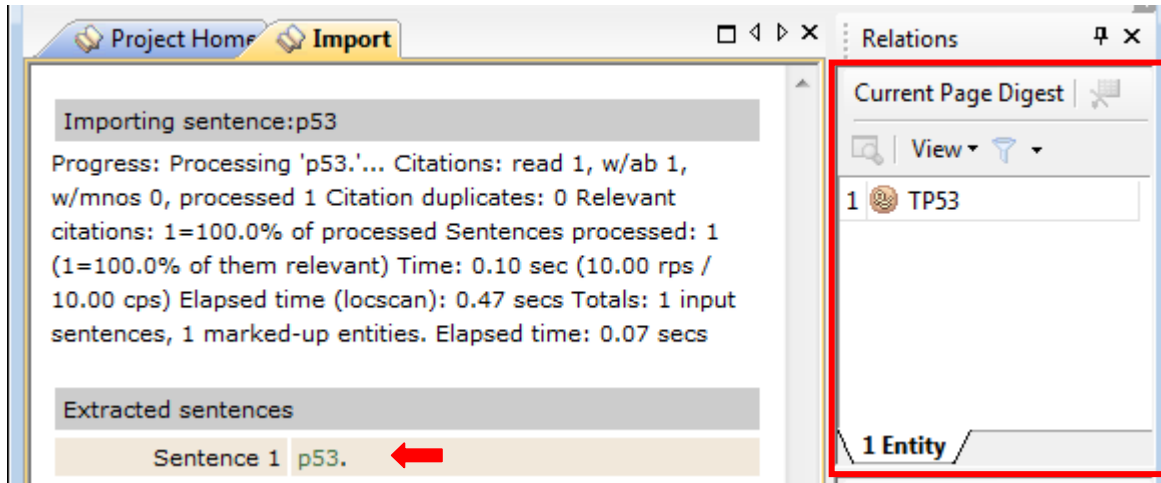
Begin Exercise: Query the MedScan Dictionary for an Entity

Objective: To use Type a Fact to query the MedScan dictionary for the presences of a desired Entity name or alias.

1. On the Project Home page, type the name of an entity, “p53” into the Type a Fact box and select “Import.”



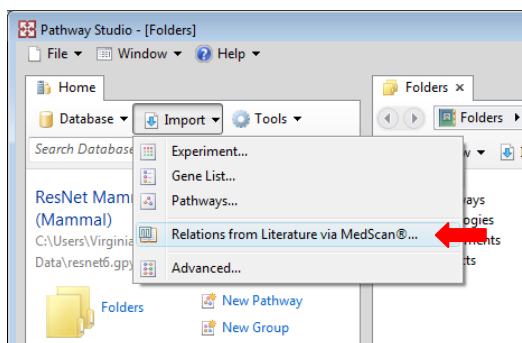
2. In the results view (Import tab) examine the marked up entity under “Extracted Sentences”. If the input entity name is highlighted in green then the entity name is in the MedScan dictionary and thus MedScan can recognize and extract facts associated with this entity in the literature.



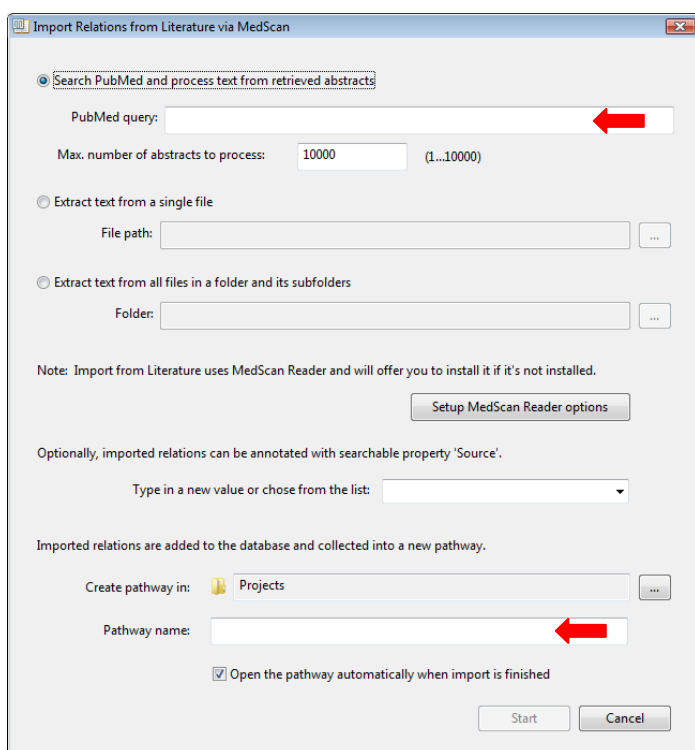
3. Next, examine the Current Digest Pane (the summary of extracted information) on the right. Does the name of the entity shown here match the marked up entity on the left? If not (as is the case in the example here) then the marked up entity name is listed as an alias and the name shown in the Current Digest Pane is the display name for the entity. The MedScan dictionary contains both names and aliases for entities.

End Exercise: Query the MedScan Dictionary for an Entity

To search up to 10,000 extracts at one time, launch MedScan from the Import menu and select “Relations from Literature via MedScan.” Note, the process of scanning ten thousand abstracts at one time takes longer than the quick import.



Provide the PubMed query and the pathway name for the resultant network.



Note: for extracting information from more than 10,000 abstracts, use a date limit search string to process consecutive units of time (and make the units small enough so each contains less than 10,000 abstracts). For example, to find abstracts that mention curcumin published only in the years 2003-2004 input the following into the PubMed query field: curcumin 2003[pdat]:2004[pdat]. (PDAT = publication date).

Go here to for more information about PubMed search descriptions.

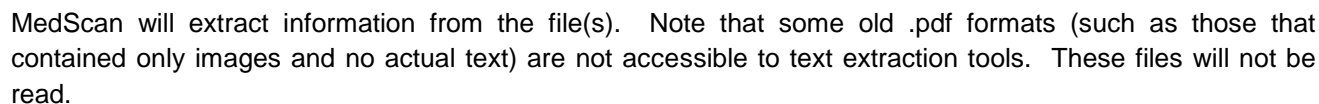
http://www.ncbi.nlm.nih.gov/books/bv.fcgi?rid=helppubmed.section.pubmedhelp.Search_Field_Descrip

When you select File > Save Project, only the facts collected in the persistent box are saved with the project.



Import a document or directory

To extract information from a document on your local drive, from the file menu select “Import Document.” Alternatively, if you would like to extract information from multiple documents at one time, if all the documents are in one directory, select “Import Directory of Documents.”



Appendix A: Definitions

Entity definitions:

Protein: The principal source of proteins and their annotation in ResNet is Entrez Gene. The level of detalization of proteins in ResNet Mammal is the Gene. This means that if proteins are encoded by the same gene, they will have the same identifier.

Small molecule: Symbolizes metabolites, drug and other chemicals of low molecular weight (< 1 KDa). It also can be used to represent non-biological polymers of larger molecular weight.

Cell object: Symbolizes cellular organelles and other sub-cellular components. A majority of cell process entities coincide with the part of the Gene Ontology cellular component classification that excludes protein complexes.

Cell process: Symbolizes biological processes. A majority of Cell process entities coincide with Gene Ontology biological processes classification.

Disease: Symbolizes diseases and other health conditions and processes.

Treatment: Symbolizes non-chemical treatments and environmental conditions such as cold shock, draught etc.

Complex: This is a container entity that symbolizes several polypeptides that form the complex via physical interaction. The complex is usually well-characterized in the literature, performs well-defined function and is referred in the literature by a specific name. A majority of complex entities coincide with part of Gene Ontology cellular classification that describes protein complexes.

Functional class: This is a container entity that symbolizes functional classes of proteins. Majority of functional class entities coincide with Gene Ontology molecular function classifications.

Relationship definitions:

MolTransport: Indicates that the regulator changes the localization of the target. Describes events of molecular translocation, export, import or release

Regulation: Indicates that the regulator changes the activity of the target. The mechanism of the regulation is either unknown or has not been specified in the sentence describing the relation.

Chemical Reaction: Symbolizes either enzyme catalyzed or spontaneous chemical reaction, i.e. transformation of one set of Small Molecules into another. Usually must have at least one substrate show with the link incoming to control and one product shown with the link outgoing from control. Enzyme is shown as undirected link between Functional Class or Protein entity and control

Binding: Physical interaction between molecules

ProtModification: Indicates that the regulator molecule changes the protein modification of the target molecule. Usually indicates the direct interaction, i.e. the regulator catalyses the chemical modification reaction.

DirectRegulation: Indicates that the regulator influences the target activity by physically interacting with it.
Expression: Indicates that the regulator changes the protein level of the target, by means of regulating its gene expression or protein stability.

PromoterBinding: Indicates that the regulator binds the promoter of the target.

MolSynthesis: Indicates that the regulator changes the concentration of the target. Usually Small Molecule is a target in MolSynthesis

Appendix B: User Roles for Enterprise Accounts

In a Pathway Studio Enterprise database, user accounts can be set up with one or more user roles. User roles are not cumulative and must be granted separately. For example, the permission to browse the database included into the role “user” is not included into the role “editor”. In most cases, the role “user” should be granted in addition to other roles.

User

- May browse and search the entire database
- Has full control over their private workspace, i.e. may create, edit or delete entities and relations stored in USERNAME Projects folder
- Can work with networks containing up to 1000 entities

Publisher

- May submit private data for publishing into the public workspace (must be approved by a manager prior to publication)

Power user

- Has access to MedScan text-mining tools through the web-client
- May visualize networks up to 5000 entities

Administrator (tools are available only in web client)

- Manages user accounts and roles
- Manages user sessions and server tools
- Manages database object properties and annotations
- New administrator accounts cannot be created; the Administrator role can be associated only with the built-in ‘admin’ account

Editor (tools are available only in webclient)

- May edit, create or delete public objects, e.g. pathways, entities, relations, etc.

Manager (tools are available only in webclient)

- Reviews pathways submitted by publisher for publishing in the shared workspace
- Approves or rejects publishers’ submissions

Graph viewer

- May not browse or search the database
- View-only access to pathways, through URLs sent by regular users

Note: User accounts can be set up with one or more user roles. User roles are not cumulative and must be granted separately. For example, the permission to browse the database included into the role “user” is not included into the role “editor”. In most cases, the role “user” should be granted in addition to other roles

Appendix C: How to Import Network Data

The import wizard “Import Relationships in Tabular Format” allows you to integrate external entity/relation data into the ResNet database. You can import relations, annotation associated with the relations and set up a template to automate the process of for consistent data sources. The initial data can be prepared in either MS Notepad or MS Excel, with the final format: one record per line with multiple fields separated with a delimiter character such as tab, comma, semicolon or a space.

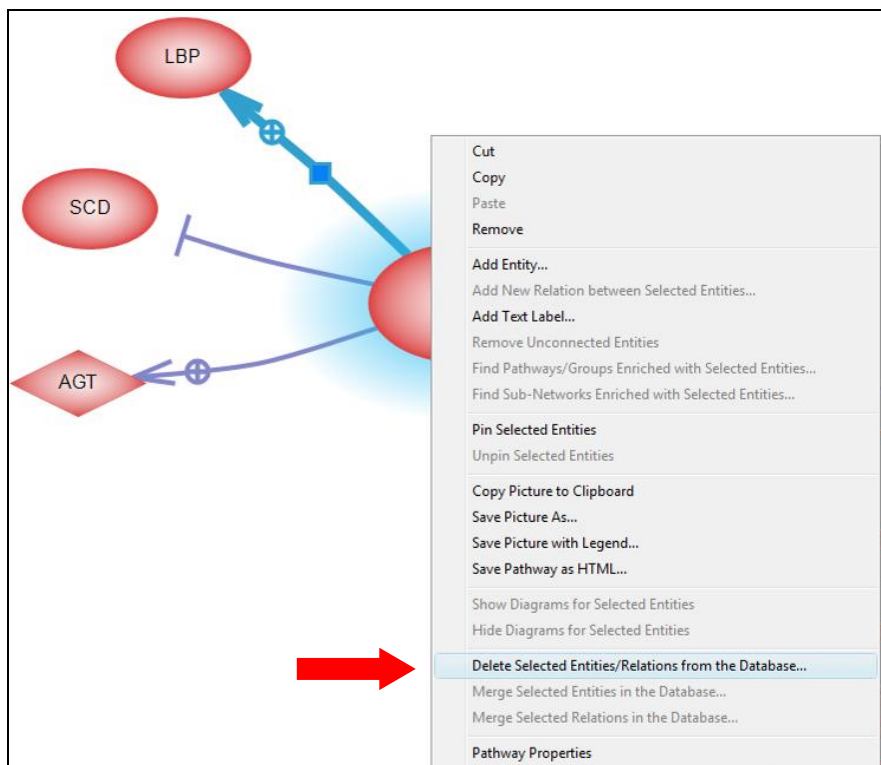
You can also specify new relation types for input using this wizard.

See the User’s Manual: Pathway Studio Tasks and Tools > Importing and Exporting > Import Pathway Wizard > Import Relationships with Tabular Format for detailed instructions on using this import wizard.

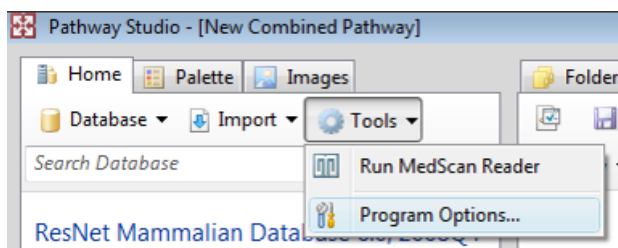
Appendix D: Deleting Entities and Relations from a Local Database (Pathway Studio Desktop)

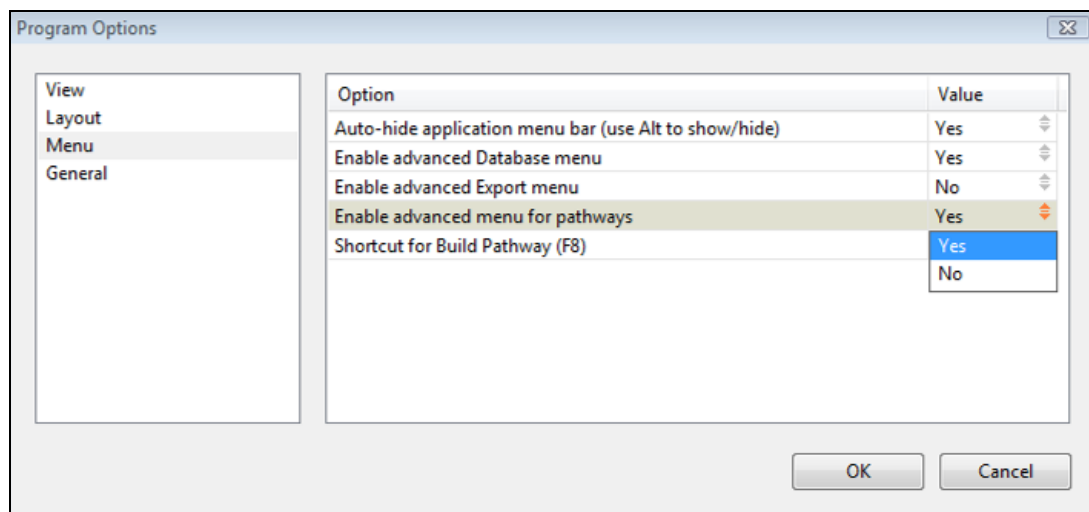
When you delete a relation or entity from a pathway view, it is only removed from that specific pathway but still remains in the ResNet database.

To permanently delete a relation or entity from the ResNet database, select the relation or entity, move the mouse cursor to an empty area of the pathway, right-click and select “Delete Selected Entities/Relations from Database.”



Note: If “Delete Selected Entities/Relations from the Database” does not appear as a menu option, in the Information Pane go to Tools > Program Options. Next, select “Menu” then change “Enable Advanced Menu for Pathways” from “No” to “Yes.”





Note: Deleting a relationship does not prevent that relationship from being added to the database again should the same source be processed again through MedScan.