Elsevier in the Classroom

Workbook for Pathway Studio®

Biological molecular pathway modeling &
experimental data annotation reference tool

ELSEVIER

This workbook contains a series of Pathway Studio teaching modules designed to provide comprehensive training in the full use of the Pathway Studio software, set within a Systems Biology context.

Additional instructional assistance is available in the form of prerecorded webinars which can be accessed through the Pathway Studio website at www.pathwaystudio.com

Please be advised that the exact numbers of entities and relations as cited in the various modules in this workbook may differ slightly from the user's experience. The reason for this stems from the frequent and ongoing updating of the underlying Pathway Studio Knowledge base. This updating provides the user a timely review of the most relevant life science literature.

Chris Cheadle, Ph.D.
Director of Genomics Research
R&D Solutions
c.cheadle@elsevier.com

ELSEVIER
www.pathwaystudio.com

//CODiE//
2015 SIIA CODiE WINNER

Elsevier Inc.
www.elsevier.com/rd-solutions

Acknowledgements

To Brad Fenwick, senior vice president, Global Strategic Alliance, Elsevier Inc., for his generous financial grants and moral support, without which this project could not have been done.

To Jaqui Hodgkinson, vice president, Product Development, Biology Products, Elsevier Inc., for her unfailingly cheerful and strong support and great belief in this project. She made it fun!

To Ancha Baranova, associate professor, School of Systems Biology, George Mason University, who helped us get real and produce a workbook that would be effective for both students and faculty.

I would also like to thank Can Kural, one of Dr. Baranova's Ph.D. graduate students, for his greatly appreciated, hands-on feedback and workflow contributions.

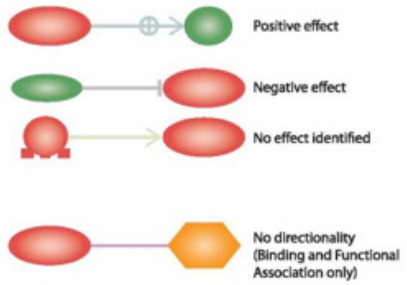
To our staff here at Elsevier—including Hongbao Cao, senior bioinformatician; Pat Morgan, product marketing manager; Kim Gordon, product marketing manager; Anton Yuryev, senior scientist; Maria Shkrob, life science consultant; and Andrey Kalinin, head of MedScan Semantics—for their substantive contributions to this workbook either directly or as I borrowed from their previous good work using the Pathway Studio software. I am indebted to each of them for their professional expertise, sound advice, and great friendship.

ADDITIONAL INFORMATION ABOUT RELATIONS

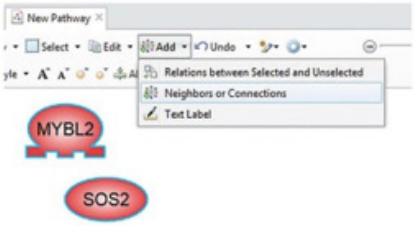
Each relation type is assigned a distinct color

Positive or negative effect for the relation are indicated by arrow head. Relations have arrows to indicate directionality (except Binding).

- Regulation
- ChemicalReaction
- PromoterBinding
- MolSynthesis
- MolTransport
- DirectRegulation
- Expression
- Binding
- ProtModification
- miRNAEffect
- QuantitativeChange
- GeneticChange
- Biomarker
- StateChange
- FunctionalAssociation
- ClinicalTrial



BUILD PATHWAY WIZARD



Build Pathway Algorithms

Find Direct Links: find the relation between two or more selected entities on the network diagram.

Find Shortest Path for a Pair of Entities: find relations between two selected entities on the network diagram, adding intermediate entities as needed to form the connection.

Expand Pathway: find entities directly connected to the entity / entities selected on the network diagram from the database.

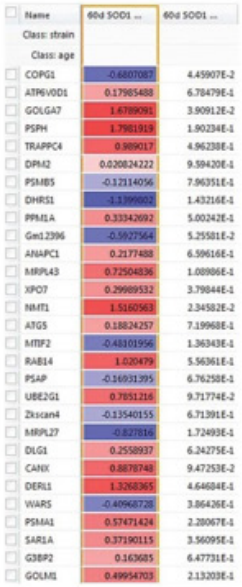
Find Common Targets: find target(s) that are regulated by at least two or more of the selected entities on the network diagram.

Find Common Regulators: find regulator(s) that regulate two or more of the selected entities on the network.

EXPERIMENTAL DATA ANALYSIS

GSEA	Gene Set Enrichment Analysis ranks experiment results by the absolute value of the fold change and identifies known gene sets (pathways and ontologies) that are statistically enriched. **Tool name:** Gene Set Enrichment Analysis.
SNEA	Sub-Network Enrichment Analysis is an extension of GSEA where the "gene sets" used in the enrichment analysis are small regulatory networks calculated de novo from the database by the algorithm. Identifies major regulators (proteins, miRNAs or small molecules), binding networks, metabolomics targets, enriched diseases and cell processes. **Tool name:** Sub-Network Enrichment Analysis.
Fisher's Exact Test	Enrichment analysis that does NOT include experimental values when calculating enrichment from a list. **Tool names:** Find Pathways/Groups Enriched; Find Enriched Genomic positions; Find Sub-Networks Enriched.

EXPERIMENTAL DATA



ALGORITHM

ENRICHED GENE SETS RESULTS

GSEA

Known Gene Sets:

Mammal: Cell Process Pathways, Disease Collections (pathways and groups), Expression Targets Pathways, Immunological Pathways, Metabolic Pathways, Nociception Pathways, Signaling Pathways, Gene Ontology, Pathway Studio® Ontology, Chromosomal localization enrichment.

Plant: AraCyc Pathways, Arabidopsis Signaling Pathways, MaizeCyc Pathways, RiceCyc Pathways, Plant Ontology, Pathway Studio® Ontology, Gene Ontology.

SNEA

De novo user-defined sub-networks: expression regulators, miRNA regulators, binding networks, metabolomics targets, disease and cell process enrichment.

Fisher's

Known Gene Sets:

Mammal: Cell Process Pathways, Disease Collections (pathways and groups), Expression Targets Pathways, Immunological Pathways, Metabolic Pathways, Nociception Pathways, Signaling Pathways, Gene Ontology, Pathway Studio® Ontology, Chromosomal localization enrichment.

Plant: AraCyc Pathways, Arabidopsis Signaling Pathways, MaizeCyc Pathways, RiceCyc Pathways, Plant Ontology, Pathway Studio® Ontology, Gene Ontology.

Fisher's

De novo user-defined sub-networks: expression regulators, miRNA regulators, binding networks, metabolomics targets, disease and cell process enrichment.

PATHWAY STUDIO®

http://help.elsevier.com/app/pathway_select/p/9048

ELSEVIER

QUICK REFERENCE

ENTITY TYPES

(mammal, mammal+CE+Dfx, and plant database)

	Cell Process * – biological processes, most coincide with Gene Ontology.
	Clinical Parameter – measured parameters of the human body used in clinical practice (mammal and mammal+CE+DFx only).
	Complex * – several polypeptides that form a complex via physical interactions.
	Disease – Mammal: health conditions and disease terms from MeSH; plant: Plant diseases.
	Functional Class * – most functional classes coincide with Gene Ontology.
	Protein – defined by Entrez Gene - represents both genes and the gene products, including proteins and miRNAs.
	Small Molecule – Mammal: naturally occurring metabolites and small molecules found in cells; ChemEffect® adds drugs (including some biologically active polypeptides that work as drugs such as monoclonal antibodies) and non-naturally occurring small molecules to the mammal database. Plant: naturally occurring metabolites and small molecules and other plant related chemicals (ex. herbicides or research related chemicals).
	Treatment – non-chemical treatments and environmental conditions, such as cold shock.

**Container Entities – these are valid entities but also can have proteins mapped to them. You can see the proteins for the container entities in the "child concepts" in the property records for the specific entity.*

PROTEIN SUB-TYPES

(mammal, mammal+CE+Dfx, and plant database)

Complexes are also "protein" entities but represent a group of proteins functioning together. In the Pathway Studio® database they function as a complex entity type so are considered separately.

- Protein (no class assigned)
- Protein kinase
- Transcription factor
- Phosphatase
- Receptor
- miRNA
- Ligands

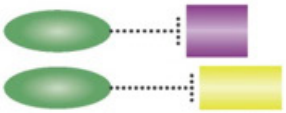
RELATION TYPES

(both mammal and plant database)

Binding	direct physical interaction between two molecules.
ChemicalReaction	enzyme catalyzed reaction involving small molecules.
DirectRegulation	influences target activity by direct physical interaction (excluding promoter binding interactions).
Expression	regulator changes protein abundance by affecting levels of transcript or protein stability.
miRNAEffect	the inhibitory effect of a miRNA on its mRNA target.
Regulation	changes the activity of the target by an unknown mechanism (may be direct or indirect). This is a less specific relation type than others provided.
MolSynthesis	regulator changes the concentrations of the target (usually a small molecule target).
MolTransport	regulator changes the localization of the target (molecular translocation, export, import etc.).
PromoterBinding	regulator binds to the promoter of a gene.
ProtModification	regulator changes the modification of the target molecule, usually by a direct interaction. **Filtering Field Name:** Mechanism **Sub-Categories:** acetylation, cleavage, deacetylation, demethylation, dephosphorylation, direct interaction, methylation, phosphorylation, posttrascriptional inhibition, proteolysis, ubiquitination.

ADDITIONAL DATA IN THE CHEMEFFECT® DATABASE

(added to Mammal)



- Relations between small molecules and diseases/cell processes.
- Relations between non-naturally occurring metabolites (small molecules), such as drugs, which are not included in the Mammal database.

ADDITIONAL DATA IN THE DISEASEFX™ DATABASE

(added to Mammal)

Additional relation types in DiseaseFx™:

Quantitative Change	Changes in abundance/ activity/expression of a gene/ protein/small molecule in a disease state (between disease-protein/complex/ functional class/small molecules). **Filtering Field name:** Quantitative Change **Sub-Categories:** Expression, Abundance, Activity
Genetic Change	Genetic changes in a gene in a disease state such as gene deletions, amplifications, mutations or epigenetic changes (between disease-protein/complex/functional class). **Filtering Field Name:** Change Type **Sub-Categories:** Gene Deletion, Mutation, Gene Amplification, Epigenic methylation
Biomarkers	Identification of proteins/ complexes/functional classes/ metabolites that are prognostic or diagnostic biomarkers for a disease (between disease-protein/complex/functional class/naturally occurring small molecules). **Filtering Field Name:** Biomarker Type **Sub-Categories:** Diagnostic, Prognostic
State Change	Changes in a protein's post-translational modification status or alternative splicing events associated with a disease (between disease-protein/complex/functional class). **Filtering Field Name:** Change Type **Sub-Categories:** Alternate Splicing, Phosphorylation
Functional Association	Different types of functional associations between a disease and a cellular process or another disease (between Disease – Cell Process) (no sub-types).
Clinical Trials	Disease/cell process relation representing clinical trials conducted for a drug against a disease (from ClinicalTrials.gov) (between Disease/Cell Process – Small Molecule) (no sub-types).

PATHWAY STUDIO®

http://help.elsevier.com/app/pathway_select/p/9048

Table of Contents

Module 1

Introduction to Pathway Studio

Contents

I LIKE IT.

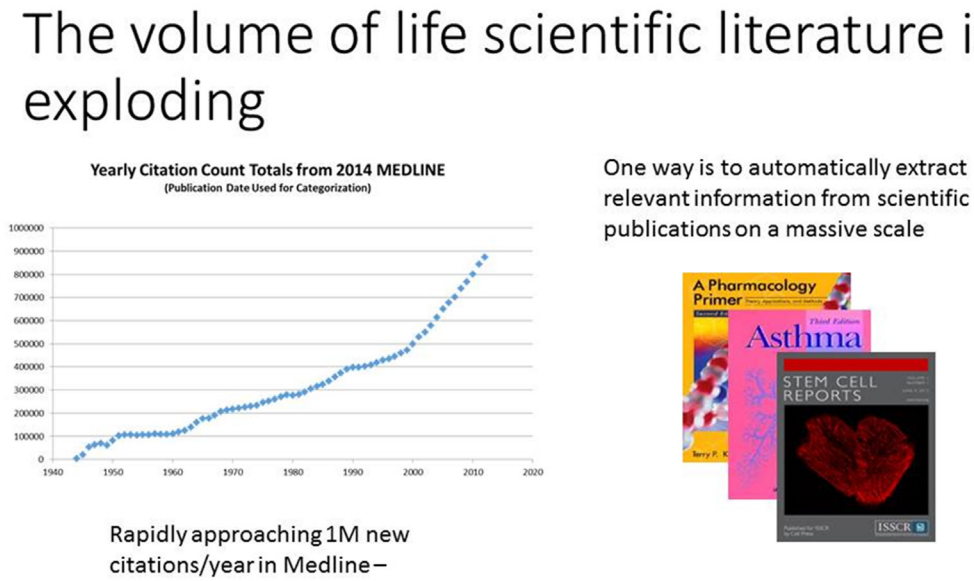
WHAT IS IT?

So people ask, "What is Pathway Studio?"

Actually that is not a trivial question. Pathway Studio, as the name implies, is first and foremost, a tool for biological pathway analysis. It can also be described as a molecular modeling software. Finally, and perhaps, most importantly, Pathway Studio is an exceptionally powerful tool for the direct data mining of the most relevant life science literature.

1.1 Volume of Scientific Literature is Exploding

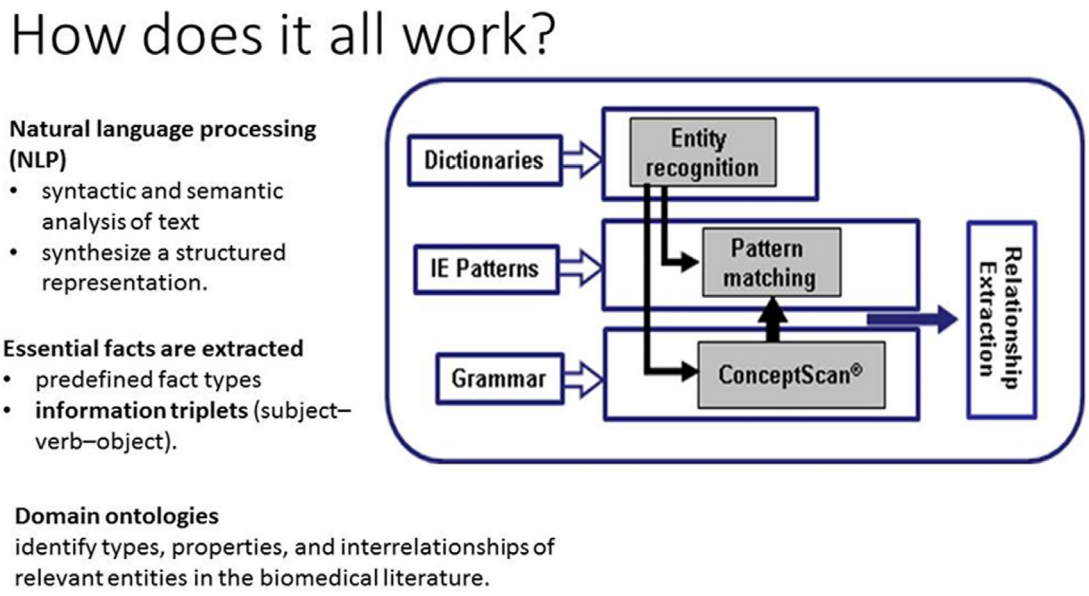
In recent years, the volume of scientific literature has increased exponentially. The rise in the total number of yearly citations from MEDLINE through 2014 is approaching more than one million new citations each year.



So the problem for scientific researchers is how to keep up. Even within their specific domains of interest, scientific researchers can be overwhelmed by the sheer volume of ongoing publications.

Elsevier deals with this problem by automatically extracting relevant information from scientific publications on a very large scale. It does this using its proprietary NLP MedScan technology.

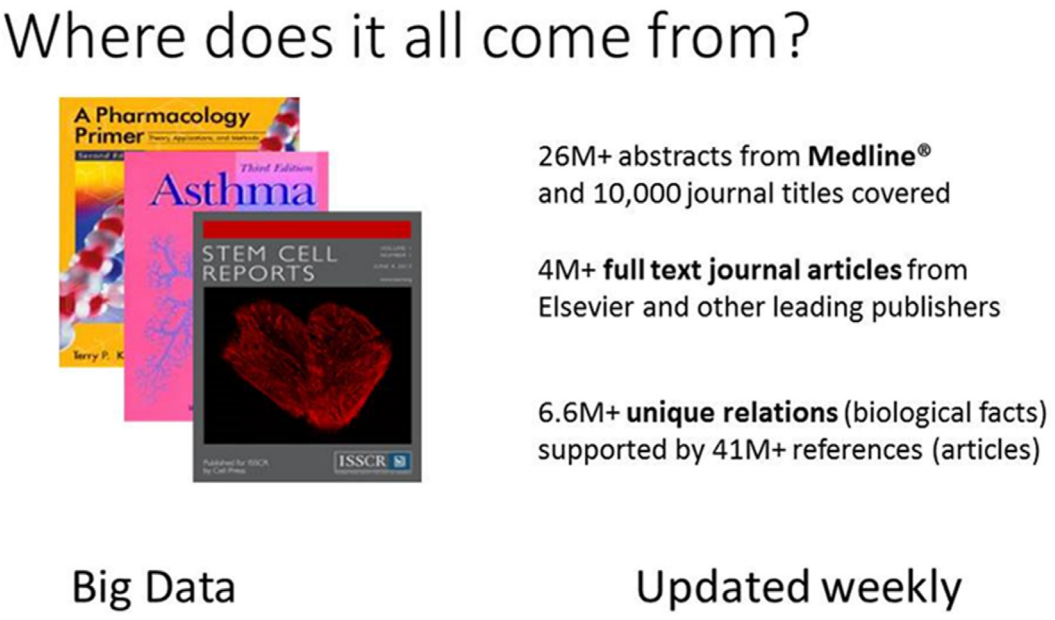
1.2 How Does It All Work?



NLP is a collection of methods for the semantic analysis of unstructured text that allows for the extraction of specific items or facts of scientific interest. Basic NLP works by recognizing and capturing information triplets in the form of subject-verb-object statements. These triplets are the basic unit of NLP technology. The recognition of these triplets is referred to as syntactic analysis (that is, recognition of these relationships in a sentence). All the different ways in which a single gene or protein could be named, for example, is referred to as the dictionary. This is one of the very powerful assets in the MedScan technology as it recognizes many synonyms that refer to the same object or entity. These information triplets once recognized must be understood in the context of particular scientific domains and definitions, and this process is referred to as semantic analysis. The definition of these domain ontologies is part of the NLP art.

1.3 Where Does It All Come From?

So, first things first, where does it all come from? Well, it comes from more than 26 million abstracts from MEDLINE and more than four million full-text journal articles from Elsevier and other major scientific literature publishers. This has resulted, to date, in more than seven million unique relations or facts supported by more than 42 million individual references or articles. The volume and timeliness (the database is updated weekly) certainly qualifies this information as "Big Data."



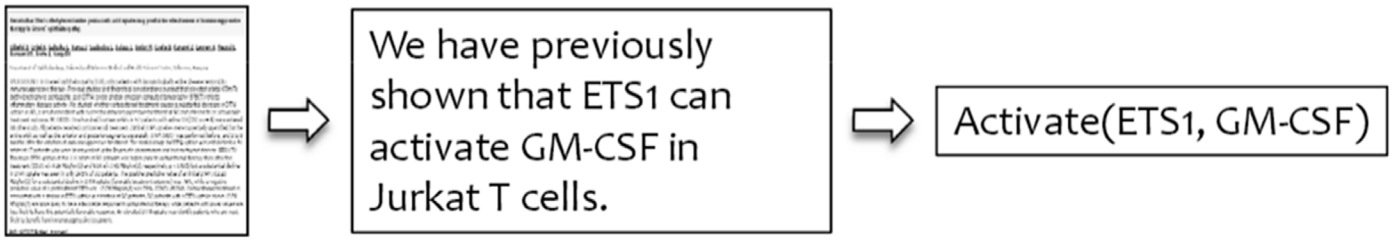
1.4 Information Extraction versus Information Retrieval

So when we talk about information extraction (IE), how do we distinguish that from information retrieval? It's easy! Information retrieval is what you typically get from an Internet search based on key words, for example, where the best documents are displayed in your search results. This is fine and many search engines are excellent, but you still have to sift through all the documents for the information that you're looking for and it can quickly become an overwhelming task. Information extraction, on the other hand, uses NLP technology to basically go from unstructured text to a structured knowledge representation (often while keeping the links back to where the information was extracted from, as in the case with Pathway Studio), which can then be curated into large databases of information.

Information Extraction (IE) vs Information Retrieval (IR)

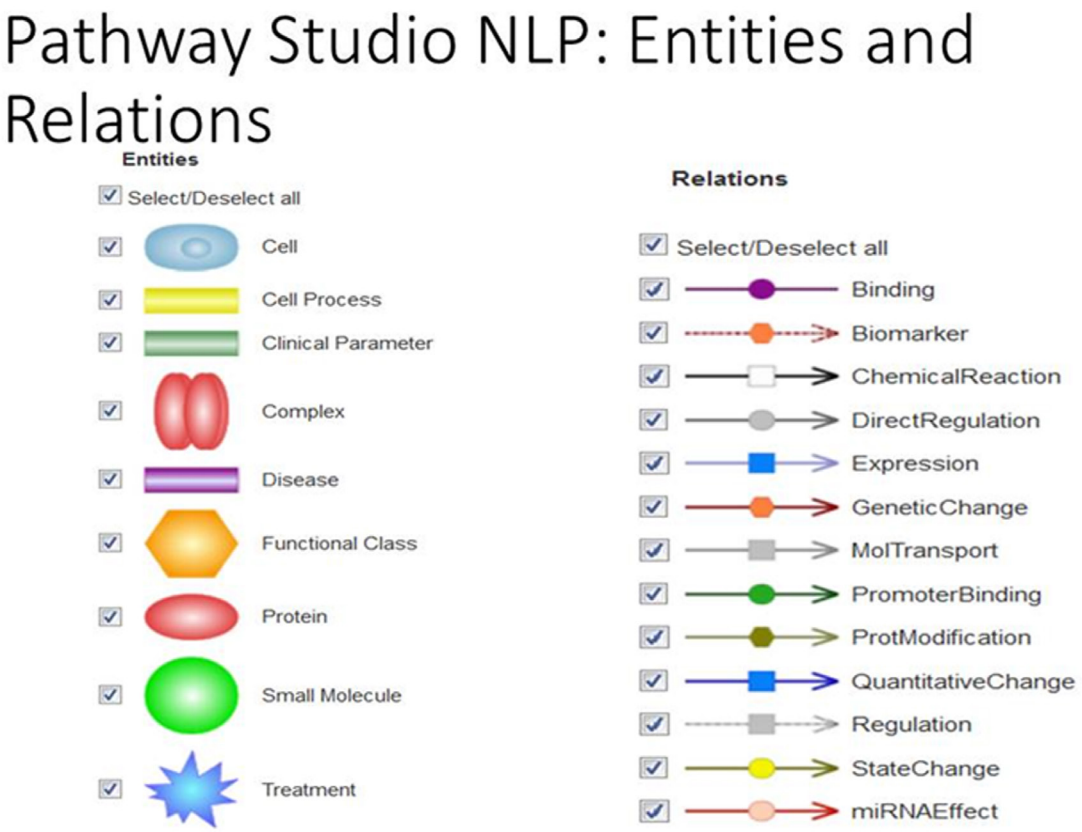
Information Retrieval (IR)	Information extraction (IE)
Returns documents.	Returns facts.
Is a classification task (each document is relevant/not relevant to a query).	Is an application of natural language processing, involving the analysis of text and synthesis of a structured representation.
Can be done without reference to syntax (treating query and indeed the documents as merely a "bag of words").	Is based on syntactic analysis and semantic analysis

As an example of how all this works, suppose you are reading through an abstract on MEDLINE, and you come across this sentence: "We have previously shown that ETS1 can activate GM-CSF in Jurkat T cells." The MedScan NLP technology will recognize that there is an activation statement in this sentence and that it combines two objects, in this case, genes or proteins. The NLP triplet then is an activate relation, which combines ETS1 and GM-CSF and preserves the direction of that interaction (ETS1->GM-CSF). It will also preserve the contextual information for this relation, which, in this case, is that this interaction was observed in Jurkat T cells.



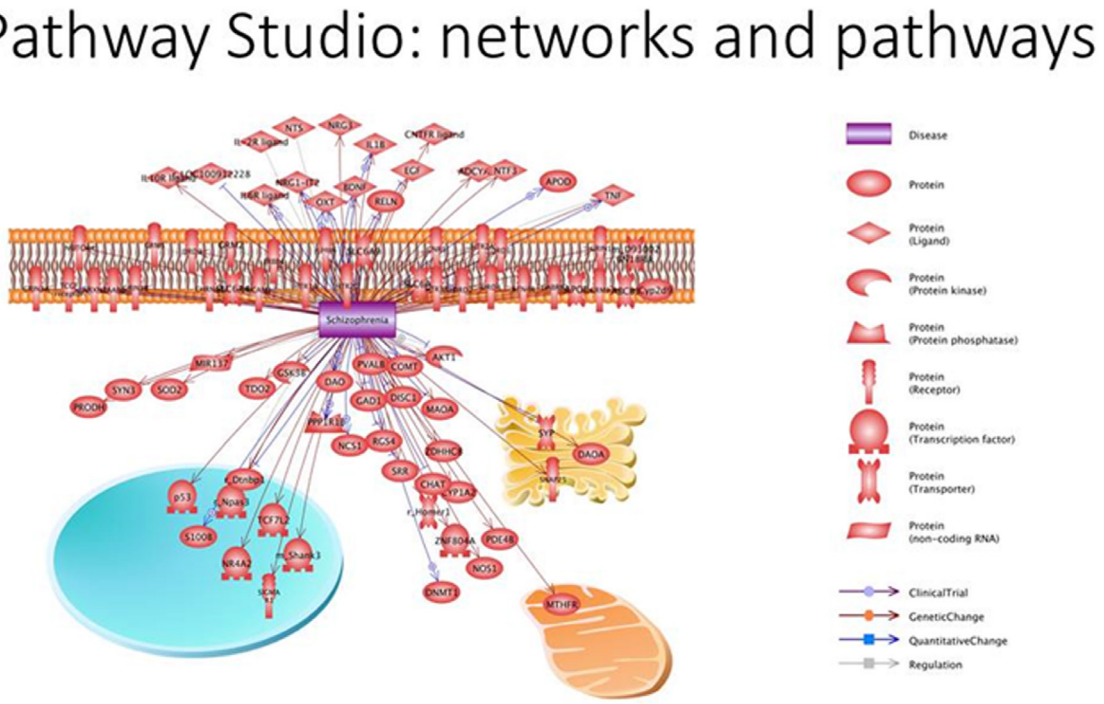
1.5 Pathway Studio NLP: Entities and Relations

The Pathway Studio database is constructed as a series of classic network interactions. Each interaction is called a relation and is composed of nodes, which are called entities, and edges, which are called relations or relation types. These relations are essentially the curated NLP triplets and can be extended in a network on the basis of common nodes. Shown here are examples of entity types such as Proteins, Small Molecules, and Cells, and common relation types such as Biomarker, Expression, and QuantitativeChange. This combination of entity and relation types allows for enormous flexibility in capturing life science content.



1.6 Pathway Studio: Networks and Pathways

An example of a constructed Pathway Studio pathway or network is presented here in this Graph View illustration of all the genes and proteins related to the disease schizophrenia, which have a minimum number of supporting articles equal to or greater than 10. The accompanying legend illustrates the different icons used to represent the various entity types such as disease and proteins. In this case, protein subtypes are also indicated by special icons such as the diamond-shaped forms that indicate protein ligands seen at the exterior of the cell. This particular visualization is displayed on the backdrop of a standard cell compartmentalization view.



1.7 Proteins/Genes Associated with Schizophrenia: Relation Table View

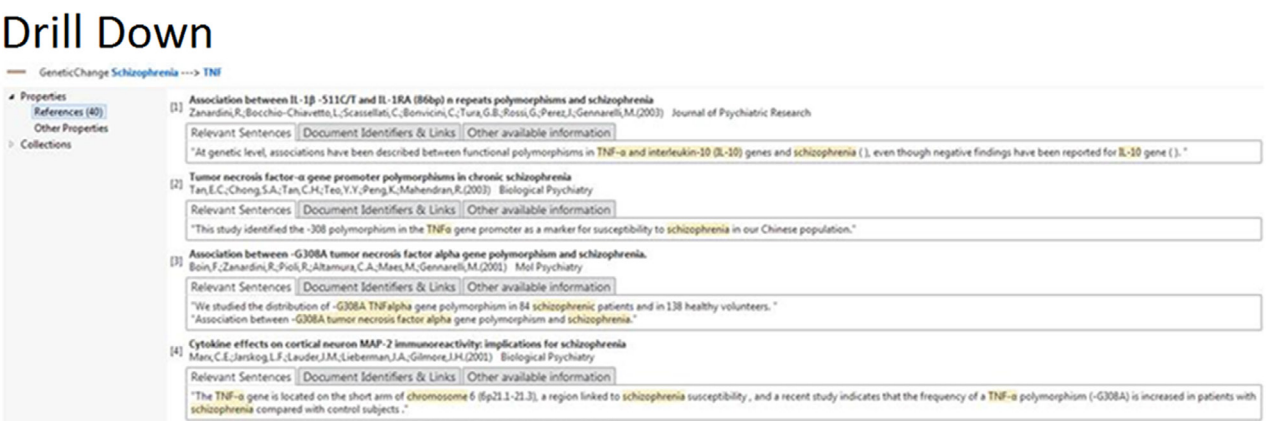
The Relation Table View shown here allows the user to inspect each relation in a given network in tabular form and view all the available annotation information for each relation, such as the number of supporting articles. In this particular view, the relations have been sorted by the number of references in descending order. From this you can easily see that, at least currently, the most studied gene with regards to schizophrenia is COMT, or Catechol-O-methyltransferase, with 325 references. The relation type is GeneticChange, and the change type is mutation, suggesting that mutations in COMT gene are considered very important in the study of schizophrenia.

Relation	Object Type	Effect	QuantitativeType	ChangeType	# of References	CellType	Title	Pub...
Schizophrenia ---> COMT	GeneticChange			mutation, mutation, mutation, m...	325	Erythrocytes, Fibroblasts, Hepatoc...	Schizophrenia from a neural circuitry p...	200...
Schizophrenia ---> DISC1	GeneticChange			mutation, mutation, mutation, m...	191	Embryonic Stem Cells, Epithelial C...	Molecular mechanisms of stress-induc...	200...
Schizophrenia ---> NRG1-IT2	GeneticChange			mutation, mutation, mutation, m...	159	Endothelial Cells, Interneurons, Le...	Molecular cloning of a brain-specific, ...	200...
Schizophrenia ---> BDNF	GeneticChange			mutation, mutation, mutation, m...	155	APUD Cells, Endothelial Cells, Leu...	Brain-Derived Neurotrophic Factor Val...	200...
Schizophrenia ---> DRD2	GeneticChange			mutation, epigenetic methylatio...	130	Blood Platelets, Insulin-Secreting ...	Differential repression by freud-1/CC2...	200...
Schizophrenia ---> HTR2A	GeneticChange			epigenetic methylation, mutatio...	120	Blood Platelets, Neurons, dopami...	Methylation and QTDT analysis of the ...	200...
Schizophrenia ---> RELN	GeneticChange			epigenetic methylation, epigene...	112	CD4-Positive T-Lymphocytes, Cajal-...	DNA (cytosine-5) methyltransferase inh...	200...
Schizophrenia ---> DRD3	GeneticChange			mutation, mutation, mutation, m...	104	CHO cell, Neurons, dopaminergic ...	S33138 [N-[4-[2-[(3aS,9bR)-8-cyano-1,3a...	200...
Schizophrenia ---> BDNF	QuantitativeChange		abundance, abundance, a...		104	Interneurons, Neurons, dopamine...	Brain-derived neurotrophic factor and ...	200...
DRD2 ---> Schizophrenia	Regulation				103	Lymphocytes, Neurons, dopaminer...	Change of dopamine receptor mRNA e...	200...
Schizophrenia ---> r_Dtnbp1	GeneticChange			mutation, mutation, mutation, m...	100	Blood Platelets, Interneurons, Neu...	Finding schizophrenia genes., Dysbind...	200...
HTR2A ---> Schizophrenia	Regulation				72	Neurons, Pyramidal Cells, dopami...	Preclinical pharmacology of FMPD [6-fl...	200...
NRG1-IT2 ---> Schizophrenia	Regulation				70	B-Lymphocytes, Interneurons, Leu...	The genetic and neurobiologic compas...	200...
Schizophrenia ---> MTHFR	GeneticChange			mutation, mutation, mutation, m...	64	Erythrocytes, neural stem cells	Influence of maternal MTHFR A1298C p...	201...
DISC1 ---> Schizophrenia	Regulation				61	Interneurons, Neurons, Oligoden...	Finding schizophrenia genes., Cyclic n...	200...
BDNF ---> Schizophrenia	Regulation				60	Neurons, PBMCs, brain cell, dopa...	The Brain-Derived Neurotrophic Factor...	200...
Schizophrenia ---> RGS4	GeneticChange			mutation, mutation, mutation, m...	57	Germ Cells, Neurons, Oligodendro...	Regulators of G-protein signaling 4 in ...	200...

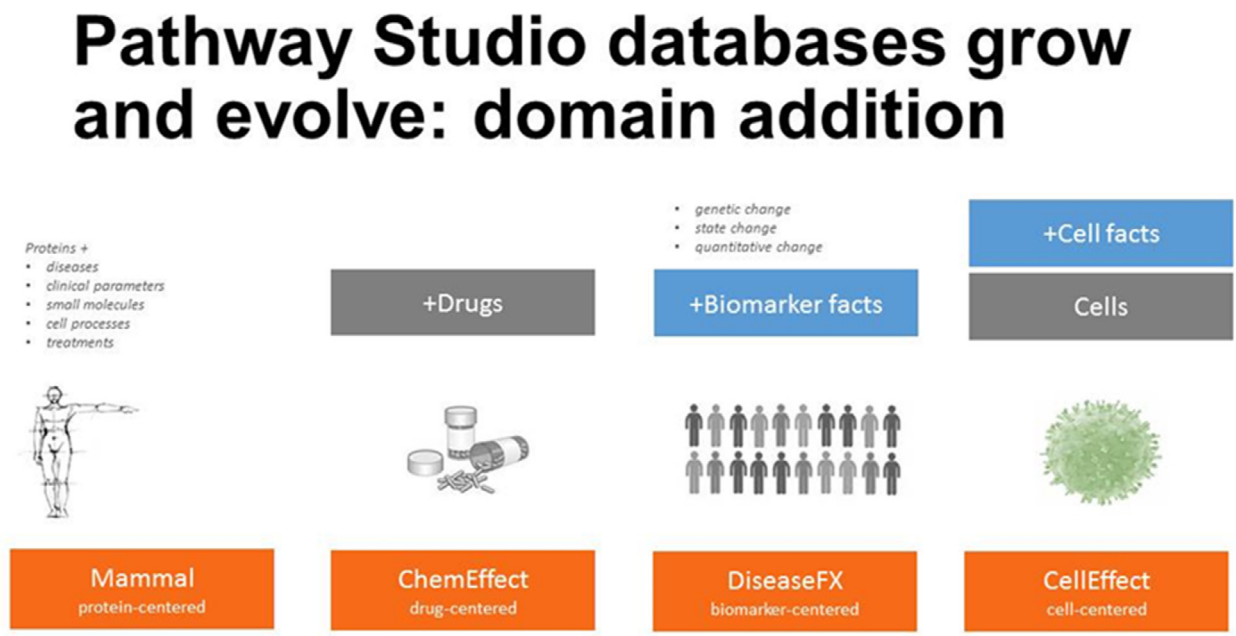
Any relation can be fully inspected from this table by double-clicking on it, which will open up the lower pane and show in detail all the references and the sentences from which this information was extracted. This is one of the great strengths of Pathway Studio, as all the evidence for a given relation is available for inspection immediately. It is even possible, by opening up the Document Identifiers & Links tab, for a reference to get links to either the PubMed abstracts or even directly to the full-text documents themselves (this would depend upon the subscription status of your institution).

As discussed before, domain knowledge is vital for the MedScan technology to correctly identify and interpret information extraction. Pathway Studio was originally designed as primarily a gene- or protein-centric database for the interpretation of gene expression for mammals (human, rat, and mouse). There is also a plant version for Pathway Studio. Over time, additional modules have been added to facilitate drug discovery research, including a drug database of more than 100,000 small molecules and accompanying new relation types (ChemEffect). Also added are more than 16,000 diseases and specific new relations (GeneticChange, QuantitativeChange, StateChange, and Biomarker) in order to facilitate new biomarker discovery (DiseaseFX). Most recently, a comprehensive new taxonomy has been added for the robust identification of more than 700 human anatomical cell types, including a major emphasis on precisely defined cells of the immune system (CellEffect).

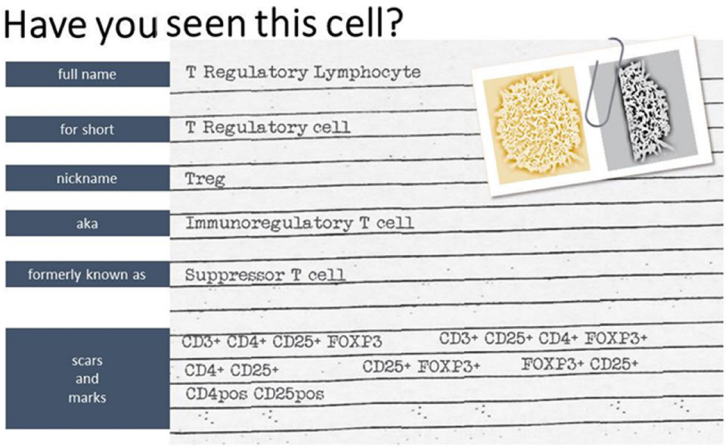
1.8 Drill Down



1.9 Pathway Studio Databases Grow and Evolve: Domain Addition

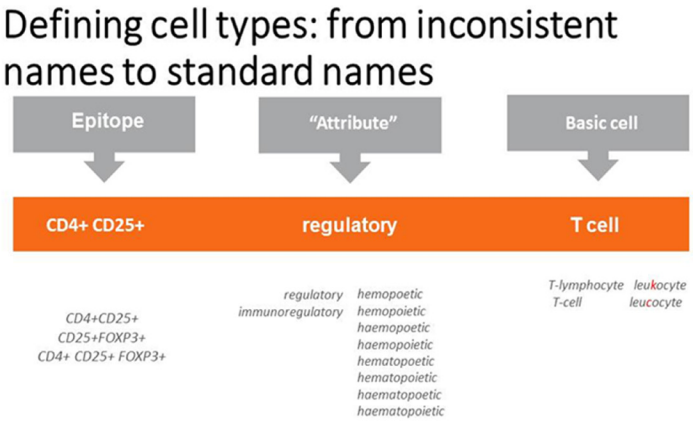


1.10 Have You Seen This Cell?



Immune cells in particular have been difficult to catalogue because of the wide variation in naming conventions used by scientists in their publications, as illustrated here for T-regulatory Lymphocytes.

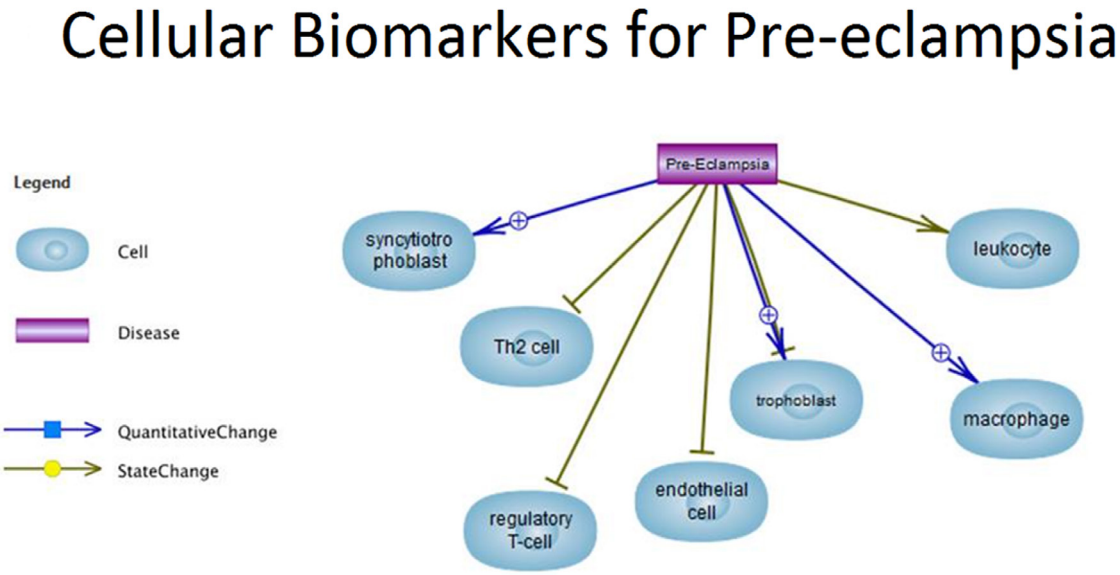
1.11 Defining Cell Types: From Inconsistent Names to Standard Names



The most reliable identification used by immunologists is to define the different immune cell types by their cell surface epitopes (identified by antibody mediated cell sorting [FACS] methods). All these possible terms have been carefully normalized by Elsevier (?) expert curators to allow the widest possible and most accurate identification of these cell types in the literature.

1.12 Cellular Biomarkers for Pre-eclampsia

An example of how they are used is shown in this figure linking the disease pre-eclampsia with key cell types including



synctiotrophoblasts. These cell types form the epithelial covering of embryonic placental villi and act as the site for nutrient exchange between the embryo and the mother. It is believed that the synctiotrophoblasts are also the cell type that mediates the transmission of Zika virus from maternally infected macrophages across the placental barrier to infect the developing fetus. This can result in disastrous developmental defects including the characteristic microcephaly or small head phenotype of these infected babies.

What did you learn today?

- Pathway Studio is a pathway analysis tool, a type of molecular modeling software, and a great method for systematic data-mining of the scientific literature.
- The sources of the Pathway Studio knowledge base are [consider recapping the sources here].
- Natural Language Processing (NLP) is [consider recapping definition here].
- The difference between data retrieval and data extraction is [consider recapping definition here].
- Pathway Studio networks contain molecular entities and biological relations.
- Pathway Studio is constantly advancing through data expansion and domain additions (e.g. ChemEffect, CellEffect, and DiseaseFX).

Study Questions 1

1. What is the basic unit of NLP technology?

2. How many unique relations or facts are contained in the Pathway Studio database?

3. What is the difference between Pathway Studio data and a Google search?

4. What is the most abundant entity type in Pathway studio? What is the most abundant relation type in Pathway Studio? (Hint: use Pathway Studio Database Content under the Start tab).

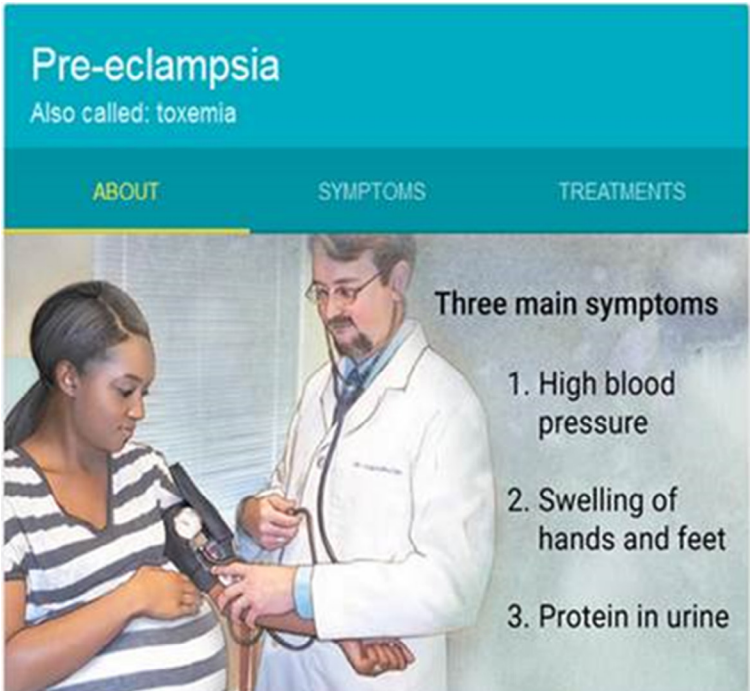
Module 2

Pre-eclampsia, a Disease-centric Analysis

Contents

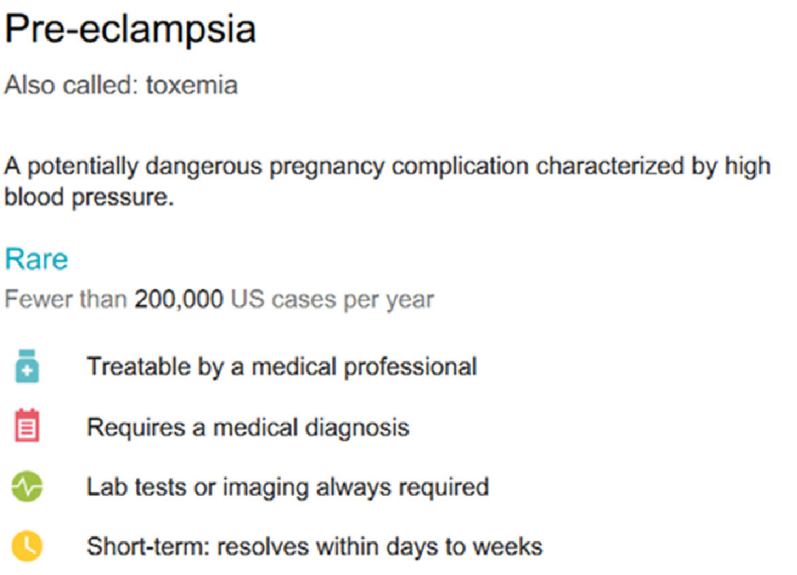
2.1 Introduction

Pre-eclampsia or preeclampsia (PE) is a disorder of pregnancy characterized by high blood pressure and a large amount of protein in the urine. The disorder usually occurs in the third trimester of pregnancy and gets worse over time. In severe disease there may be red blood cell breakdown, a low blood platelet count, impaired liver function, kidney dysfunction, swelling, shortness of breath due to fluid in the lungs, or visual disturbances. Preeclampsia increases the risk of poor outcomes for both the mother and the baby. If left untreated, it may result in seizures at which point it is known as eclampsia.



Pre-eclampsia usually begins after 20 weeks of pregnancy in a woman, whose blood pressure had been normal. It can lead to serious, even fatal complications for both mother and baby.

There may be no symptoms. High blood pressure and protein in the urine are key features. There may also be swelling in the legs and water retention, but this can be hard to distinguish from normal pregnancy.

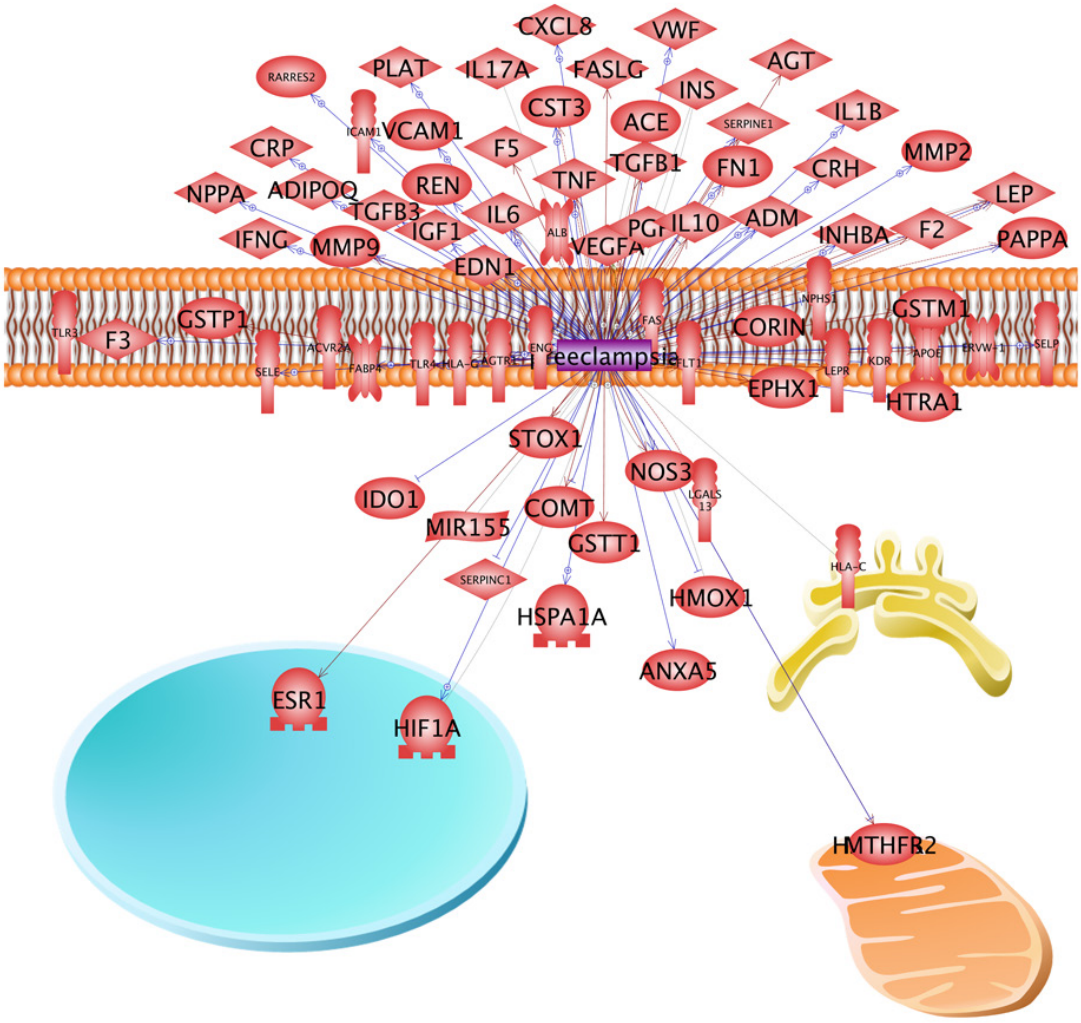


Pre-eclampsia can often be managed with oral or IV medications until the baby is sufficiently mature to be delivered. This often requires weighing the risks of early delivery versus the risks of continued pre-eclampsia symptoms.

2.2 Proteins Related to Pre-eclampsia

What are the most important proteins related to pre-eclampsia as found in the scientific literature?

Find all pre-eclampsia disease relations with protein as an entity type, select all reference >=10

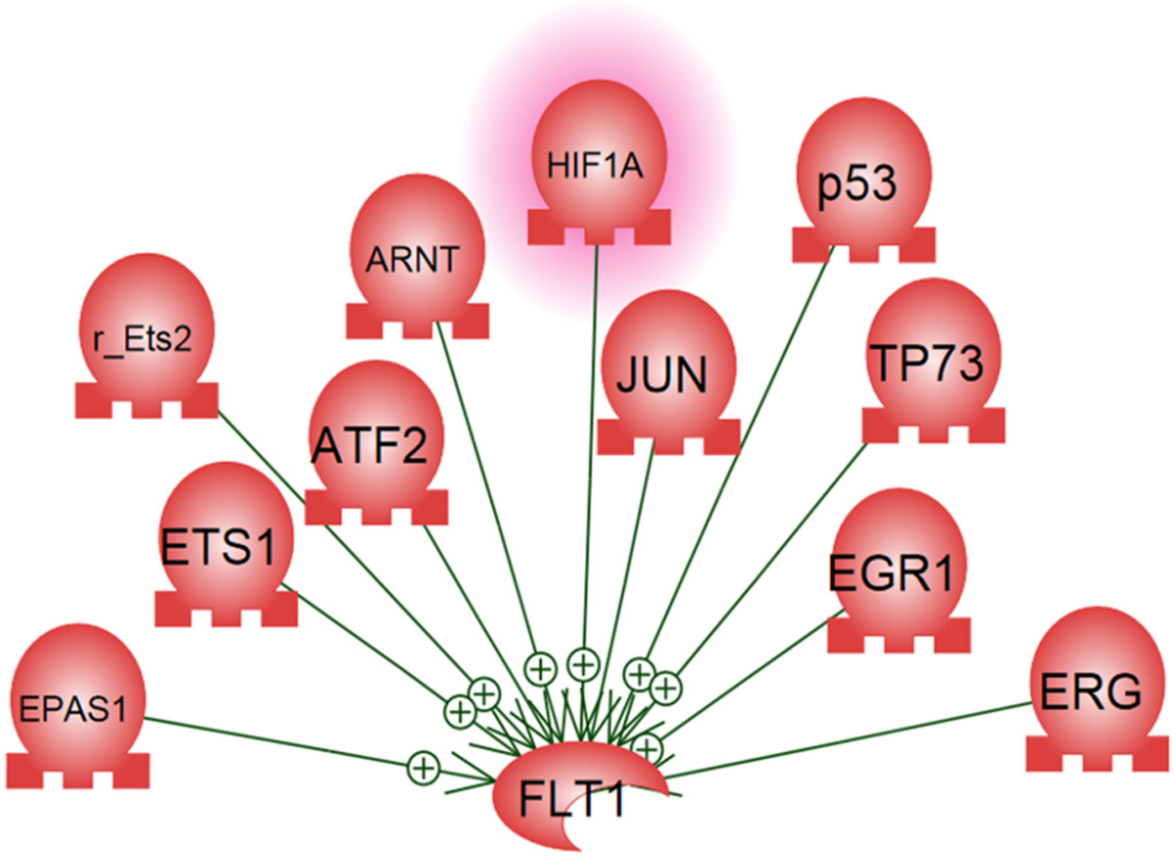


What is the single most highly cited protein related to pre-eclampsia and what is its most common transcription factor?

Select protein most highly cited in connection with pre-eclampsia (FLT1).

Find proteins (transcription factors) that bind to the promoter of FLT1.

Highlight most cited transcription factor for FLT1 (HIF1A).

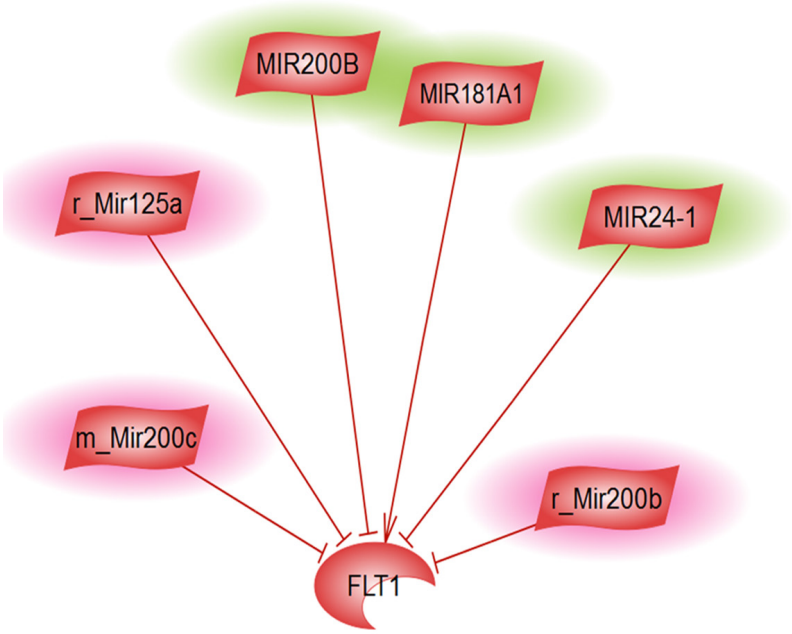


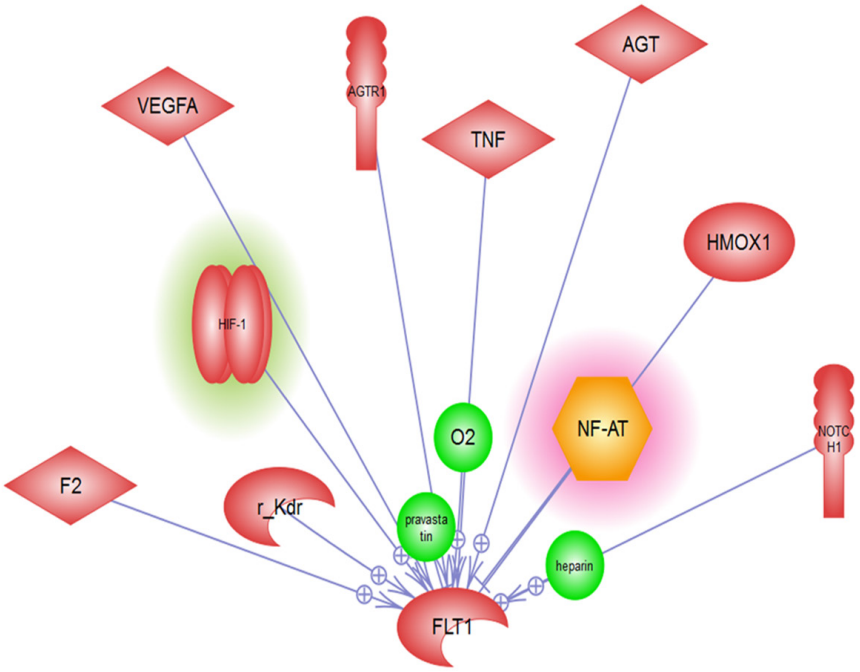
2.3 MiRNAs Affecting FLT1 Gene Expression

Are there any miRNAs that affect the expression of the FLT1 gene?

Find predicted miRNAs that may regulate expression of FLT1.

Highlight predicted (red) vs experimental (green) miRNAs for FLT1.



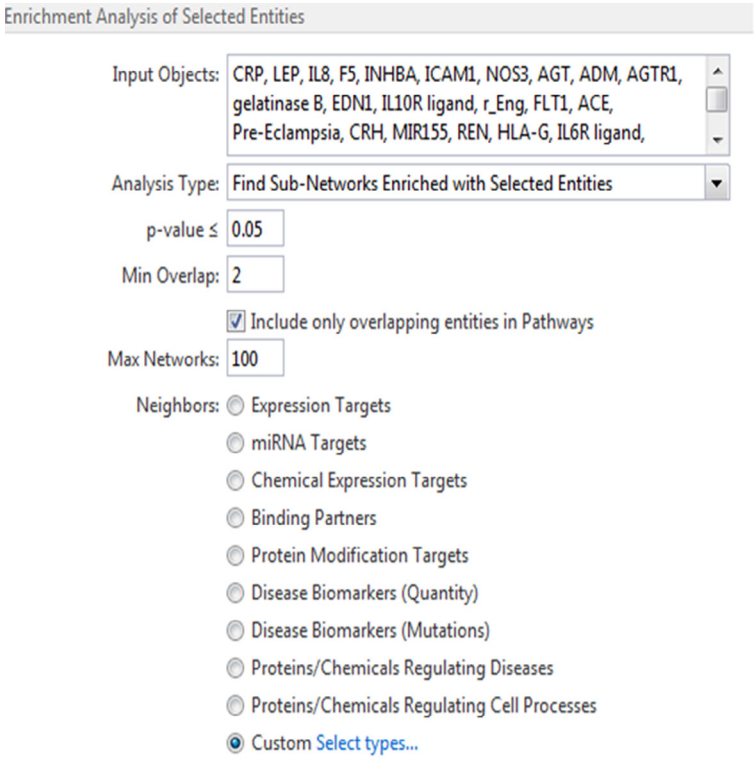


Find proteins, protein complexes, protein functional classes, and small molecules that are involved in the expression FLT1, either directly or indirectly (references >=10).

Highlight protein functional classes (red) and protein complexes (green).

2.4 Sub-Network Enrichment Analysis—Common Expression Regulators

Identify groups of genes (from the list of proteins most strongly associated with pre-eclampsia) that share common expression regulators using Sub-Network Enrichment Analysis (SNEA).



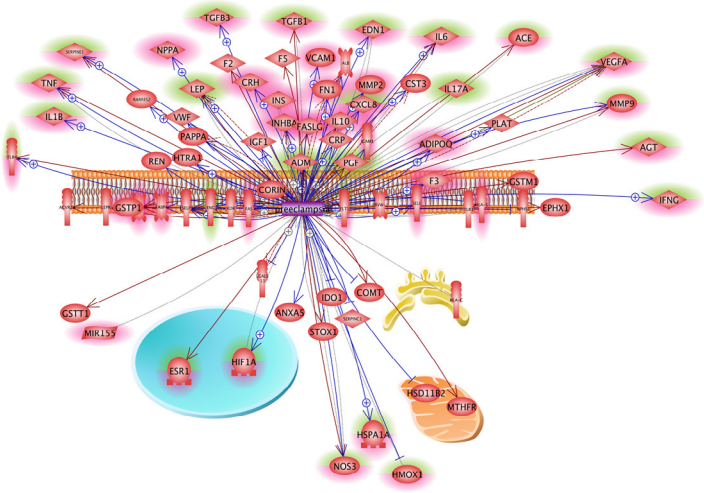
Under Custom Select types, use "Upstream" as the direction, "Protein" as the entity (seed), and "Promoter Binding" as the relation.

Name	Total # of Neighbors	Gene Set Seed	Overlap	Percent Overlap	Overlappi	p-value	Jaccard sir	Hit type
Downstream Neighbors of JUN	331	JUN	24	7	EDN1;INH	1.79E-19	0.069565	Downstre
Downstream Neighbors of HIF1A	248	HIF1A	17	6	EDN1;ADM	7.76E-13	0.063197	Downstre
Downstream Neighbors of ATF2	89	ATF2	12	13	ACE;HIF1A	1.5E-12	0.104348	Downstre
Downstream Neighbors of EP300	220	EP300	16	7	EDN1;VCA	1.89E-12	0.066116	Downstre
Downstream Neighbors of CEBPB	300	CEBPB	16	5	CRP;EDN1	2.16E-10	0.049689	Downstre
Downstream Neighbors of ETS1	170	ETS1	13	7	VCAM1;AC	2.16E-10	0.066667	Downstre
Downstream Neighbors of FOS	171	FOS	13	7	VCAM1;M	2.33E-10	0.066327	Downstre
Downstream Neighbors of STAT3	309	STAT3	16	5	CRP;ADM;	3.36E-10	0.048338	Downstre
Downstream Neighbors of EGR1	232	EGR1	14	6	ACE;HIF1A	8.51E-10	0.054688	Downstre

How many of the pre-eclampsia genes (from the first network in this module) share either of the top two FLT1 transcription factor regulators?

Highlight the pre-eclampsia genes for the two most common upstream expression regulators (JUN in red, HIF1A in green) as revealed in the SNEA analysis (See below for help).

Now for....Pathway Studio Trick #1!



PATHWAY STUDIO TRICK #1

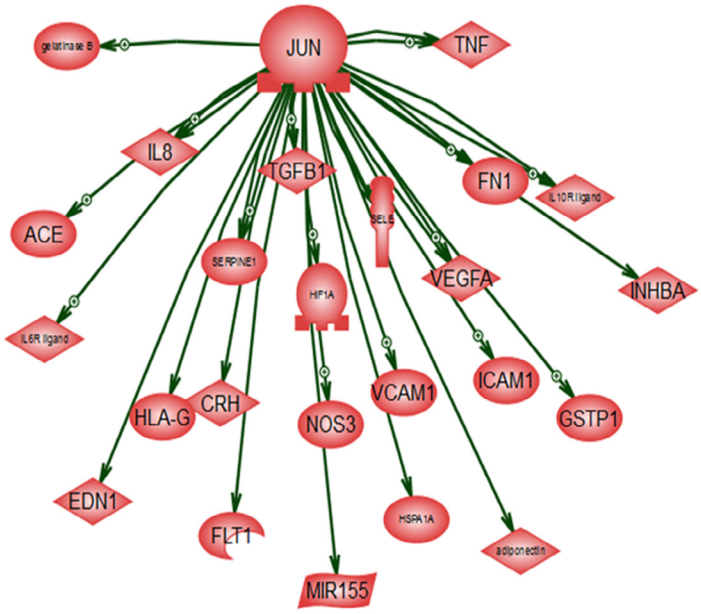
This may not be obvious (at first!), but the way to highlight proteins/genes in a pathway view with information from another group is actually quite easy. Just follow the trail of bread crumbs below.

In the current example, a list of the most common upstream regulators for the pre-eclampsia-related genes were generated using the SNEA tool. The top two entries, which are sorted by p-value, in that list (as found in the table below the pathway viewer) are JUN and HIF1A.

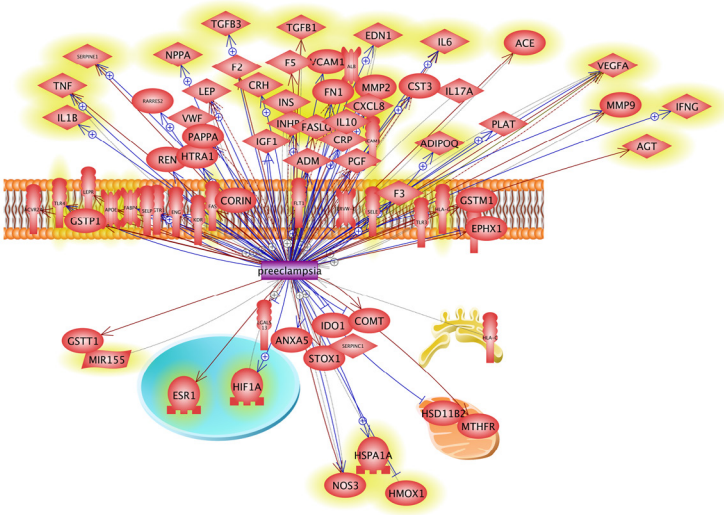
Name	Total # of Neighbors	Gene Set Seed	Overlap	Percent Overlap	Overlappi	p-value	Jaccard sir	Hit type
Downstream Neighbors of JUN	331	JUN	24	7	EDN1;INH	1.79E-19	0.069565	Downstre
Downstream Neighbors of HIF1A	248	HIF1A	17	6	EDN1;ADM	7.76E-13	0.063197	Downstre
Downstream Neighbors of ATF2	89	ATF2	12	13	ACE;HIF1A	1.5E-12	0.104348	Downstre
Downstream Neighbors of EP300	220	EP300	16	7	EDN1;VCA	1.89E-12	0.066116	Downstre
Downstream Neighbors of CEBPB	300	CEBPB	16	5	CRP;EDN1	2.16E-10	0.049689	Downstre
Downstream Neighbors of ETS1	170	ETS1	13	7	VCAM1;A	2.16E-10	0.066667	Downstre
Downstream Neighbors of FOS	171	FOS	13	7	VCAM1;M	2.33E-10	0.066327	Downstre
Downstream Neighbors of STAT3	309	STAT3	16	5	CRP;ADM;	3.36E-10	0.048338	Downstre
Downstream Neighbors of EGR1	232	EGR1	14	6	ACE;HIF1A	8.51E-10	0.054688	Downstre

You can view the overlapping genes from your target list with the list of all the potential targets. In this case shown below, the overlapping genes and potential targets of the JUN transcription factor are viewed by double-clicking on the particular list entry in the table, and….voila, now those genes (and those genes only) are displayed in the pathway viewer.

Downstream neighbors of JUN among the pre-eclampsia genes.



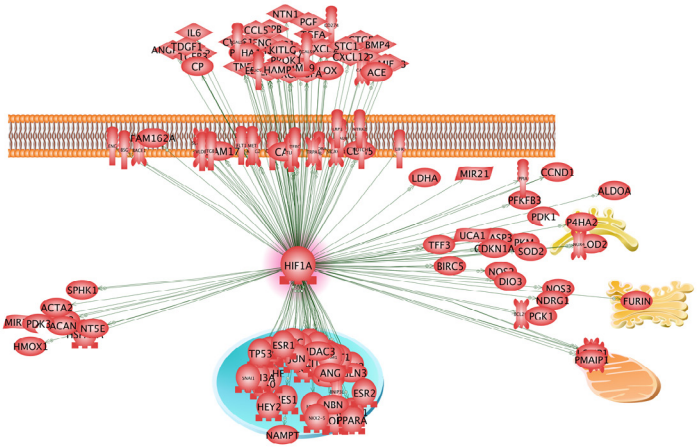
Now, here's the big trick! If you want to highlight just those genes in another pathway, all you have to do is: select and copy them, go to the other pathway (drum roll, please!), and … select clipboard content and then highlight with the color of your choice (Hint: if you are going to highlight more than once, the second time use a mix-in contrasting color so you can see both highlights together).



See how easy that was? Now, why don't you try it with HIF1A!

2.5 Small Molecule Regulators of Pre-eclampsia

HIF1A is not only one of the most important regulators of the FLT1 gene—which is the most highly cited gene in reference to pre-eclampsia in the literature—it also appears to be a major regulator for a significant number (17/41) of all the pre-eclampsia associated genes.



This is beginning to look interesting, so what can be observed about the biological processes controlled by HIF1A?

Well, first of all, you could identify all the genes with promoters known to be bound by HIF1A.

Hint: Copy and paste HIF1A into new pathway, select HIF1A, select "Add Neighbors," (downstream), and select Protein as the entity type and Promoter Binding as the relation type.

Now you can ask yourself: of all the genes under the potential control of HIF1A, what are some of the most common biological processes involved?

Select all the genes from the HIF1A transcriptional network, go to Tools, select "Enrichment Analysis," Analysis Type = "Find Pathways."

Go to "Ontologies," choose "biological process" from the GO sets, and then select "Find."

Name	# of Entities	Expanded # of Entities	Overlap	Percent Overlap	Overlapping Entities	p-value	Jaccard similarity	Hit type
response to hypoxia	249	250	41	16	SLC11A2, PPARA, ABCB...	4.03737E-35	9.27602E-2	biological_process
response to drug	494	495	44	8	ABCG2, GPX3, ABCB1, C...	3.70069E-26	6.43275E-2	biological_process
positive regulation of cell proliferation	513	513	41	7	GPER, r_Sox9, NAMPT, N...	1.27948E-22	5.81560E-2	biological_process
cellular response to hypoxia	105	105	21	20	SLC11A2, PMAIP1, PTGS...	3.04355E-20	6.62461E-2	biological_process
response to organic cyclic compound	239	239	26	10	GPX3, ABCB1, NAMPT, ...	7.13389E-18	5.82960E-2	biological_process
response to lipopolysaccharide	222	222	25	11	S100A9, SLCO1B3, SLC1...	1.38156E-17	5.81395E-2	biological_process
negative regulation of apoptotic process	642	642	39	6	r_Sox9, NKX2-5, HIF1A, ...	2.27499E-17	4.66507E-2	biological_process
positive regulation of transcription from RNA polym...	846	846	44	5	NR1H3, r_Sox9, PPARA, ...	5.17094E-17	4.25121E-2	biological_process
aging	210	210	22	10	NTRK2, NTRK1, DDIT3, T...	6.13156E-15	5.22565E-2	biological_process
positive regulation of apoptotic process	359	359	27	7	GPER, PMAIP1, ITGB1, S...	1.75947E-14	4.77876E-2	biological_process

The most enriched GO biological process for the HIF1A transcriptome is...?

Of what diagnostic parameter (see introduction) of pre-eclampsia does this finding make the most sense? Discuss with class.

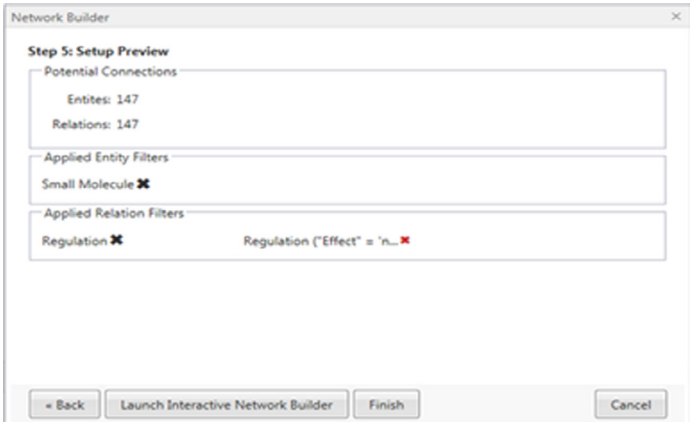
Are you ready for a little bit more?

What about the drugs and possible drug treatments that are used for handling patients (i.e. pregnant women) at risk for pre-eclampsia?

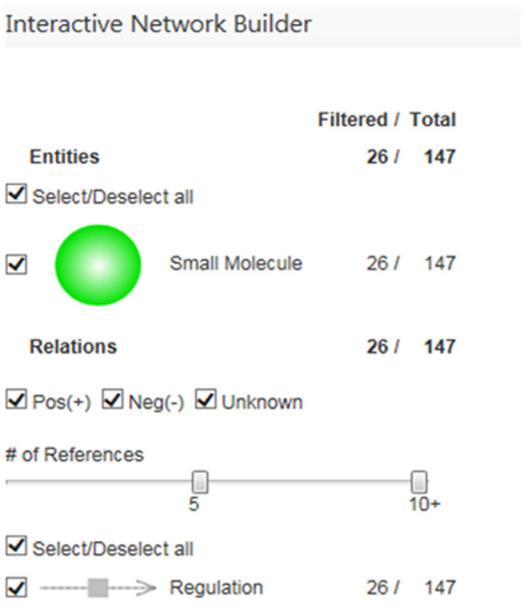
You have two quick ways (at least!) of investigating this question using Pathway Studio.

1. Test all small molecules associated in the literature with pre-eclampsia.
2. And then look for any clinical trials reported for pre-eclampsia treatments.

Create a new pathway using the pre-eclampsia disease entity as a starting point. Add small molecules as an entity type and relation = regulation, effect = negative.



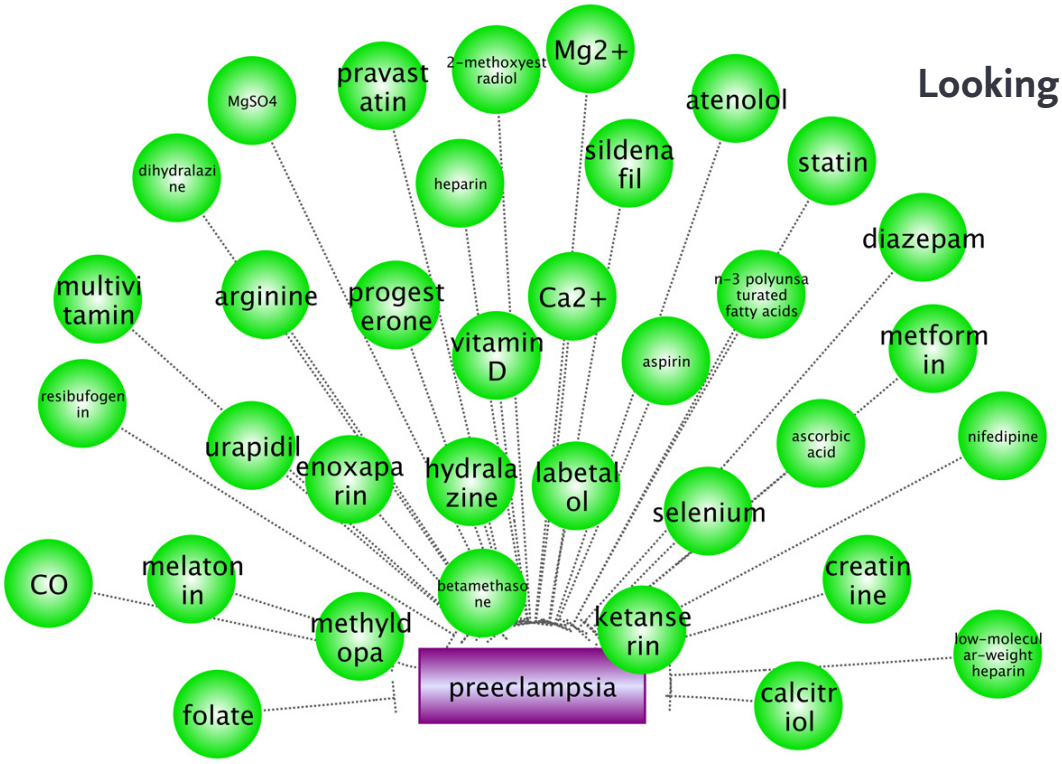
This will still give you a lot of relations!



Filter that down a little bit by going to the "Interactive Network Builder" and selecting for relations with five or more supporting articles.

Now that's a little more manageable!

Next, see how the graph view looks.



If you look at the "Relation Table View" and sort by the highest number of references, you find...aspirin!

Try Googling aspirin and pre-eclampsia and tell the class what you find.

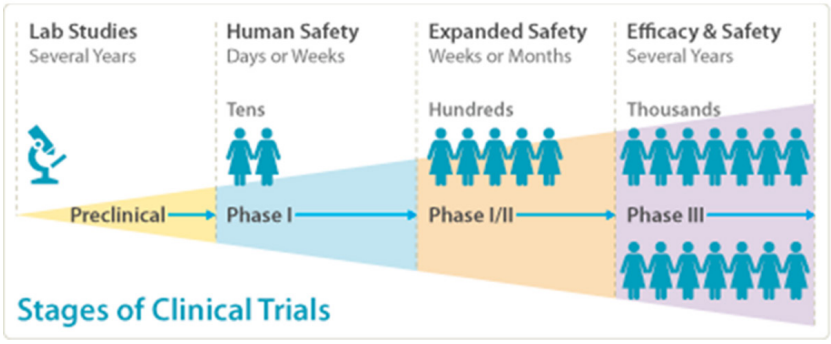
How is $MgSO_4$ used in the treatment of pre-eclampsia?

For what disease condition was the drug pravastatin originally developed?

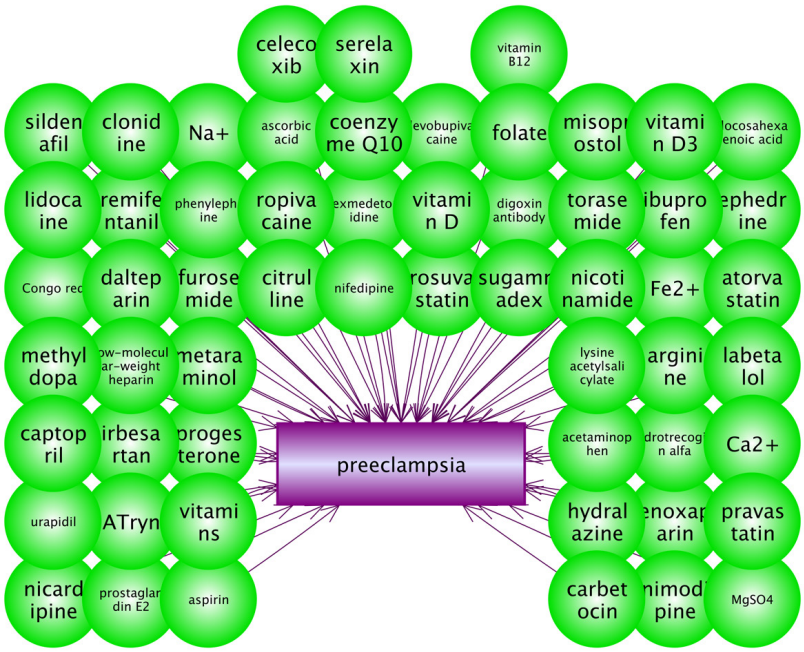
Relation	Object Type	Effect	Mechanism	# of References
aspirin ---| Pre-Eclampsia	Regulation	negative		197
MgSO4 ---| Pre-Eclampsia	Regulation	negative		167
Ca2+ ---| Pre-Eclampsia	Regulation	negative		94
Mg2+ ---| Pre-Eclampsia	Regulation	negative		66
ascorbic acid ---| Pre-Eclampsia	Regulation	negative		50
pravastatin ---| Pre-Eclampsia	Regulation	negative		37
arginine ---| Pre-Eclampsia	Regulation	negative		30
folate ---| Pre-Eclampsia	Regulation	negative		24
nifedipine ---| Pre-Eclampsia	Regulation	negative		24
heparin ---| Pre-Eclampsia	Regulation	negative		22
hydralazine ---| Pre-Eclampsia	Regulation	negative		21
low-molecular-weight heparin ---| Pre...	Regulation	negative		18
labetalol ---| Pre-Eclampsia	Regulation	negative		16
2-methoxyestradiol ---| Pre-Eclampsia	Regulation	negative		15
methyldopa ---| Pre-Eclampsia	Regulation	negative		14
ketanserin ---| Pre-Eclampsia	Regulation	negative		13
CO ---| Pre-Eclampsia	Regulation	negative		13
metformin ---| Pre-Eclampsia	Regulation	negative		12
progesterone ---| Pre-Eclampsia	Regulation	negative		10
multivitamin ---| Pre-Eclampsia	Regulation	negative		7

2.6 Drugs in Clinical Trials

And finally, clinical trials!



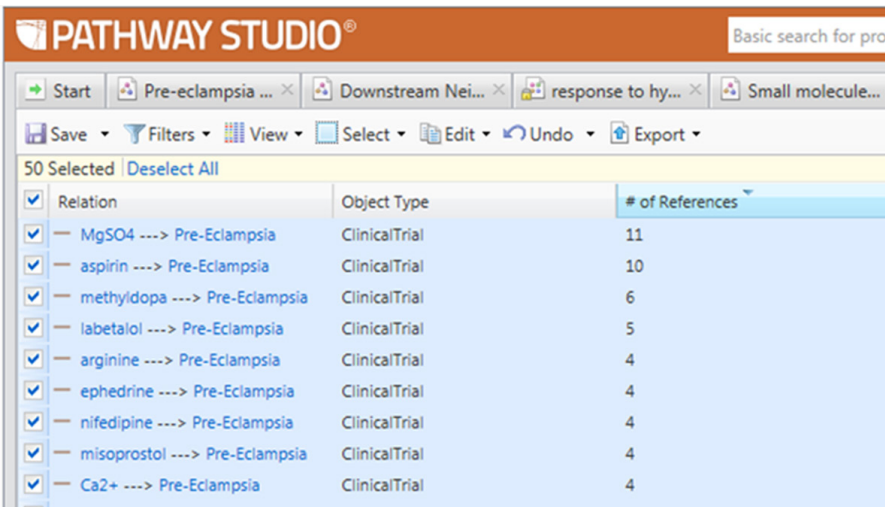
Select small molecules, same as before, and for relation type, pick ClinicalTrial.



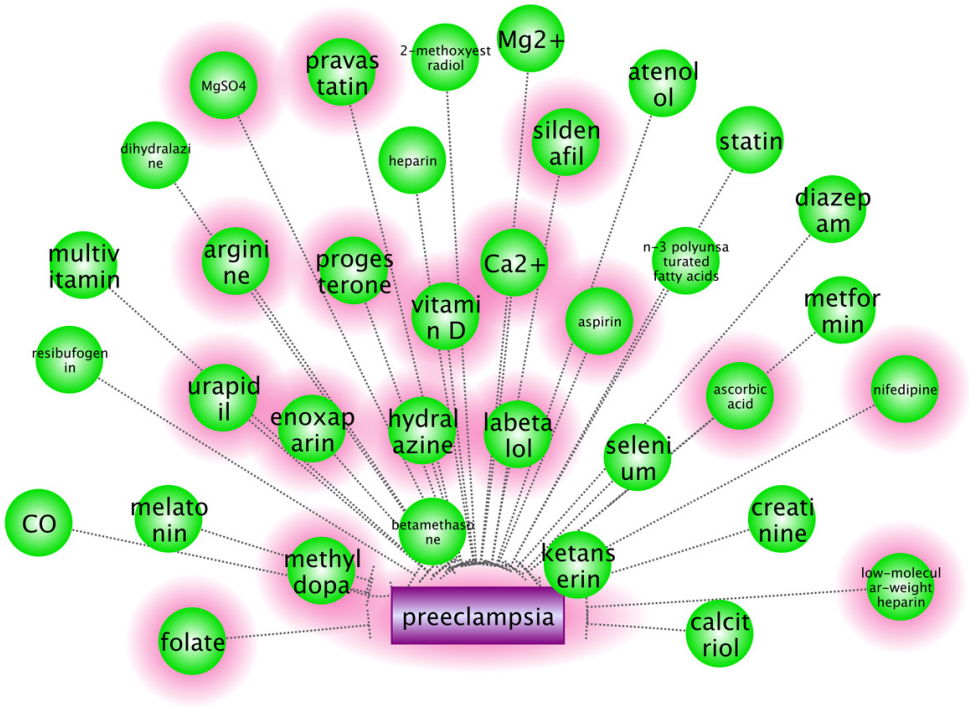
What are the most studied drugs in terms of number of clinical trials?

What is the rationale for the use of pravastatin in treating pre-eclampsia (requires linking out to the clinical trials record for answer)?

Hint: check Detailed Description field



Overlap of Clinical Trials and Small Molecules



Wait! Before you go, what did you learn today?

- "Network Builder" can be used to add relations to entities.
- "Relation Table View" and "# of References" can be used to find most cited relations.
- You can find transcription factors that bind to gene promoters or miRNAs that affect gene expression.
- Enriched upstream regulators can be identified for a group of genes using Sub-Network Enrichment Analysis (SNEA).
- Possible drug treatments can be identified using Small Molecules and Clinical Trials.

And that's a lot!

Study Questions 2

1. Identify genes/proteins that are linked to OCD. How many of them?
2. Identify drugs/small molecules linked to OCD. How many of them?
3. Identify transcription factors of protein FLT1. List top three by reference number.
4. What is the most cited transcription factor of FLT1?
5. Find miRNAs that regulate the protein FLT1? List top five by reference number. Which are predicted? How many are from the literature? List top two by reference number.
6. Identify the number of small molecules and protein functional classes related to the expression of the protein FLT1. List top three by reference number. How many have a negative effect on expression? How many have a positive effect on expression? (Hint: use Interactive Network Filter).
7. Identify top five enriched diseases by p-value using genes linked to pre-eclampsia (10+ references only). (Hint: use SNEA).
8. Identify top five by p-value enriched GO terms by p-value for genes linked to pre-eclampsia (10+ references only)? (Hint: use GSEA).

How Do I ... Gene/Protein Expression:

Exercise 2.1: Find proteins (transcription factors) that bind to the promoter of a gene(s)?

Finds transcription factors for genes (directly binding to promoters)

Step 1: Create New Pathway or within Pathway, select a protein(s)

Step 2: Select Add > Network Builder

Step 3: Select Expand Pathway, Advanced Expand Pathway Tool

Step 4: Select Direction: Upstream

Step 5: Entities: Select "Protein" Relations: Select "Promoter Binding"

Continue to Network Builder Step 5 and click Launch Interactive Network Builder or Finish

Exercise 2.2: Find predicted miRNAs that may regulate expression of a gene(s)?

Finds predicted miRNA targets (from public prediction datasets - identified in "source" annotation field of the relation)

Step 1: Create New Pathway or within Pathway, select a protein(s)

Step 2: Select Add > Network Builder

Step 3: Select Expand Pathway, Advanced Expand Pathway Tool

Step 4: Select Direction: Upstream

Step 5: Entities: Select "Protein" Relations: Select "miRNAEffect"

Continue to Network Builder Step 5 and click Launch Interactive Network Builder or Finish

Switch to the Relation Table view: add "Source" as column to the table and sort on that column. Relations with Source annotation of public datasets "miRanda" "TargetScan" "PicTar" "TarBase" etc. are predicted miRNA targets.

Exercise 2.3: Find known miRNAs that regulate expression of a gene(s)?

Finds literature confirmed miRNA targets. (Differentiate from predicted miRNAEffect relations by reference annotation)

Step 1: Create New Pathway or within Pathway, select a protein(s)

Step 2: Select Add- > Network Builder

Step 3: Select Expand Pathway, Advanced Expand Pathway Tool

Step 4: Select Direction: Upstream

Step 5: Entities: "Protein" Relations: Select "miRNAEffect"

Continue to Network Builder Step 5 and click Launch Interactive Network Builder or Finish

Switch to the Relation Table view, add "Sentence" as a column to the table and sort on that column. Any relation with a supporting sentence is a literature confirmed relation.

Exercise 2.4: Find proteins that are involved in the expression of a gene(s)?

Finds both direct expression regulators (promoterbinding) and proteins with possibly an indirect effect on expression

Step 1: Create New Pathway or within Pathway, select a protein(s)

Step 2: Select Add- > Network Builder

Step 3: Select Expand Pathway, Advanced Expand Pathway Tool

Step 4: Select Direction: Upstream

Step 5: Entities: Select "Protein" Relations: Select "Promoter Binding + Expression"

Continue to Network Builder Step 5 and click Launch Interactive Network Builder or Finish

Exercise 2.5: Does a group of genes share common expression regulators?

Finds common expression regulators that impact multiple targets in a select group. If Promoter Binding + Expression gives too many results, try examining only Promoter Binding. If results are too low, use the back button before you launch Interactive Network Builder to step 2 and expand to 2 or greater steps

Step 1: Create New Pathway or within Pathway, select a protein(s)

Step 2: Select Add- > Network Builder

Step 3: Select Expand Pathway, Expression Regulators

Continue to Network Builder Step 5 and click Launch Interactive Network Builder or Finish

Exercise 2.6: Does a group of proteins regulate expression of similar gene(s)?

Finds targets that share a common regulator from a select group. If Promoter Binding + Expression gives too many results, try examining only Promoter Binding. If results are too low, use the back button before you launch Interactive Network Builder to step 2 and expand to 2 or greater steps.

Step 1: Create New Pathway or within Pathway, select a protein(s)

Step 2: Select Add- > Network Builder

Step 3: Select Expand Pathway, Expression Targets

Continue to Network Builder Step 5 and click Launch Interactive Network Builder or Finish

Module 3

Variant Analysis I

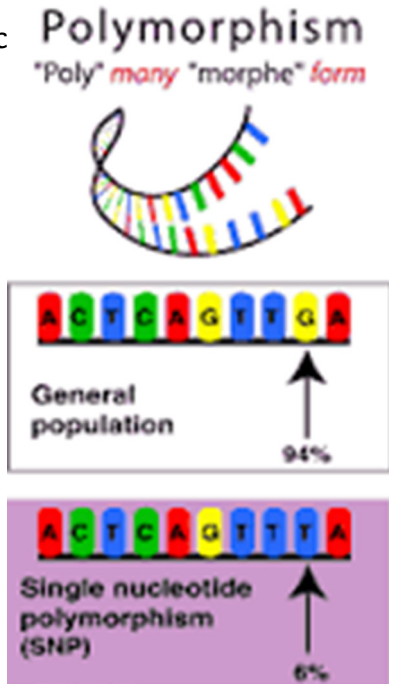
Contents

3.1 Introduction to Variant Analysis in Genomic Sequence Data

The dbSNP database [http://www.ncbi.nlm.nih.gov/snp] currently has more than 200 million known human SNPs, or single nucleotide polymorphisms. While many SNPs may have no biological impact, and others may simply provide the basis for benign differences between individuals, it is the variations in human genomic sequences that lead to medically relevant phenotypes that are of great interest to researchers. These include DNA nucleotide changes that cause diseases, greater susceptibility to medical conditions, and variations in responses to medications.

3.2. The dbSNP Database

In this training module, you will learn to search the dbSNP database to find variants known to be associated with a specific disease.

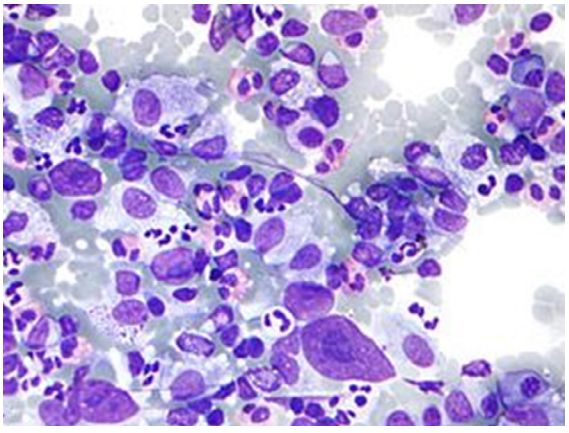


As there are a large number of SNPs in any given genome, identifying specific variants of interest involves applying multiple filters, which can be based on a variety of factors such as:

- Is the mutation within the coding region of a gene?
- Does the SNP affect the protein's sequence?
- Is the mutation in a specific protein's amino acid position known to be generally conserved?
- Is the protein mutation within a region that is known to be conserved across species other than just human?
- How common or frequently is a particular SNP found within the standard 1000 Genome Project reference data? [http://www.1000genomes.org/

3.3 Hodgkin's Lymphoma—Deleterious SNPs

Hodgkin's Lymphoma (HL) is a type of lymphoma in which cancer originates from a specific type of white blood cells called lymphocytes.



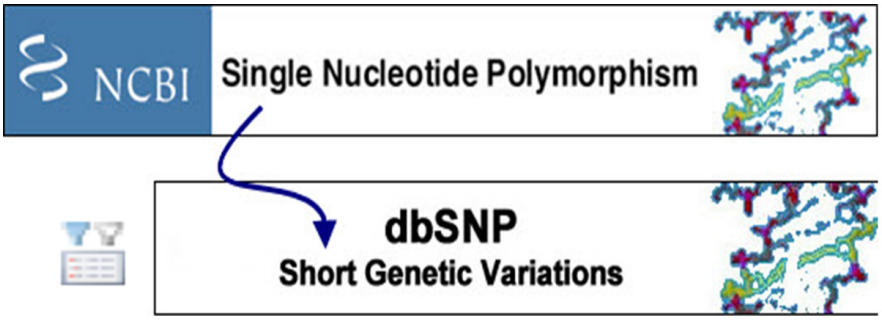
Micrograph showing Hodgkin's Lymphoma (Field stain)

A history of infectious mononucleosis due to infection by Epstein–Barr virus (EBV) may increase risk of HL, but the precise contribution of Epstein–Barr virus remains largely unknown.

Hodgkin's Lymphoma is characterized by the orderly spread of disease from one lymph node group to another and by the development of systemic symptoms with advanced disease.

Let's find some SNP(s) that are potentially deleterious to the function of proteins that are known to be associated with Hodgkin's Lymphoma.

Let's get started!

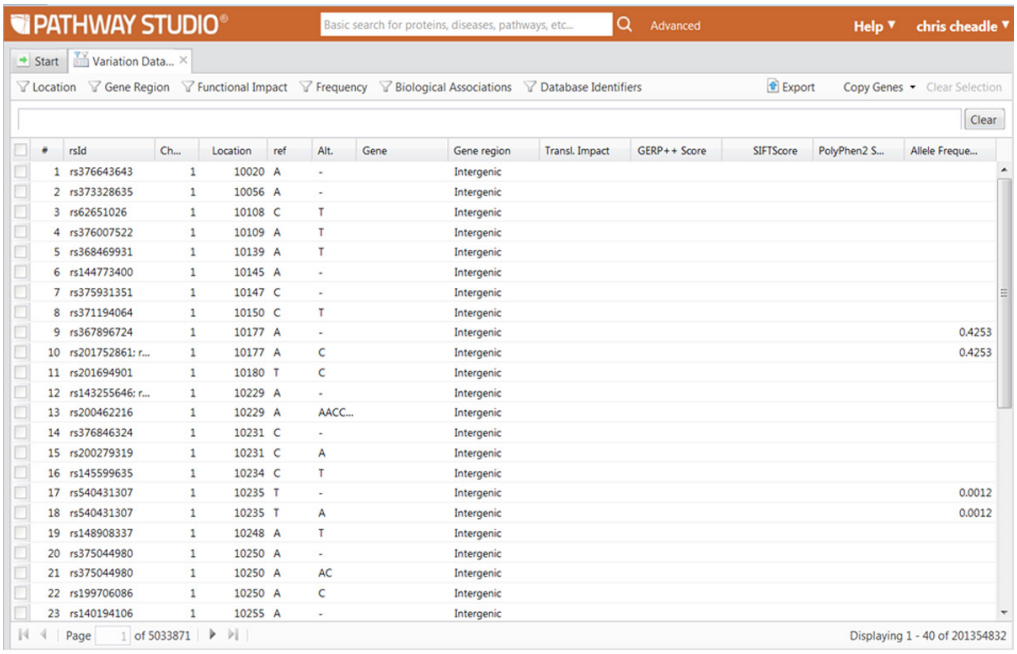


And what better place to begin than at the beginning,

in this case, the Pathway Studio Start Tab.

Click on the Variation database icon:

Which will bring you to the variation database (from the dbSNP database) and it looks like this:



If you check the lower right-hand edge of your screen you will see the message displaying "1-40 of 201354832".

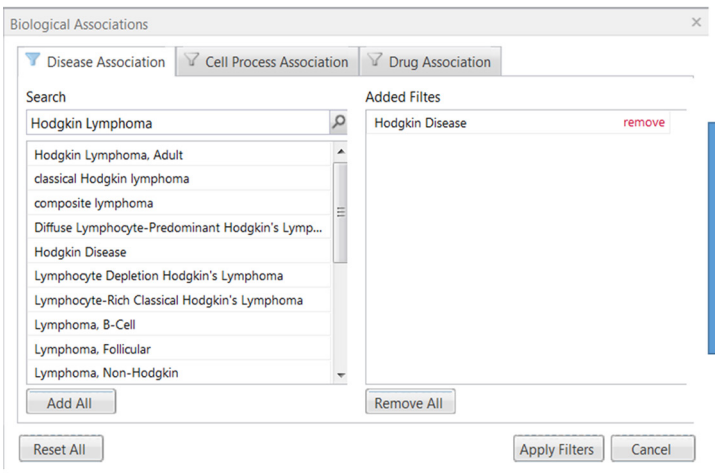
That's because there have been over 200 million SNPs identified and reported to date across the world.

Along the top ruler are the filters that we will use to drill down for getting useful data.

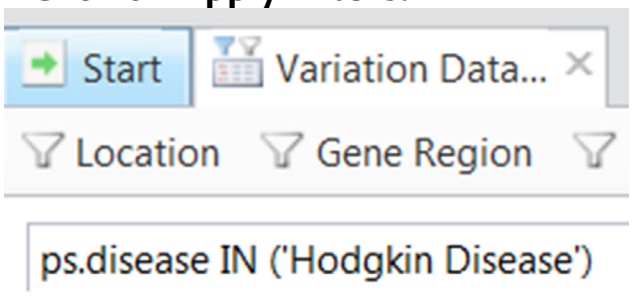
3.4 Biological – Disease Associations

Our first stop will be Biological Associations where we will select Disease Association and type in Hodgkin's Lymphoma (we could also have picked Cell Process Association for things like "apoptosis" or Drug Association for small molecule drugs, all of which are annotated in the main Pathway Studio database).

The search will return a number of related diseases (you could select more than one). For this example, just select "Hodgkin Disease"



Click on Apply Filters:



Click on Apply Filters

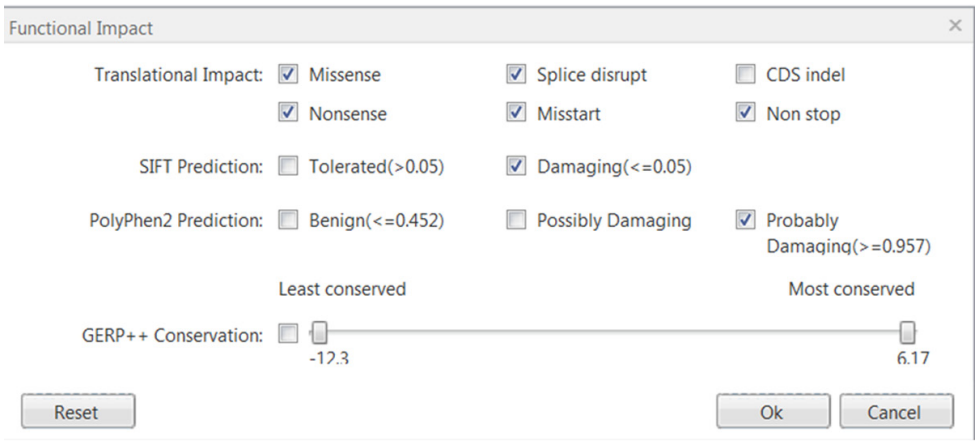
This search will be tracked in the Added Filters bar just below the Filter tabs. This is very helpful to follow as you add more filters. (note: This bar is not interactive, i.e. you cannot add or delete anything by typing. The only way to go back is to clear everything, but don't worry - the filters are easy to use).

Now we're down to about 2 million SNPs (check it and see!).

What we've done is selecting all those SNPs that are mapped to the subset of genes associated with Hodgkin's Lymphoma in the Pathway Studio database (from the scientific literature).

But that's still way too much, so let's keep filtering!

3.5 Functional Impact Filter



Next select the **type of SNP** with respect to the functional impact on the protein.

From the tool bar select the Functional Impact filter. In this example, missense, splice disrupt, nonsense, misstart and non-stop mutation types are selected for Translational Impact.

Also, for SIFT, select "damaging" and for PolyPhen2 select "probably damaging."

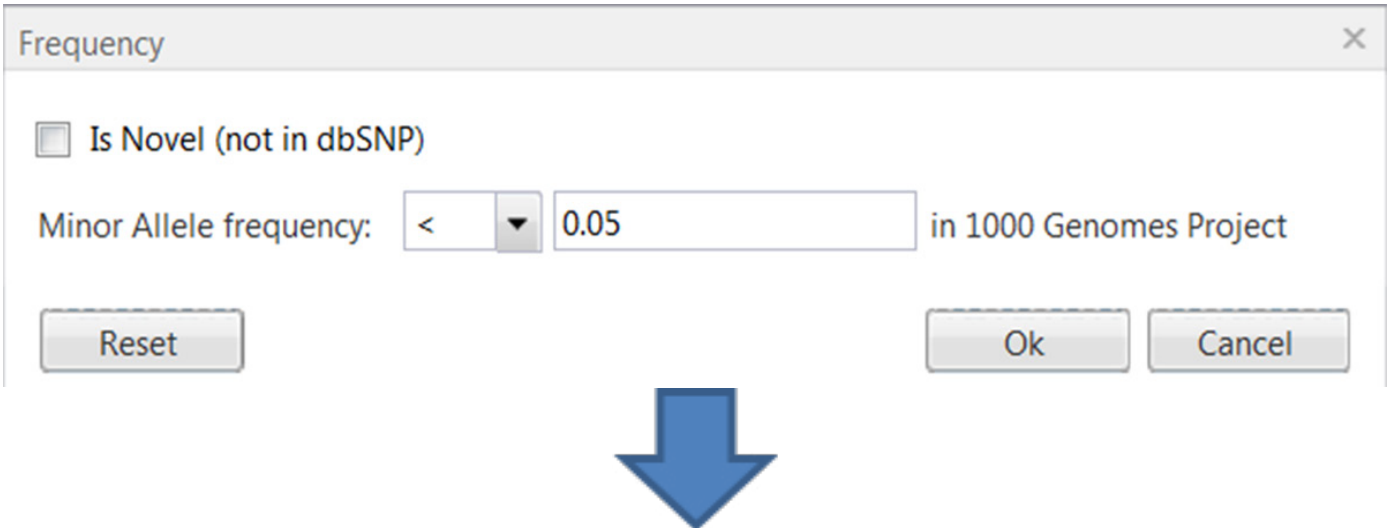
- SIFT predicts whether an amino acid substitution affects protein function based on sequence homology and the physical properties of amino acids.

- PolyPhen2 predicts possible impact of amino acid substitutions on the structure and function of human proteins using straightforward physical and evolutionary sequence comparative considerations.

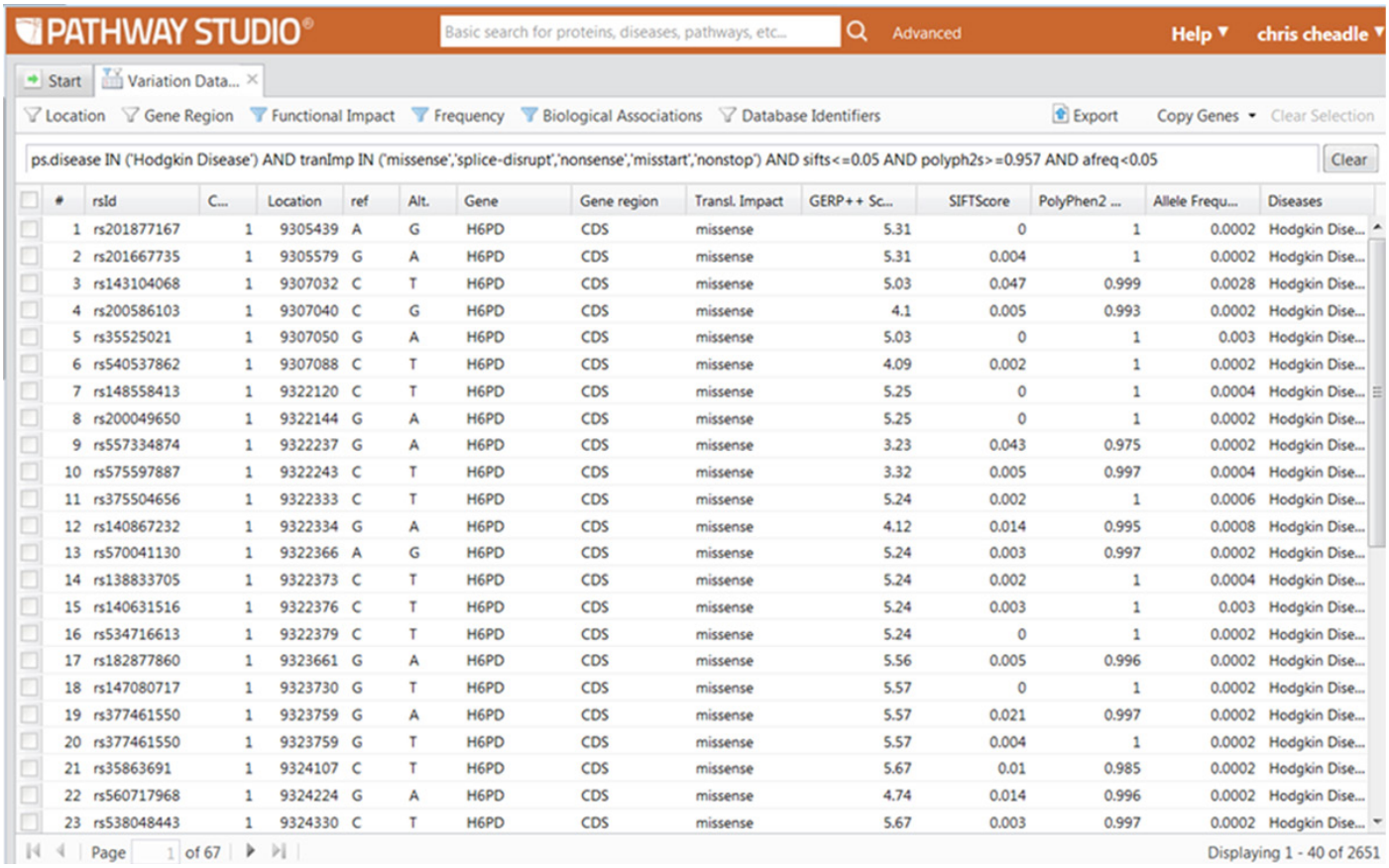
Now, we're down to a little over 500,000 SNPs to examine. That's still a lot! We have to keep going!

Let's reduce this number further by examining only those SNPs that are not commonly found within the 1000 Genome Project data (SNPs associated with disease are expected to be **uncommon** in the general population).

In this example only SNPs that are present in less than 5% of the genomes in the 1000 Genome Project will be considered.



Now we're down to a manageable number of SNPs (2738).



#	rsId	C...	Location	ref	Alt.	Gene	Gene region	Transl. Impact	GERP++ Sc...	SIFTScore	PolyPhen2 ...	Allele Frequ...	Diseases
1	rs201877167	1	9305439	A	G	H6PD	CDS	missense	5.31	0	1	0.0002	Hodgkin Dise...
2	rs201667735	1	9305579	G	A	H6PD	CDS	missense	5.31	0.004	1	0.0002	Hodgkin Dise...
3	rs143104068	1	9307032	C	T	H6PD	CDS	missense	5.03	0.047	0.999	0.0028	Hodgkin Dise...
4	rs200586103	1	9307040	C	G	H6PD	CDS	missense	4.1	0.005	0.993	0.0002	Hodgkin Dise...
5	rs35525021	1	9307050	G	A	H6PD	CDS	missense	5.03	0	1	0.003	Hodgkin Dise...
6	rs540537862	1	9307088	C	T	H6PD	CDS	missense	4.09	0.002	1	0.0002	Hodgkin Dise...
7	rs148558413	1	9322120	C	T	H6PD	CDS	missense	5.25	0	1	0.0004	Hodgkin Dise...
8	rs200049650	1	9322144	G	A	H6PD	CDS	missense	5.25	0	1	0.0002	Hodgkin Dise...
9	rs557334874	1	9322237	G	A	H6PD	CDS	missense	3.23	0.043	0.975	0.0002	Hodgkin Dise...
10	rs575597887	1	9322243	C	T	H6PD	CDS	missense	3.32	0.005	0.997	0.0004	Hodgkin Dise...
11	rs375504656	1	9322333	C	T	H6PD	CDS	missense	5.24	0.002	1	0.0006	Hodgkin Dise...
12	rs140867232	1	9322334	G	A	H6PD	CDS	missense	4.12	0.014	0.995	0.0008	Hodgkin Dise...
13	rs570041130	1	9322366	A	G	H6PD	CDS	missense	5.24	0.003	0.997	0.0002	Hodgkin Dise...
14	rs138833705	1	9322373	C	T	H6PD	CDS	missense	5.24	0.002	1	0.0004	Hodgkin Dise...
15	rs140631516	1	9322376	C	T	H6PD	CDS	missense	5.24	0.003	1	0.003	Hodgkin Dise...
16	rs534716613	1	9322379	C	T	H6PD	CDS	missense	5.24	0	1	0.0002	Hodgkin Dise...
17	rs182877860	1	9323661	G	A	H6PD	CDS	missense	5.56	0.005	0.996	0.0002	Hodgkin Dise...
18	rs147080717	1	9323730	G	T	H6PD	CDS	missense	5.57	0	1	0.0002	Hodgkin Dise...
19	rs377461550	1	9323759	G	A	H6PD	CDS	missense	5.57	0.021	0.997	0.0002	Hodgkin Dise...
20	rs377461550	1	9323759	G	T	H6PD	CDS	missense	5.57	0.004	1	0.0002	Hodgkin Dise...
21	rs35863691	1	9324107	C	T	H6PD	CDS	missense	5.67	0.01	0.985	0.0002	Hodgkin Dise...
22	rs560717968	1	9324224	G	A	H6PD	CDS	missense	4.74	0.014	0.996	0.0002	Hodgkin Dise...
23	rs538048443	1	9324330	C	T	H6PD	CDS	missense	5.67	0.003	0.997	0.0002	Hodgkin Dise...

3.6 Export Genes to Pathway Studio

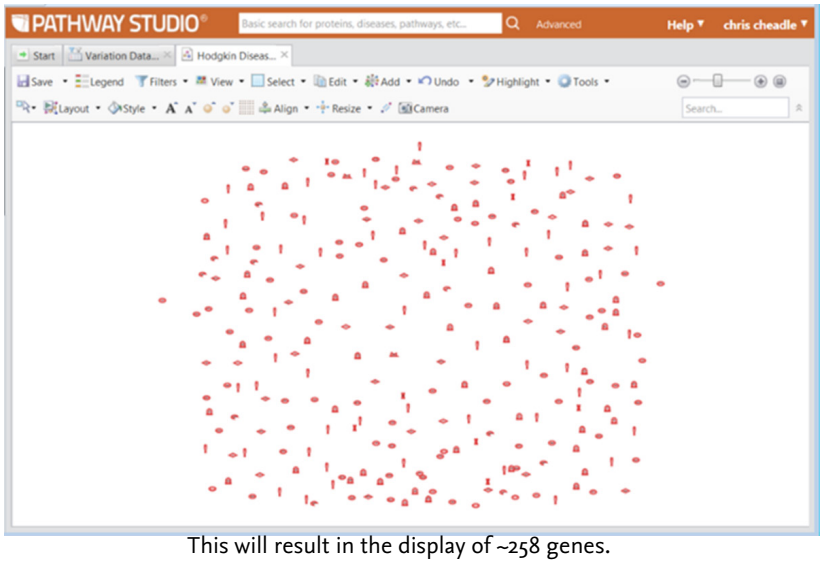
Note that the filter commands are reprised on the Added Filters bar and reflected as well in the table annotations below for the individual SNPs.

Our next step will be to export the genes* to which these SNPs are mapped into the Pathway Studio main program for further analysis.

*As many of the genes in the list have multiple SNPs, the number of genes to examine is far less than 2700.

To further examine these identified genes, go to: Copy Genes > Copy First 1000 Genes. This will copy the gene names to the clipboard.

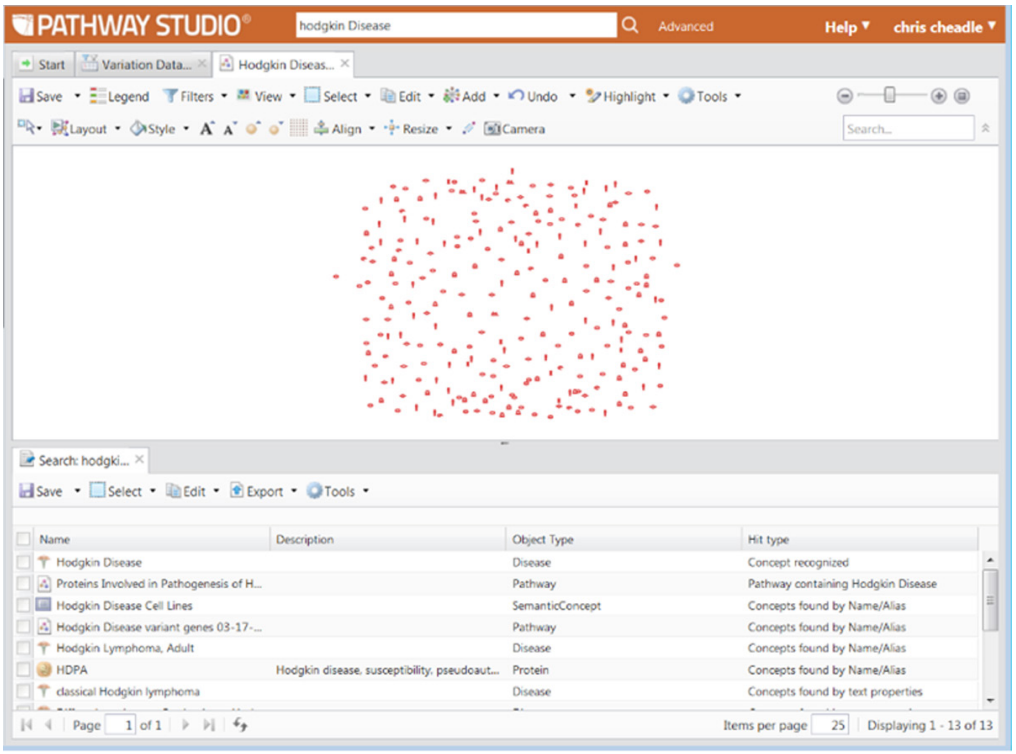
Go to Start tab, Create New Pathway and paste the gene list into the Pathway Viewer window (this will take a minute, so be patient, there are a lot of genes!).



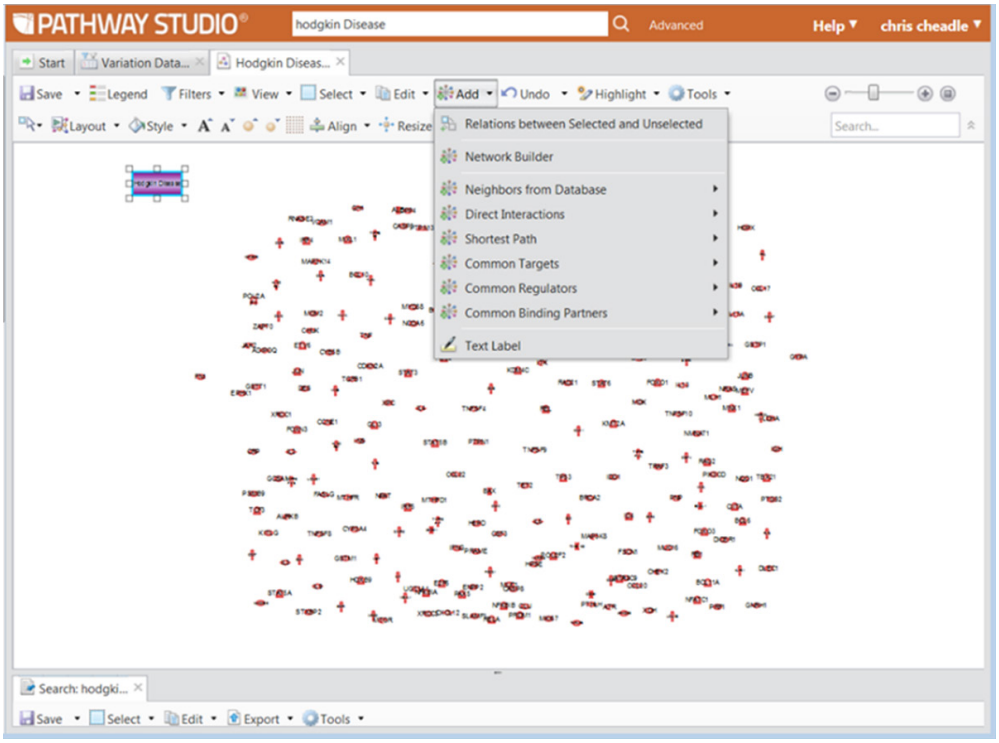
This will result in the display of ~258 genes.

Next, type in "Hodgkin Lymphoma" into the Search box at top of page

Select "Hodgkin Disease" and proceed to the "Add" Tab. **Select, copy, and paste the return (from table below) into the Pathway**

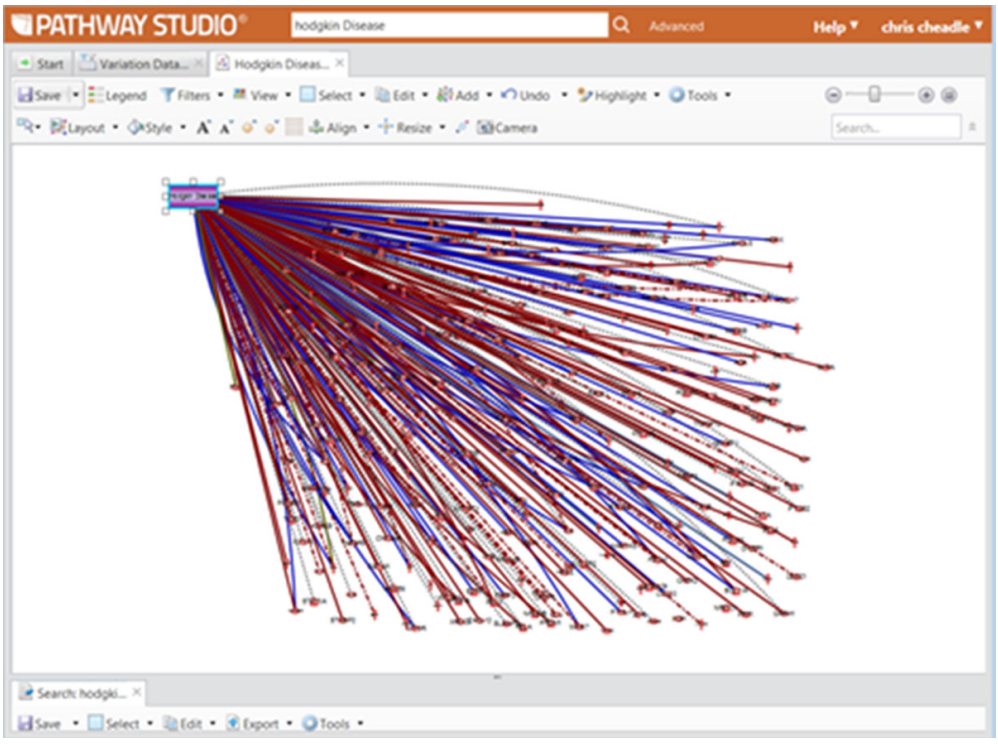


Add “Relations between Selected and Unselected”



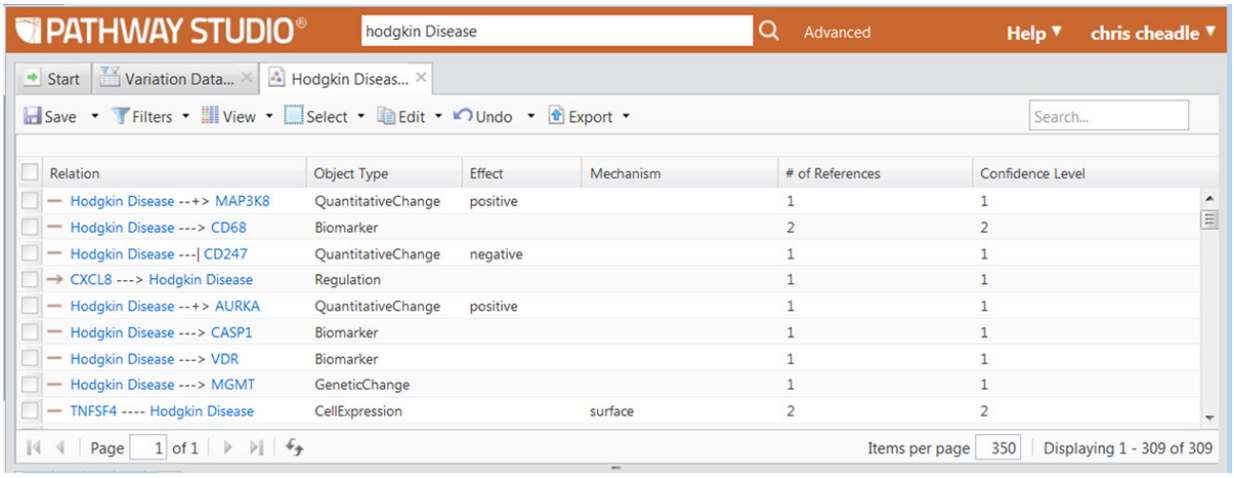
What we’ve done is to connect all the Hodgkin Disease variant genes to the Disease entity “Hodgkin Disease” in the Pathway Studio database using all the literature based relationships found in the database between these genes and that disease.

You should get something that looks like this:



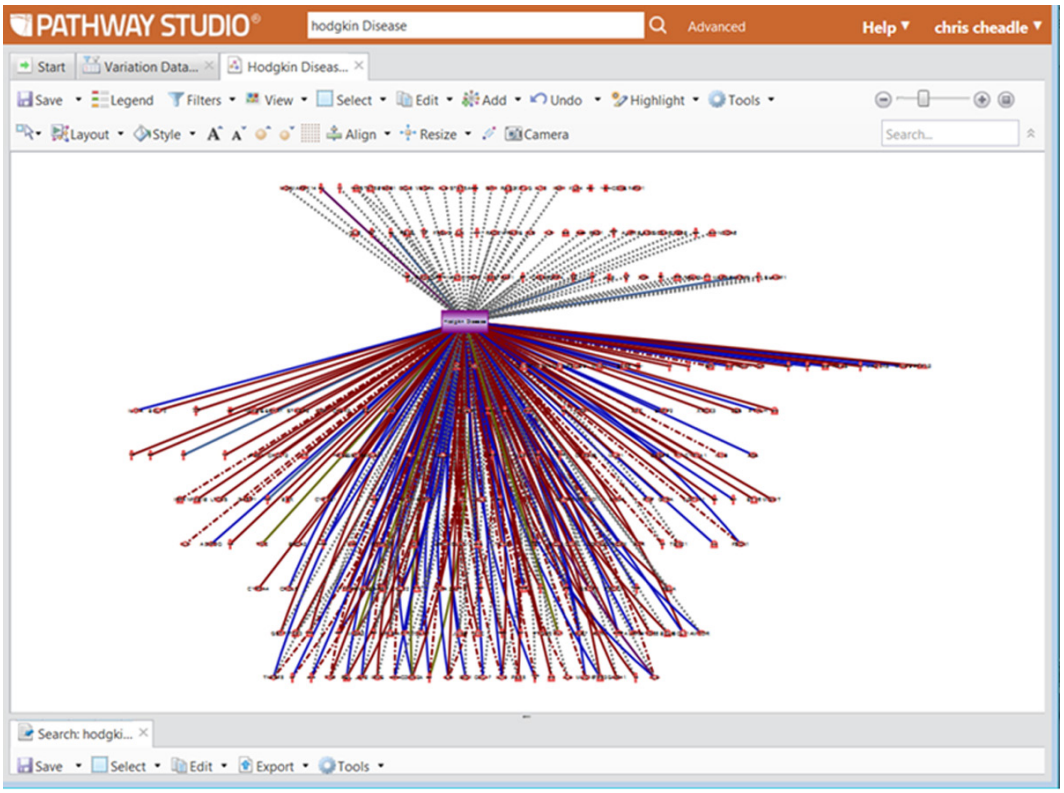
If you click View -> Entity Table View, you will find a list of the genes (246) plus one disease (Hodgkin Disease)

There's a lot of useful information in this table!



If you click View -> Relation Table View, you will find a list of all the relations (309). Note: There are more relations for this network than there are entities because a single entity can have multiple relations in the network.

Hint: to get this look, select Layout -> Hierarchical.



OK, if you got this far alright (right!), now, we can begin to systematically explore the biology of our selected genes using some of the many Pathway Studio enrichment analysis tools.

3.7 Associate Genes with Hodgkin's Disease

So you'll recall that we selected these genes in the variant analysis tool precisely because they are associated in some way with Hodgkin Disease (see above).

Let's test that right now;

Select all the genes in the Pathway Viewer (or Entity Viewer if you are in that window).

Click on Tools -> Enrichment Analysis of Selected Entities

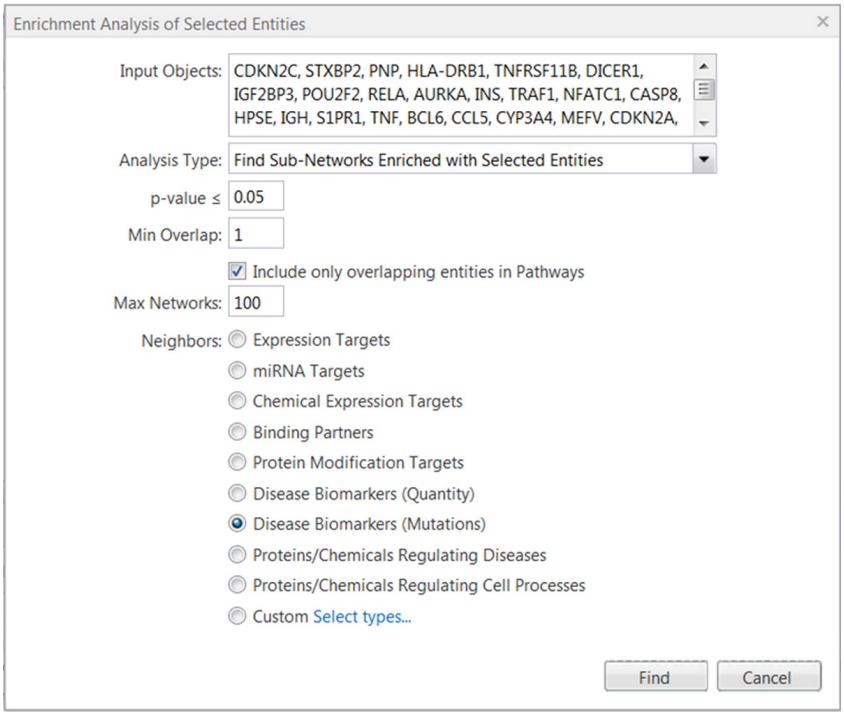
The Input Objects are the genes you selected (scroll down in this box and you will find the total number of genes selected [Total: 246], this is a good way to double-check your selection).

The Analysis Type is Find Sub-Networks Enriched with Selected Entities

Skip down to Neighbors (we'll accept all the default parameters for this example) and click the radio button for Disease Biomarkers (Mutations).

3.8 Sub-Network Enrichment Analysis, Disease and Cell Process

Sub-Network Enrichment Analysis (SNEA) is a powerful form of causal reasoning for network analysis*. In this case, the "Gene Set Seed" type selected is Disease. The SNEA algorithm will rapidly evaluate the submitted gene list for enrichment of all disease types in the database found upstream of these genes and connected to them by the relation "GeneticChange."

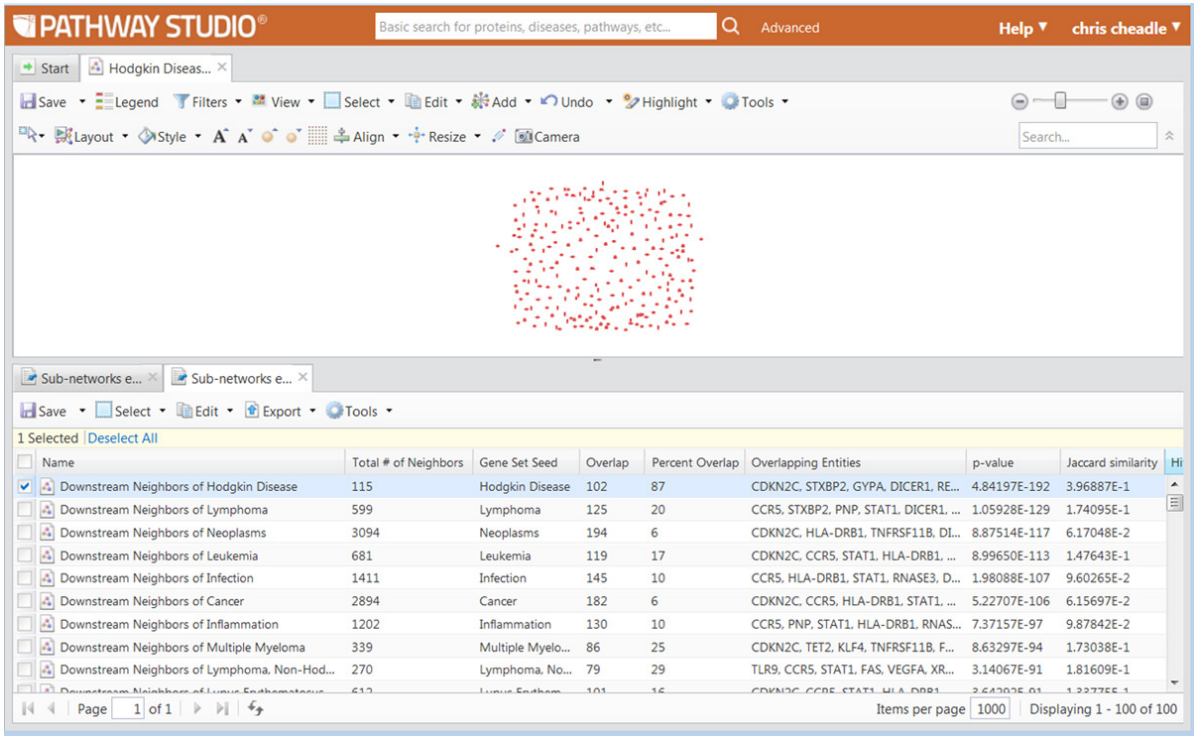


In a sense, it is asking the simple question: "Are any particular diseases over-represented in this particular list of genes?"

*Sivachenko AY, Yuryev A, Daraselia N, Mazo I (2007) Molecular networks in microarray analysis. Journal of bioinformatics and computational biology 5: 429–456.

And the answer is...? You got it! **Hodgkin's Disease**, followed by Lymphoma, Neoplasms Leukemia, and Infection. Clearly, mutations in these variant-selected genes are highly involved in multiple diseases for both cancer and inflammation.

There's a lot of useful information in this table!

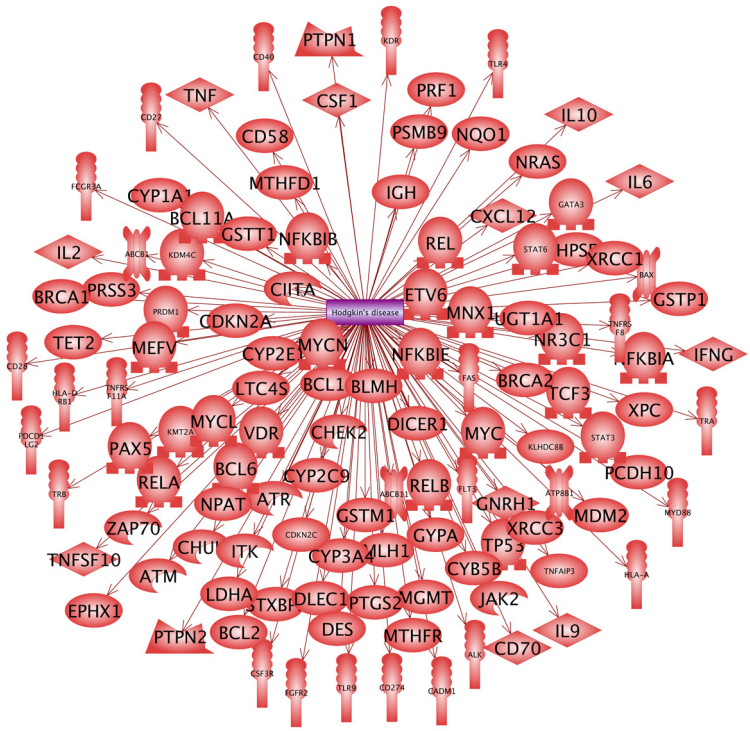


So, there is a high degree of overlap (108/257, 42%) between our variant gene list and the genes annotated as mutated in Hodgkin Disease in the Pathway Studio database.

Once again, this isn't surprising because this group is a subset of the group of all Hodgkin Disease associated genes which we used originally to filter the variant database (pX).?

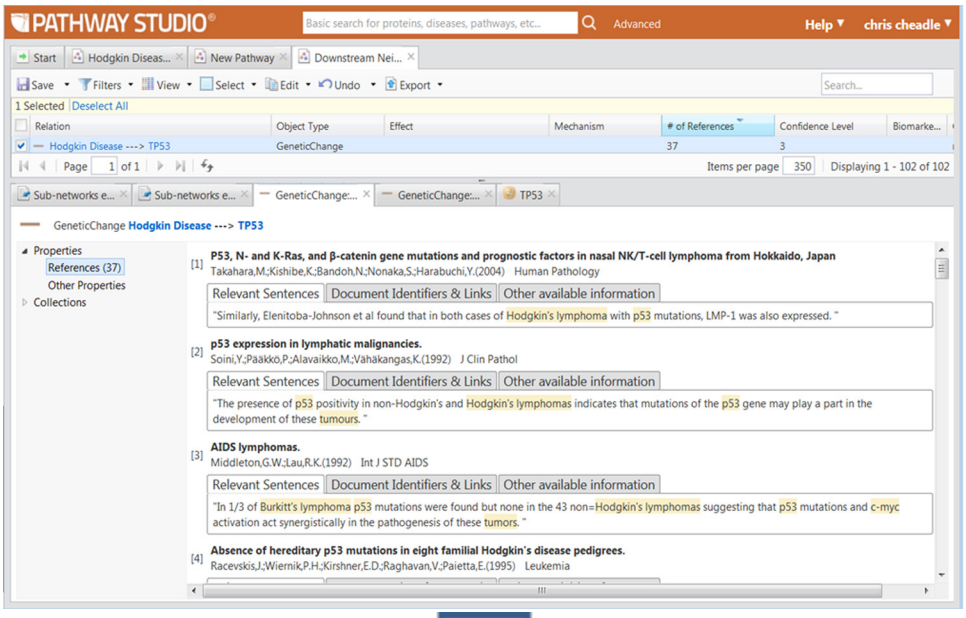
But it is reassuring – everything is working!

Let's now take a peek at these SNEA-identified genes by double-clicking on the Overlapping Entities box on the Downstream Neighbors of Hodgkin Disease row. This will open a new window for these genes in the Pathway Viewer.

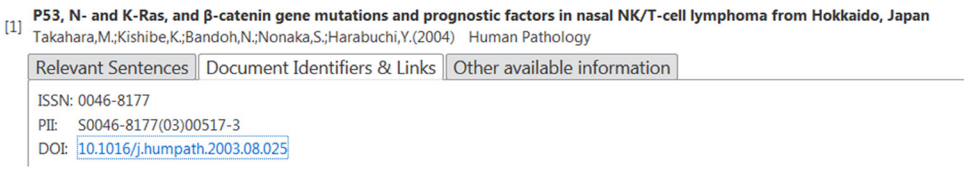


3.9 Drill Down!

If we go to View -> Relation Table View and sort by # of References (Descending) we can immediately see the most studied genetic changes associated with Hodgkin Disease.

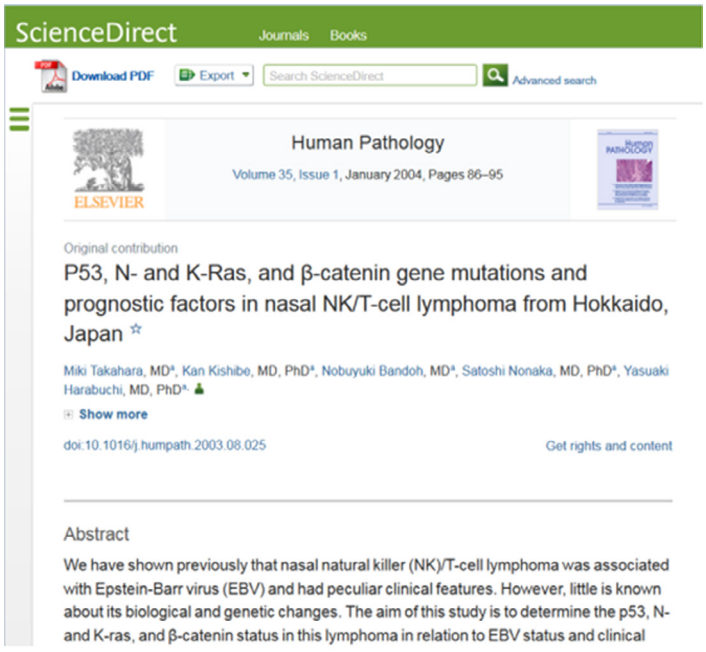


For any relation in the Relation Table View you can drill down and see the documents and even the sentences used to derive those biological facts.



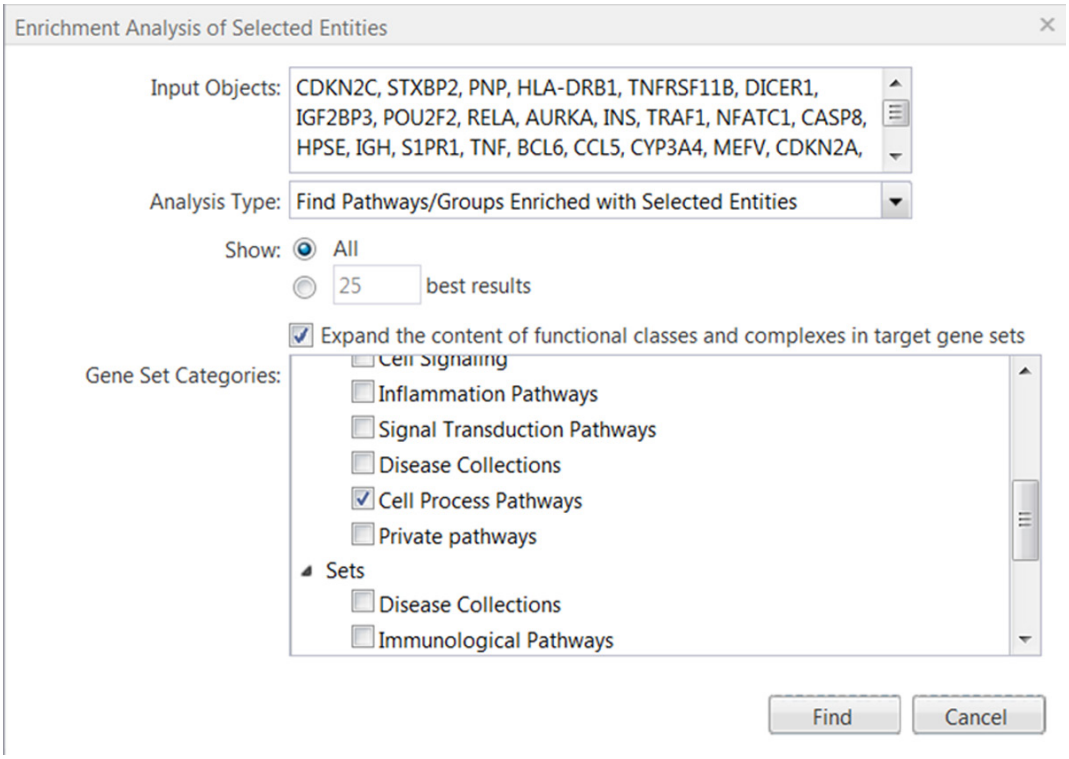
Link-out to PubMed and full text articles*

*depending on user subscription status



Let's finish up with a quick analysis of the potential biology for all the variant selected genes.

Just like we did before, click on Tools -> Enrichment Analysis of Selected Entities.



The **Input Objects** are the same genes as before (246).

The **Analysis Type,** this time, is Find Pathways/Groups Enriched with Selected Entities.

Under **Gene Set Categories**, check" Cell Process Pathways" (make sure nothing else is selected!) and then click Find…

Name	# of Entities	Expanded # of Entiti...	Overlap	Percent Overlap	Overlapping Entities	p-value
Adipokines Production by Adipocyte	58	99	18	18	RELA, MAP3K14, N...	1.77383E-14
Apoptosis	94	171	22	12	RELA, MAP3K14, N...	2.67863E-14
Cell Cycle	140	304	20	6	RB1, CDKN2A, AUR...	9.36480E-8
G2/M DNA Damage Checkpoint	22	30	5	16	CHEK2, ATR, ATM, T...	1.29757E-4
G0/G1 Cell Cycle Phase Transition	52	101	8	7	RELA, RB1, STAT1, ...	2.82067E-4
G1/S DNA Damage Checkpoint	27	81	7	8	CHEK2, MDM2, AT...	4.05942E-4

We can see that 5 out of the 6-top enriched cellular processes for the Hodgkin Disease DNA variants involve mutations in genes controlling apoptosis and the cell cycle, key regulatory events in cancer initiation and progression. …and we're….

But wait, before we go let's do a quick review of what we learned today.

We learned how to access the Pathway Studio variant database.

We learned how to perform a Sub-Network Enrichment Analysis (SNEA) to find diseases enriched in a list of genes by mutation relations.

We learned how to drill down to the actual article references underlying relations.

All in all, not a bad day's work!

And **now** we're done, see you next time!

Study Questions 3

1. Identify SNPs that are related to the disease SCA3, and tell the total number of SNPs.

2. Apply “Functional Impact” filter mentioned in the text of Module 3. How many SNPs are left?

3. Apply “Frequency” mentioned in the text of Module 3. How many SNPs are left?

4. Use “Copy Genes” feature with the remaining SNPs in question 3. How many genes do you have?

5. Use SNEA to identify top five diseases sorted by p-value related to the genes in Question 4.

For Physical Interaction with Proteins, How Do I ...

Exercise 3.1: Find proteins that bind to a protein?

Identifies protein binding partners (no additional regulatory event known). Binding relations have no Direction (Direct Regulation is regulation through a direct physical interaction and can also be considered here.)

Step 1: Create New Pathway or within Pathway, select a protein(s)

Step 2: Select Add- > Network Builder

Step 3: Select Expand Pathway, Physical Interactions

Continue to Network Builder Step 5 and click Launch Interactive Network Builder or Finish filter: Binding

Exercise 3.2: Find small molecules that bind to a protein?

Finds small molecules that regulate the activity of a protein through a direct physical interaction (Drugs/non-naturally occurring small molecules included in ChemEffect data)

Step 1: Create New Pathway or within Pathway, select a protein(s)

Step 2: Select Add- > Network Builder

Step 3: Select Expand Pathway, Advanced Expand Pathway Tool

Step 4: Select Direction: Upstream

Step 5: Entities: Select “Small Molecule” Relations: Select “Direct Regulation”

Continue to Network Builder Step 5 and click Launch Interactive Network Builder or Finish

Exercise 3.3: Find proteins that regulate a protein through a direct physical interaction?

Finds proteins that regulate the activity of a target protein through a direct physical interaction. Can also consider "Protmodification" relations.

Step 1: Create New Pathway or within Pathway, select a protein(s)

Step 2: Select Add- > Network Builder

Step 3: Select Expand Pathway, Advanced Expand Pathway Tool

Step 4: Select Direction: Upstream

Step 5: Entities: Select "Protein" Relations: Select "Direct Regulation"

Continue to Network Builder Step 5 and click Launch Interactive Network Builder or Finish

Exercise 3.4: Find small molecules that regulate a protein through direct physical interactions?

Finds small molecules that regulate the activity of a protein through a direct physical interaction (Drugs/non-naturally occurring g small molecules included in ChemEffect data)

Step 1: Create New Pathway or within Pathway, select a protein(s)

Step 2: Select Add- > Network Builder

Step 3: Select Expand Pathway, Advanced Expand Pathway Tool

Step 4: Select Direction: Upstream

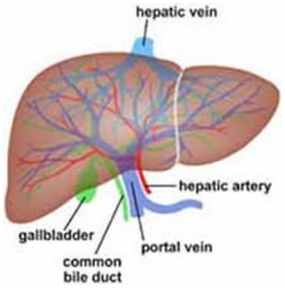
Step 5: Entities: Select "Small Molecule" Relations: Select "Direct Regulation"

Continue to Network Builder Step 5 and click Launch Interactive Network Builder or Finish

Module 4

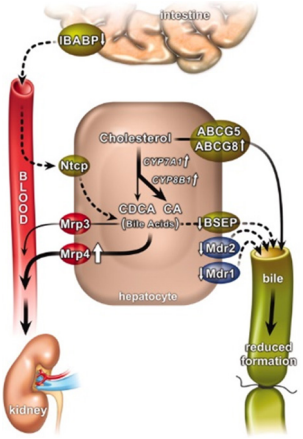
Toxicology Workflow

Contents



4.1 Introduction

Cholestasis is a condition where bile cannot flow from the liver to the duodenum. There are two basic of cholestasis. One is an obstructive type of cholestasis, where there is a mechanical blockage in the duct system that can occur from a gallstone or malignancy. The other, which includes metabolic types of cholestasis, involves disturbances in bile formation that can occur because of genetic defects or acquired as a side effect of many medications.

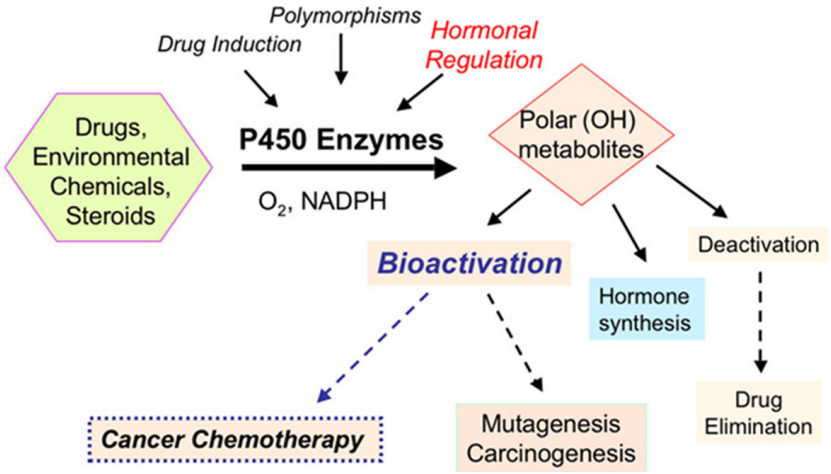


4.2 Cytochrome P450 Genes

Enzymes produced from cytochrome P450 genes are involved in the formation (synthesis) and breakdown (metabolism) of various molecules and chemicals within cells. Cytochrome P450 enzymes play a role in the synthesis of many molecules including steroid hormones, certain fats (cholesterol and other fatty acids), and acids used to digest fats (bile acids). Additional cytochrome P450 enzymes metabolize external substances, such as medications that are ingested, and internal substances, such as toxins that are formed within cells. There are approximately 60 CYP genes in humans.

Common variations (polymorphisms) in cytochrome P450 genes can affect the function of the enzymes. The effects of polymorphisms are most prominently seen in the breakdown of medications. Depending

Cytochrome P450 Enzymes and their Regulation

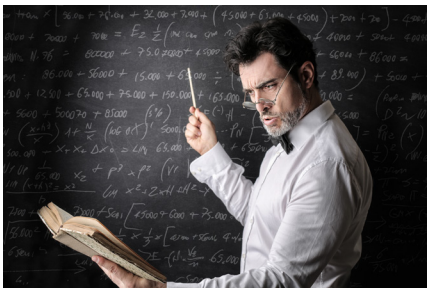


on the gene and the polymorphism, drugs can be metabolized quickly or slowly. If a cytochrome P450 enzyme metabolizes a drug slowly, the drug stays active longer and less is needed to get the desired effect. A drug that is quickly metabolized is broken down sooner, and a higher dose might be needed to be effective. Cytochrome P450 enzymes account for 70 percent to 80 percent of enzymes involved in drug metabolism.

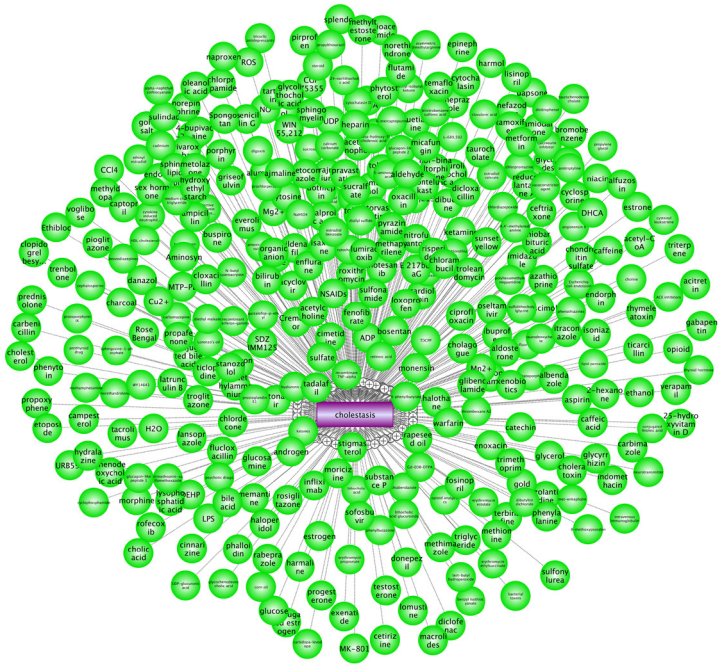
Common variations (polymorphisms) in cytochrome P450 genes can affect the function of the enzymes. **The effects of polymorphisms are most prominently seen in the breakdown of medications.** Depending on the gene and the polymorphism, drugs can be metabolized quickly or slowly. If a cytochrome P450 enzyme metabolizes a drug slowly, the drug stays active longer and less is needed to get the desired effect. A drug that is quickly metabolized is broken down sooner and a higher dose might be needed to be effective. **Cytochrome P450 enzymes account for 70 percent to 80 percent of enzymes involved in drug metabolism.**

Each cytochrome P450 gene is named with CYP, indicating that it is part of the cytochrome P450 gene family. The gene is also given a number associated with a specific group within the gene family, a letter representing the gene's subfamily, and a number assigned to the specific gene within the subfamily. For example, the cytochrome P450 gene that is in group 27, subfamily A, gene 1 is written as CYP27A1.

Diseases caused by mutations in cytochrome P450 genes typically involve the buildup of substances in the body that are harmful in large amounts or that prevent other necessary molecules from being produced. OK, enough chalk talk. You want to get to some hands-on learning!



4.3 Map Small Molecules Inducing Cholestatic Effects



First, let's map all small molecules reported to induce cholestatic effects: Add neighbors to cholestasis

- Upstream;
- Small Molecules;
- Regulation;
- Effect = positive

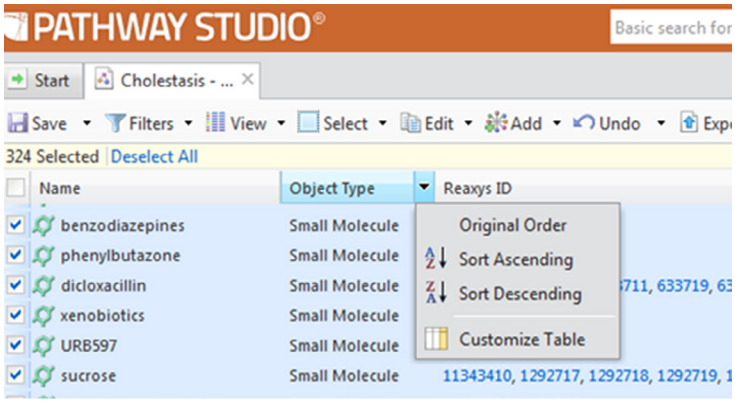
Save pathway (Hint: Give it a name that will remind you what is, such as "Cholestasis - all small molecules."). Entity # = 352

4.4 Filter Small Molecules for Drugs Only

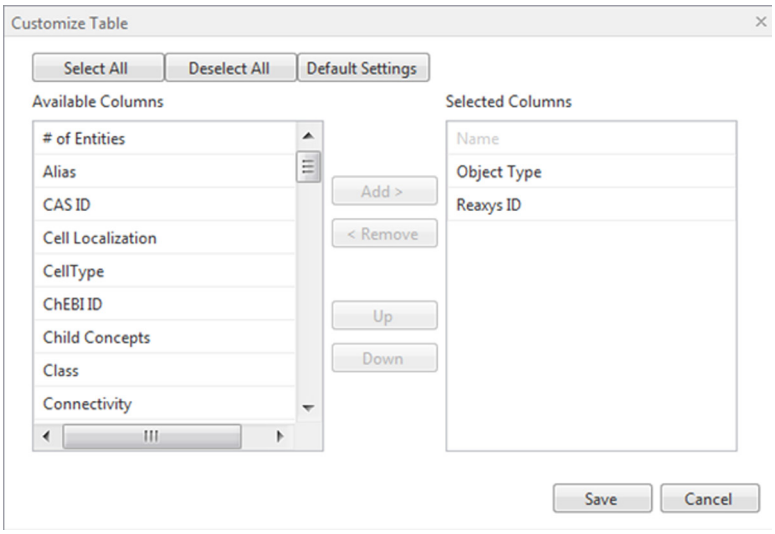
Next, you need to filter out just drugs from all small molecules, and this will require a quick excursion outside of Pathway Studio into Excel and back again.

First, go to the Entity Table View and, if it's not here already, add Reaxys ID to the columns

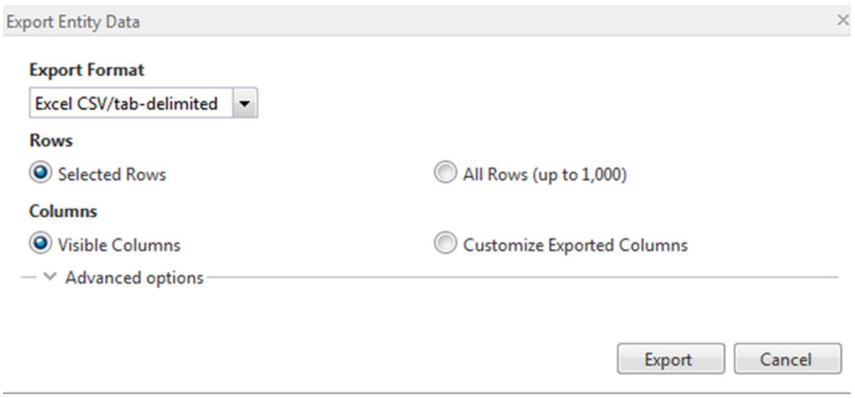
(Note: You can always add or remove columns from either the Entity Table View or the Relation Table View by clicking on any column header and choosing Customize Table from the dropdown menu.)



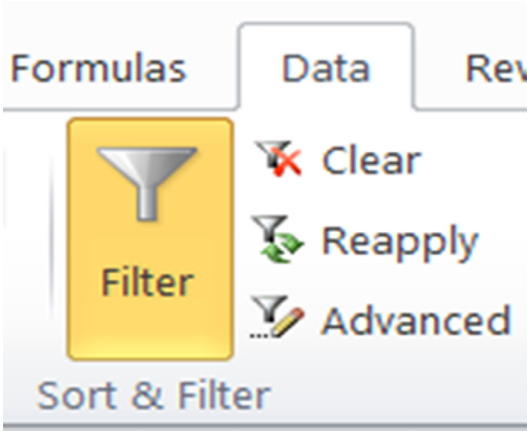
Select and add Reaxys ID (Note: Column options will change depending on the entities and relations in the table.)



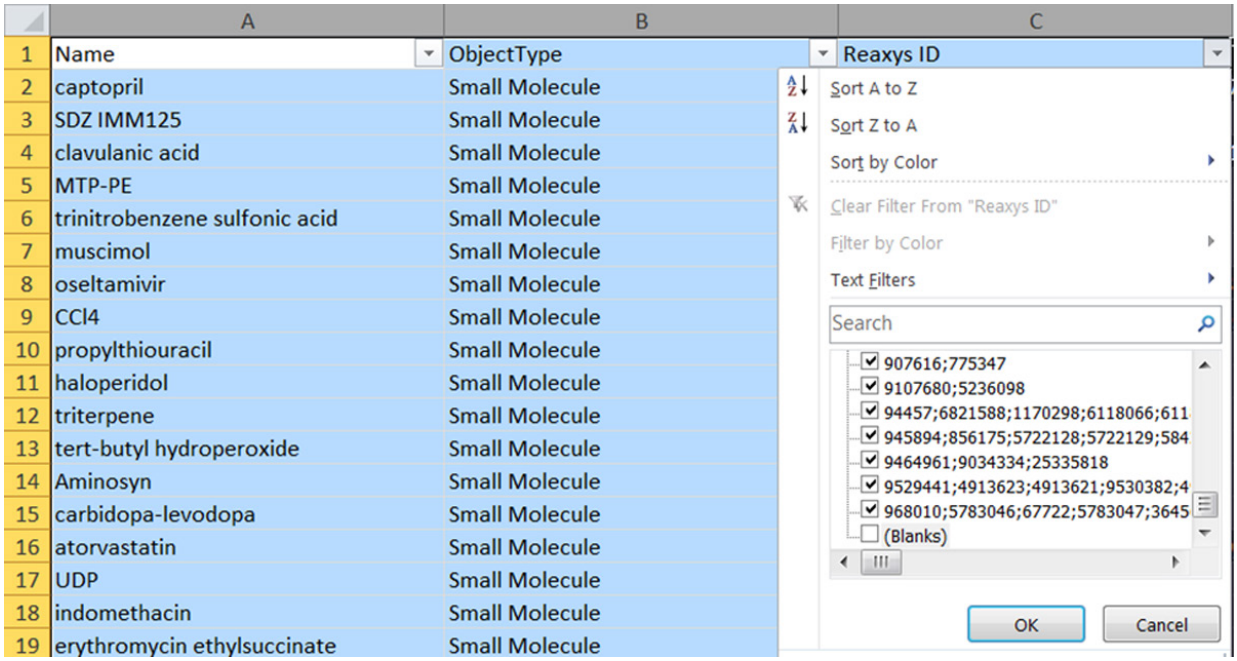
Export entity data (will automatically open in Excel).

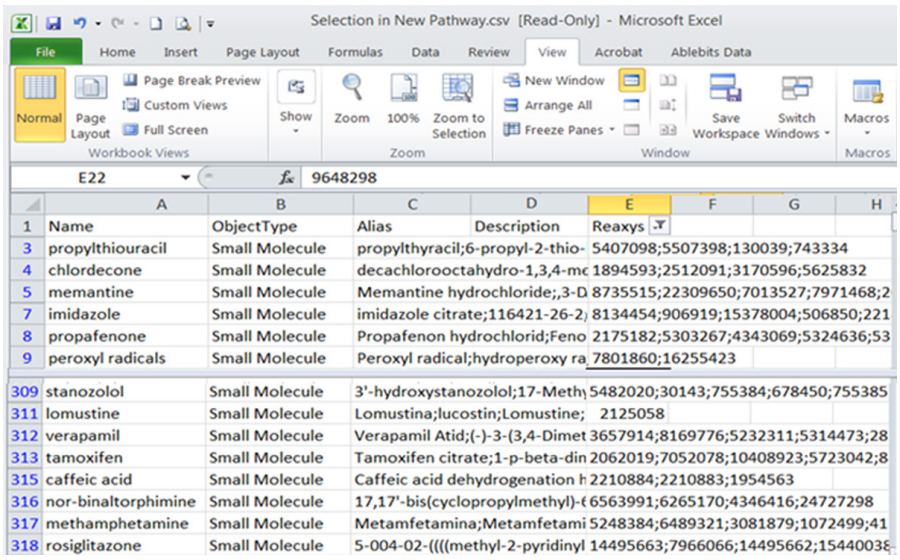


Select Reaxys ID column, and under Data tab -> select Filter.



Scroll down and uncheck "Blanks."





This will filter data for only those small molecules that have a Reaxys ID entry and thus will enrich those small molecules that are also recognized as drugs.

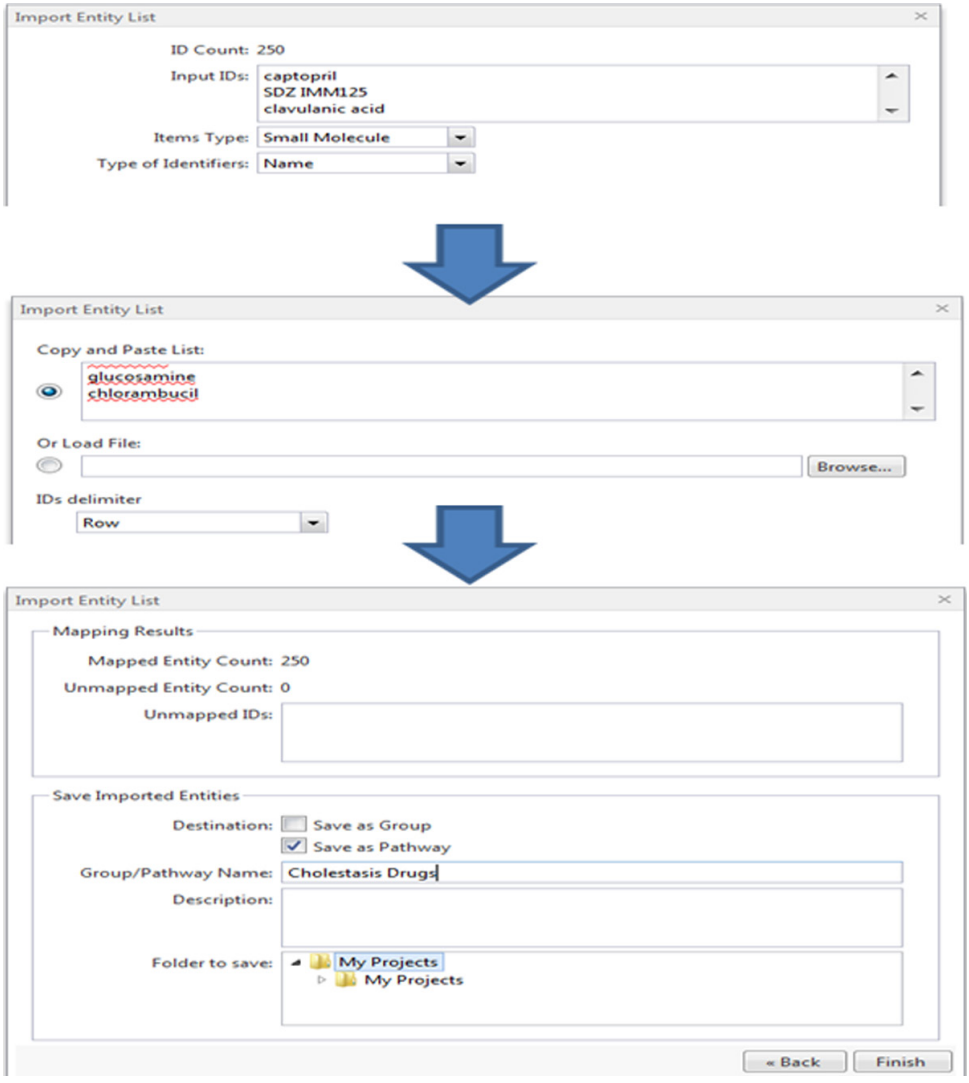
Copy (from Column A) all filtered drug names.

Now, you're going to go back into Pathway Studio and import these entities in as a group or pathway (Note: For this purpose, it doesn't matter which type.)

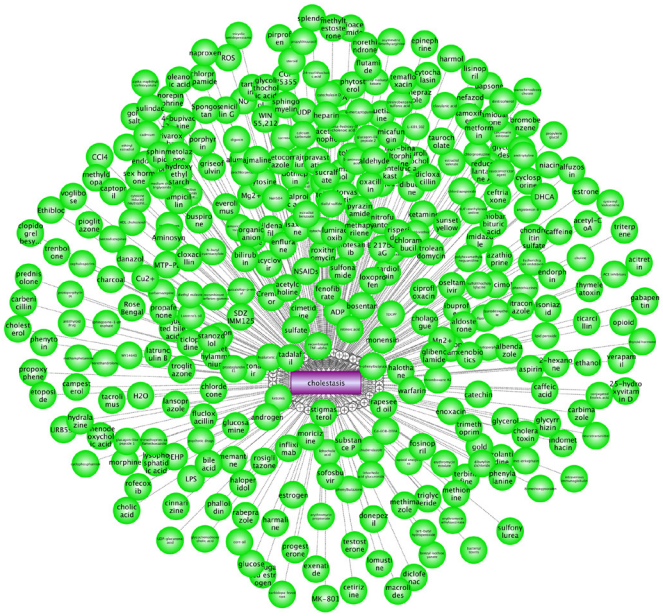
Now, you will quickly reimport the filtered list of Entity Names from Excel back into Pathway Studio by copy/paste using the Import Entity function:

Item type = Small Molecule"

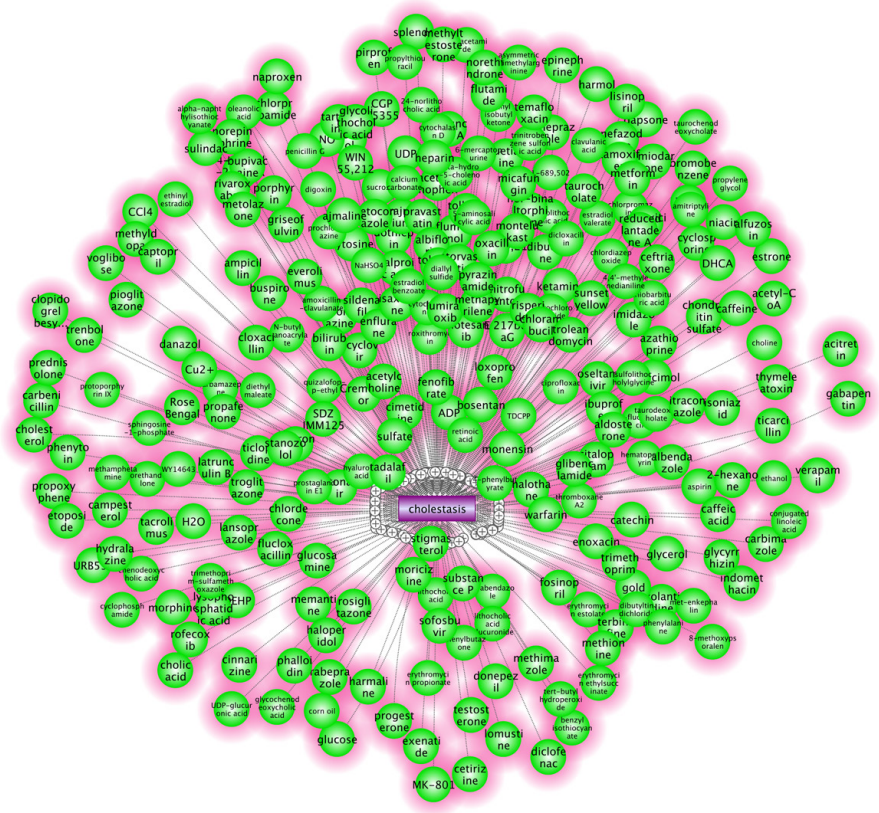
Give it a name!



Next, select all of the imported entities, copy, and return to the original "Cholestasis – all small molecules" pathway. Open in Graph View.



Select "clipboard content," and highlight in red.



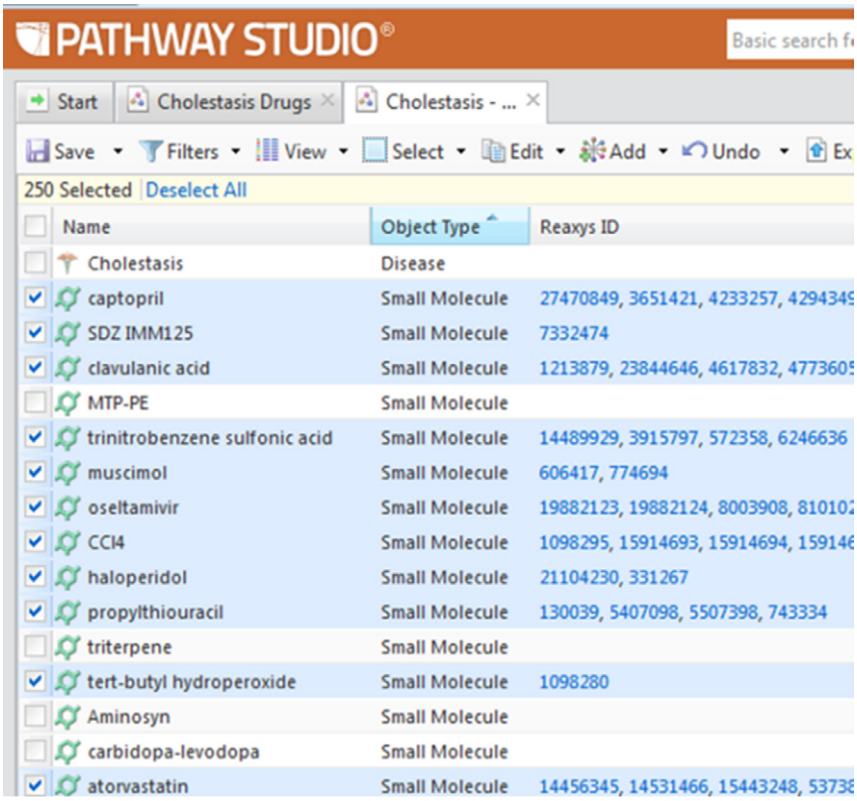
Now, remove all the non-drug small molecules.

Keeping the current selection, shift click on the cholestasis disease icon (Note: This will preserve all the remaining relations after the removal of unwanted entities in the next step.)

Select the Entity Table View (you'll notice some entries are highlighted, while others are not).

You'll also notice (if you're being particularly observant!) that all the highlighted entries have a Reaxys entry while the non-highlighted entries do not. Those are the ones we want to remove, but how?

...Get ready for it!

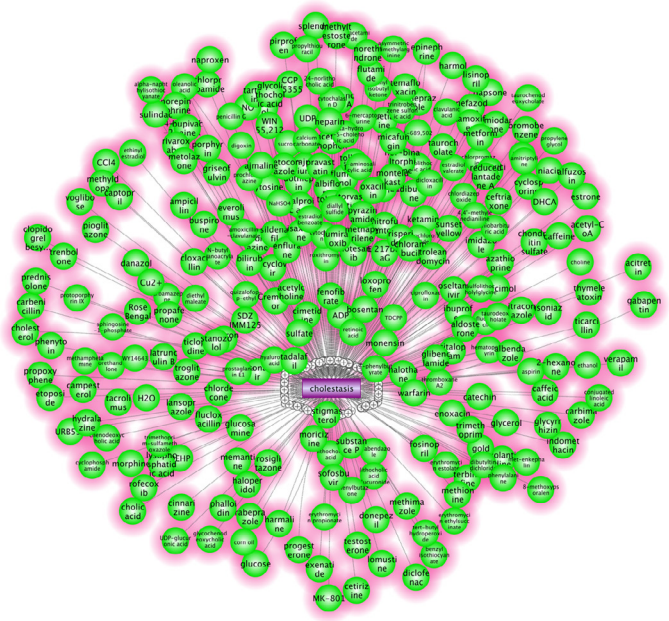


PATHWAY STUDIO TRICK #2

You're going to use something called "invert selection" found under the Select tab (just do it and see what happens!).

Did you see what just happened? Now, all the non-Reaxys entries are highlighted and all you have to do is...hit the Remove command (under the Edit tab) and voila - all gone!

View your results in the Graph View mode. See, now the only small molecules are the highlighted drugs.



Entity # = 272 +1 (drugs + disease)

Don't forget to save your work!

OK, time to get pumped up!

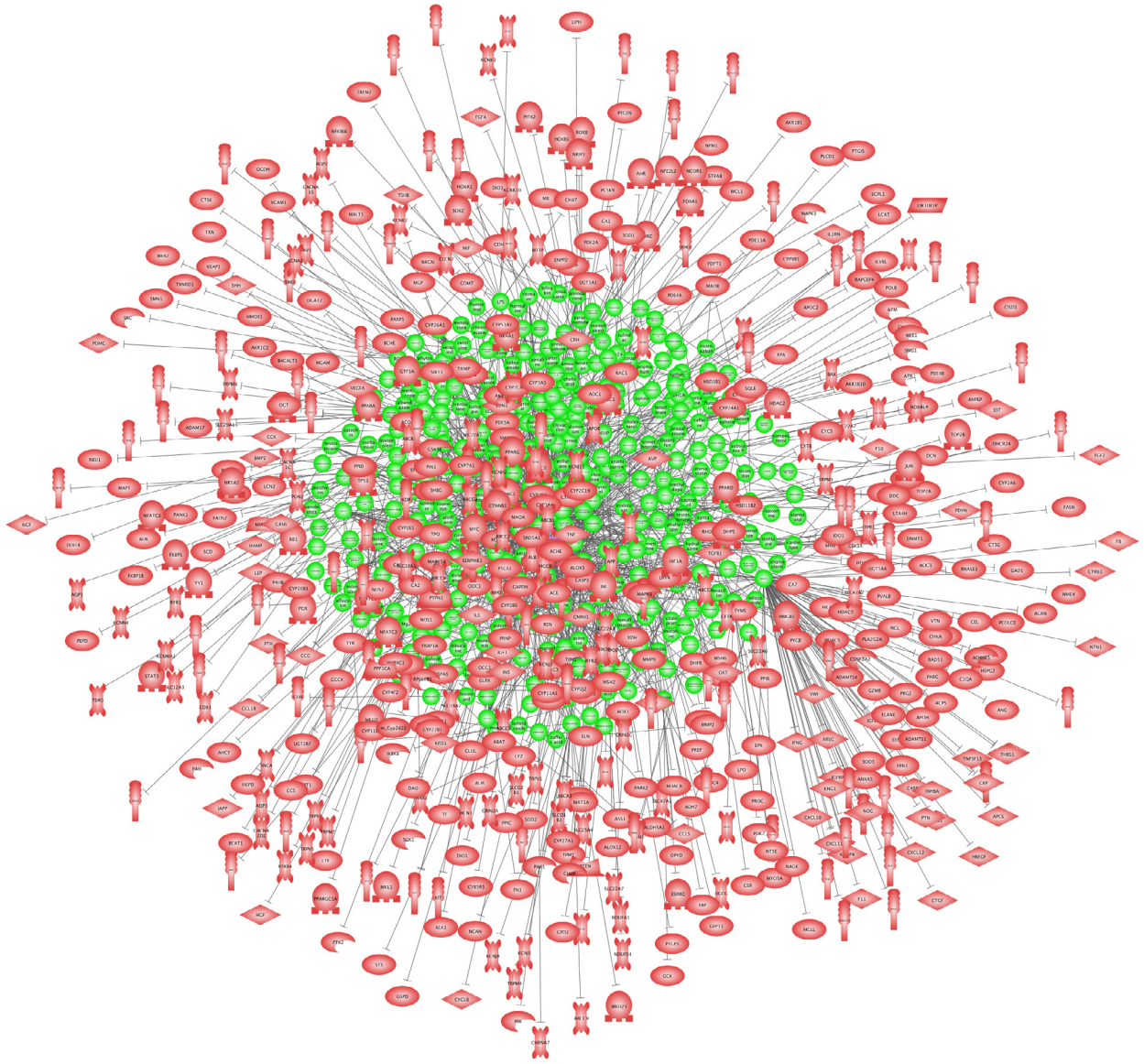
Are you ready for some serious networking?
Are you ready?

4.5 Find All Proteins Inhibited by Cholestasis-Related Drugs

You're now going to find all the proteins that are inhibited by these cholestasis related drugs.

Add neighbors to cholestasis related drugs:

1. Select -> All
2. Add -> Network Builder
3. Expand pathway -> Advanced Expand Pathway Tool -> Next
4. Direction = Downstream -> Next
5. Select Protein as Entity, Direct Regulation as Relation, "Effect" = 'Negative' -> Next
6. Build Network from Entities In…. Entire Database -> Next -> Finish
7. **View -> Graph View**



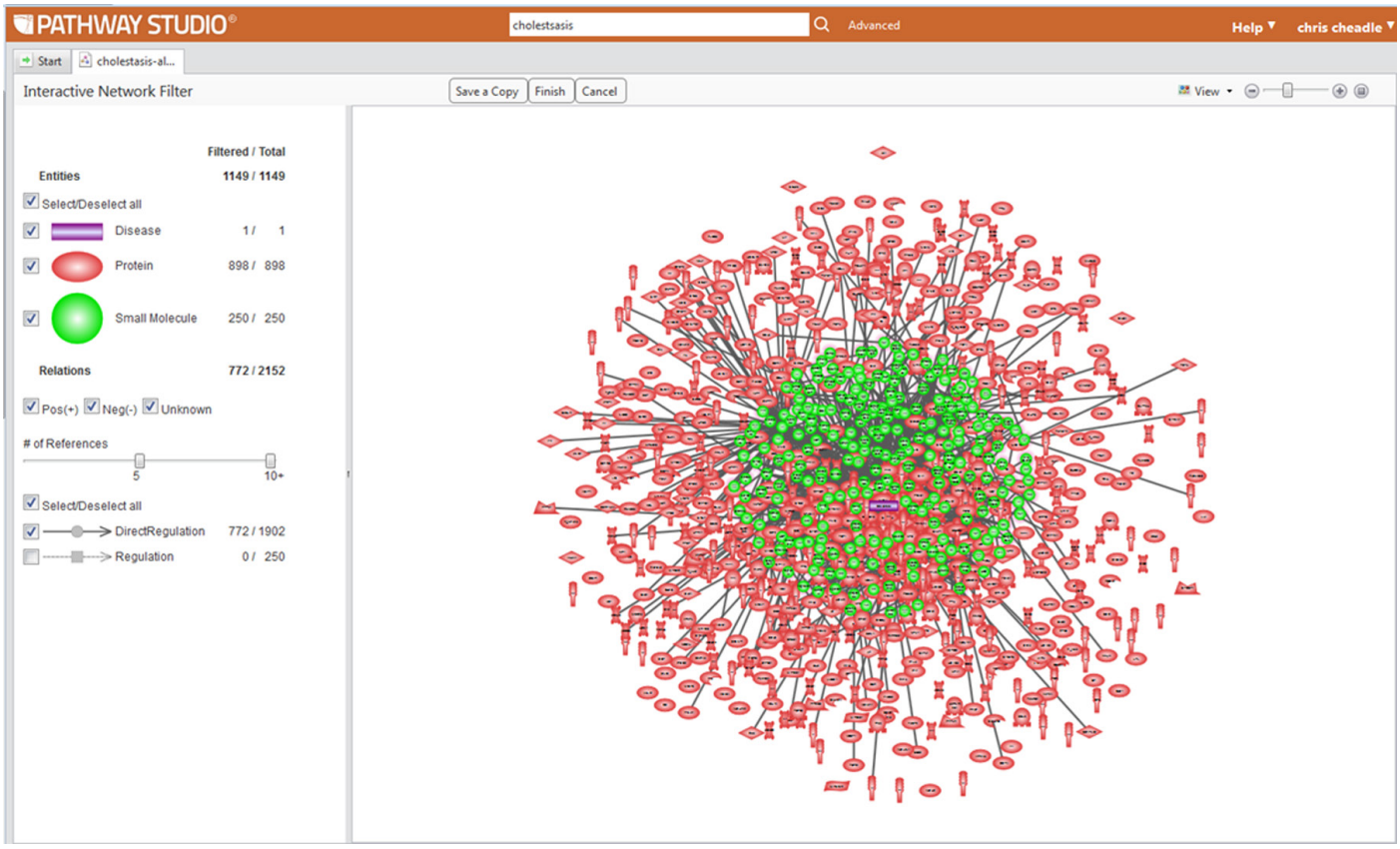
Entity # = 1029, Relation # = 2310

That's a lot of proteins! But then again, you started with a lot of drugs! Filter this down a little bit and GET IT UNDER CONTROL!

4.6 Using the Interactive Network Filter

Select "Basic" under the "Filters" tab. This will open up the "Interactive Network Filter." This is an extremely useful feature in Pathway Studio 11.2. You can filter by entities and relations, effect (positive, negative, or none), and reference number.

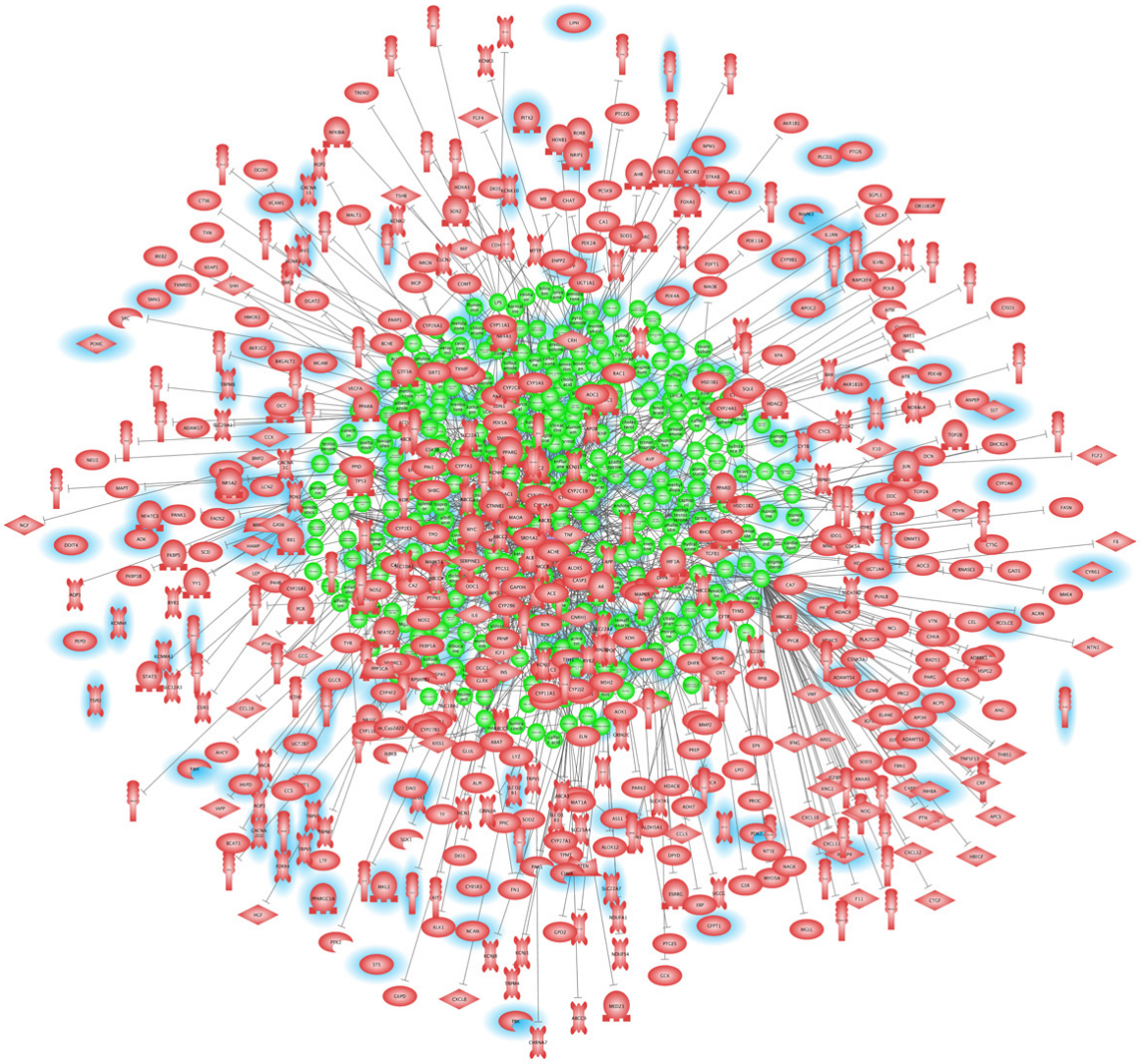
All your choices are reversible until you press "Finish." This allows you to rapidly inspect the network size and complexity based on your filtering choices.



Caution: Once you click on "Finish," you can't go back! So save your work before you filter. It will save you time and trouble if you need to backtrack!

Set the # of References = 7 (on the slider bar). This means each relation will have seven or more separate articles in support of the basic observation, and this will ensure that only the most well-studied drug/protein interactions will be considered for further analysis. This is fine for your purposes here.

Wait a second, it doesn't look like anything's changed! But, if you look a little closer, you see that there are now entities in the network that are unconnected, i.e. they no longer have relations between certain proteins and any of the drugs. Why is that? (Hint: Something to do with reference #.)



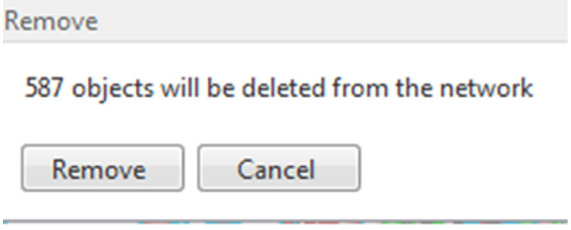
And that brings us to

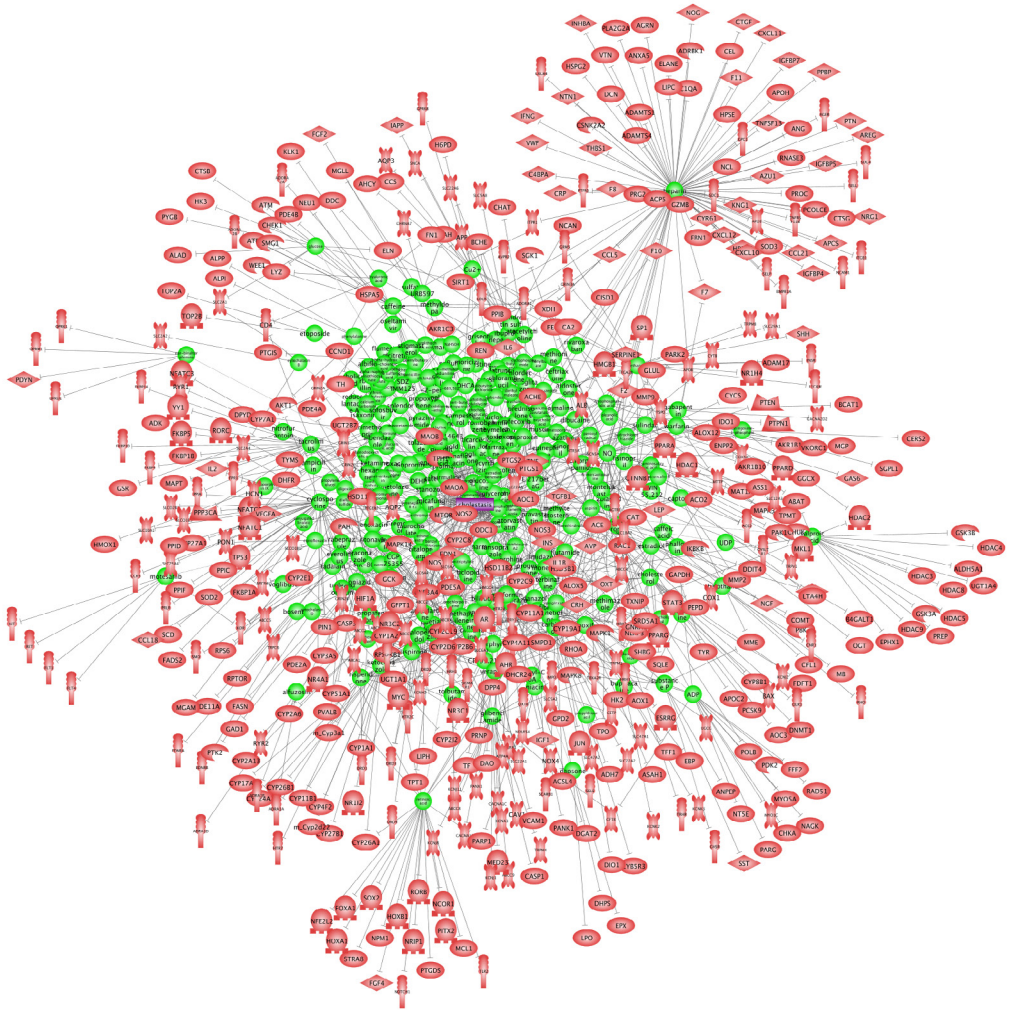
PATHWAY STUDIO TRICK #3

Watch this! Go to the “Select” tab and select “Unconnected Entities” then go to edit and click remove.

AND....

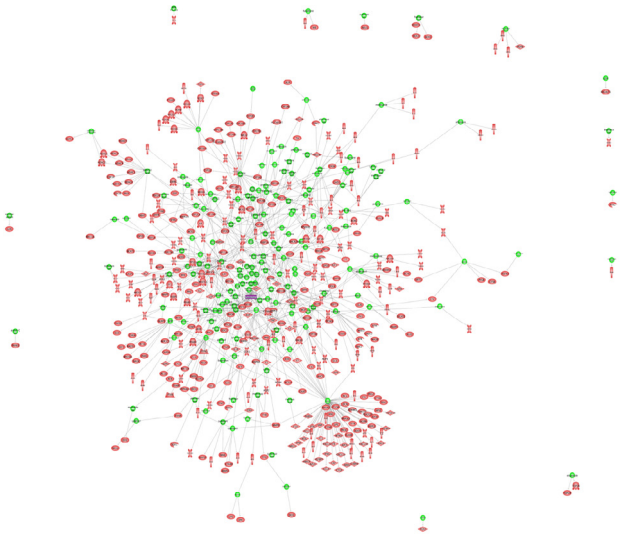
Voila ! --> a greatly reduced network.





Entity # = 606 Relation # = 884

Now...



Go to “Layout” and click “Direct Force.”

Same data, different look.

See if you can mouse over and identify drugs that affect the largest number of different proteins (Hint: Heparin is one.).

That's interesting, but what you'd really like to do is to identify those proteins that are the most frequently targeted by cholestasis-inducing drugs. In other words, instead of drugs connected to the largest number of different proteins (as seen in the previous slide), you'd like to identify individual proteins that are connected to the largest number of different drugs. They may be (by inference) the most common mediators of adverse cholestatic events.

How can you do this?

Well, it's going to take a little work and another trip outside Pathway Studio.

Are you ready? C'mon it's fun!

Yes, we are going to Excel (in every way).

4.7 Identify Proteins Most Frequently Targeted by Cholestasis-Inducing Drugs

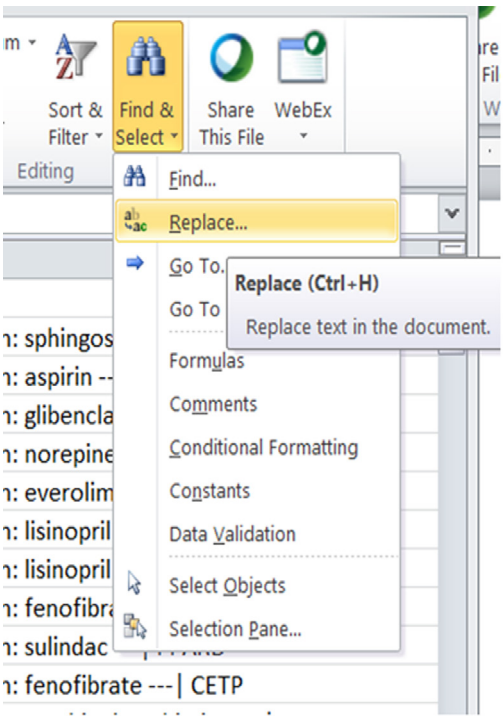
First, from "Relation Table View" in Excel, click on "Export Relation Data," open in Excel. You should see something like this. Remove all but the first column and follow along carefully!

A	B	C	D	E	F	G
RelationSymbolicName	ObjectType	RelationNumber(	RelationConfi	BiomarkerType	ChangeType	Source
negative DirectRegulation: sphingosine-1-phosphate ---| HDAC1	DirectRegulation	9	3			
negative DirectRegulation: aspirin ---| TBXA2R	DirectRegulation	8	3			
negative DirectRegulation: glibenclamide ---| ABCC9	DirectRegulation	6	3			
negative DirectRegulation: norepinephrine ---| HTR2A	DirectRegulation	10	3			
negative DirectRegulation: everolimus ---| EGFR	DirectRegulation	7	3			
negative DirectRegulation: lisinopril ---| MMP9	DirectRegulation	6	3			
negative DirectRegulation: lisinopril ---| ACE	DirectRegulation	19	3			
negative DirectRegulation: fenofibrate ---| SCARB1	DirectRegulation	5	3			
negative DirectRegulation: sulindac ---| PPARD	DirectRegulation	8	3			
negative DirectRegulation: fenofibrate ---| CETP	DirectRegulation	9	3			
negative DirectRegulation: nor-binaltorphimine ---| OPRK1	DirectRegulation	599	3			
negative DirectRegulation: haloperidol ---| SIGMAR1	DirectRegulation	62	3			

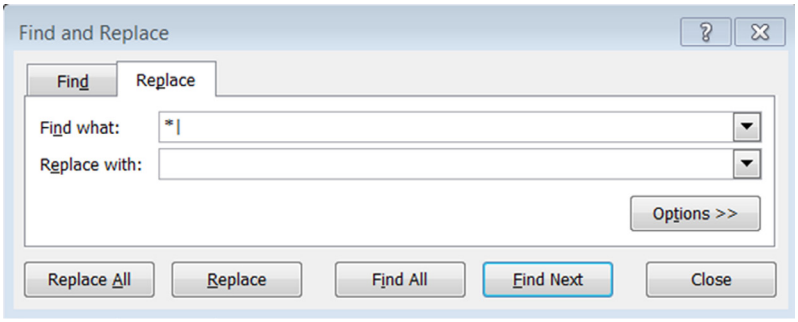
Copy and Paste the first column into two additional blank columns to the right.

A	B	C
RelationSymbolicName	RelationSymbolicName	RelationSymbolicName
negative DirectRegulation: sphingosine-1-phosphate ---| HDAC1	negative DirectRegulation: sphingosine-1-phosphate ---| HDAC1	negative DirectRegulation: sphingosine-1-phosphate ---| HDAC1
negative DirectRegulation: aspirin ---| TBXA2R	negative DirectRegulation: aspirin ---| TBXA2R	negative DirectRegulation: aspirin ---| TBXA2R
negative DirectRegulation: glibenclamide ---| ABCC9	negative DirectRegulation: glibenclamide ---| ABCC9	negative DirectRegulation: glibenclamide ---| ABCC9
negative DirectRegulation: norepinephrine ---| HTR2A	negative DirectRegulation: norepinephrine ---| HTR2A	negative DirectRegulation: norepinephrine ---| HTR2A
negative DirectRegulation: everolimus ---| EGFR	negative DirectRegulation: everolimus ---| EGFR	negative DirectRegulation: everolimus ---| EGFR
negative DirectRegulation: lisinopril ---| MMP9	negative DirectRegulation: lisinopril ---| MMP9	negative DirectRegulation: lisinopril ---| MMP9
negative DirectRegulation: lisinopril ---| ACE	negative DirectRegulation: lisinopril ---| ACE	negative DirectRegulation: lisinopril ---| ACE
negative DirectRegulation: fenofibrate ---| SCARB1	negative DirectRegulation: fenofibrate ---| SCARB1	negative DirectRegulation: fenofibrate ---| SCARB1
negative DirectRegulation: sulindac ---| PPARD	negative DirectRegulation: sulindac ---| PPARD	negative DirectRegulation: sulindac ---| PPARD
negative DirectRegulation: fenofibrate ---| CETP	negative DirectRegulation: fenofibrate ---| CETP	negative DirectRegulation: fenofibrate ---| CETP
negative DirectRegulation: nor-binaltorphimine ---| OPRK1	negative DirectRegulation: nor-binaltorphimine ---| OPRK1	negative DirectRegulation: nor-binaltorphimine ---| OPRK1
negative DirectRegulation: haloperidol ---| SIGMAR1	negative DirectRegulation: haloperidol ---| SIGMAR1	negative DirectRegulation: haloperidol ---| SIGMAR1
negative DirectRegulation: glibenclamide ---| NLRP3	negative DirectRegulation: glibenclamide ---| NLRP3	negative DirectRegulation: glibenclamide ---| NLRP3

Select column B and do a "replace all."



Type in an asterisk (in Excel, * is a wild card), a pipe (|), and a space (don't forget the space!), and then click "Replace All."

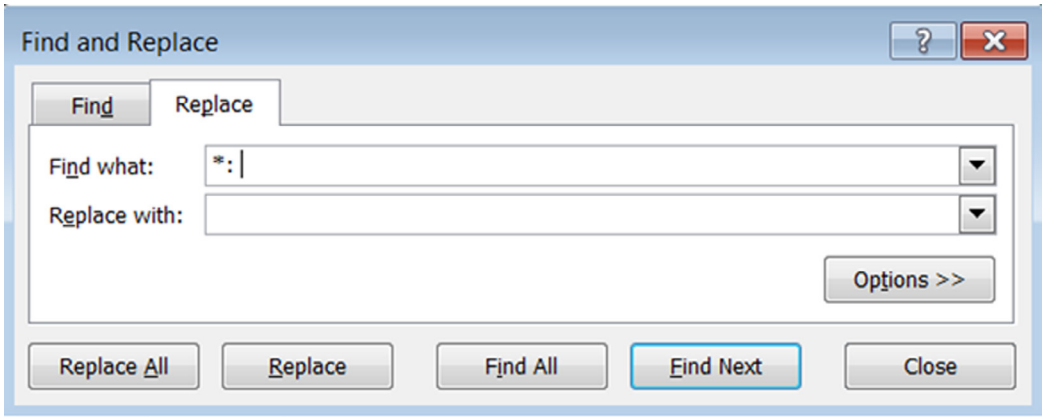


Column B should now look like this:

A	B	C
RelationSymbolicName	RelationSymbolicName	RelationSymbolicName
negative DirectRegulation: sphingosine-1-phosphate ---| HDAC1	HDAC1	negative DirectRegulation: sphingosine-1-phosphate ---| HDAC1
negative DirectRegulation: aspirin ---| TBXA2R	TBXA2R	negative DirectRegulation: aspirin ---| TBXA2R
negative DirectRegulation: glibenclamide ---| ABCC9	ABCC9	negative DirectRegulation: glibenclamide ---| ABCC9
negative DirectRegulation: norepinephrine ---| HTR2A	HTR2A	negative DirectRegulation: norepinephrine ---| HTR2A
negative DirectRegulation: everolimus ---| EGFR	EGFR	negative DirectRegulation: everolimus ---| EGFR
negative DirectRegulation: lisinopril ---| MMP9	MMP9	negative DirectRegulation: lisinopril ---| MMP9
negative DirectRegulation: lisinopril ---| ACE	ACE	negative DirectRegulation: lisinopril ---| ACE
negative DirectRegulation: fenofibrate ---| SCARB1	SCARB1	negative DirectRegulation: fenofibrate ---| SCARB1
negative DirectRegulation: sulindac ---| PPARD	PPARD	negative DirectRegulation: sulindac ---| PPARD
negative DirectRegulation: fenofibrate ---| CETP	CETP	negative DirectRegulation: fenofibrate ---| CETP
negative DirectRegulation: nor-binaltorphimine ---| OPRK1	OPRK1	negative DirectRegulation: nor-binaltorphimine ---| OPRK1
negative DirectRegulation: haloperidol ---| SIGMAR1	SIGMAR1	negative DirectRegulation: haloperidol ---| SIGMAR1
negative DirectRegulation: glibenclamide ---| NLRP3	NLRP3	negative DirectRegulation: glibenclamide ---| NLRP3
negative DirectRegulation: aspirin ---| ITGA2B	ITGA2B	negative DirectRegulation: aspirin ---| ITGA2B
negative DirectRegulation: ticlopidine ---| P2RY12	P2RY12	negative DirectRegulation: ticlopidine ---| P2RY12
negative DirectRegulation: isoniazid ---| CYP3A4	CYP3A4	negative DirectRegulation: isoniazid ---| CYP3A4

Select column C and do a "replace all."

Type an asterisk, a colon (:), and a space (don't forget the space). then click "Replace All."



This will get you to here:

C
RelationSymbolicName
sphingosine-1-phosphate ---| HDAC1
aspirin ---| TBXA2R
glibenclamide ---| ABCC9
norepinephrine ---| HTR2A
everolimus ---| EGFR
lisinopril ---| MMP9
lisinopril ---| ACE
fenofibrate ---| SCARB1
sulindac ---| PPARD

And again, one more time, type in two dashes, and an asterisk, then click "Replace All."

That will get you to here:

B	C
RelationSymbolicName	RelationSymbolicName
HDAC1	sphingosine-1-phosphate
TBXA2R	aspirin
ABCC9	glibenclamide
HTR2A	norepinephrine
EGFR	everolimus
MMP9	lisinopril
ACE	lisinopril
SCARB1	fenofibrate
PPARD	sulindac
CETP	fenofibrate

So what have you accomplished by this maneuver? Well, you now have extracted both protein and drug information from each of the imported relations (as is still reflected in column A).

Feel free to re-label the headers for columns B & C, proteins and drugs, respectively.

By the way, why do you think you used two dashes instead of one in the last "replace all"? That's right (you are so smart)! You had to use two because many of the drug names have an embedded dash, and that would have messed up everything!

B	C
Proteins	**Drugs**
ABAT	valproic acid
ABAT	gabapentin
ABCA1	glibenclamide
ABCA1	cyclosporine
ABCB1	chlorpromazine
ABCB1	amiodarone
ABCB1	itraconazole
ABCB1	atorvastatin
ABCB1	tamoxifen
ABCB1	fenofibrate
ABCB1	amitriptyline
ABCB1	verapamil

Now, you can sort alphabetically by protein name.

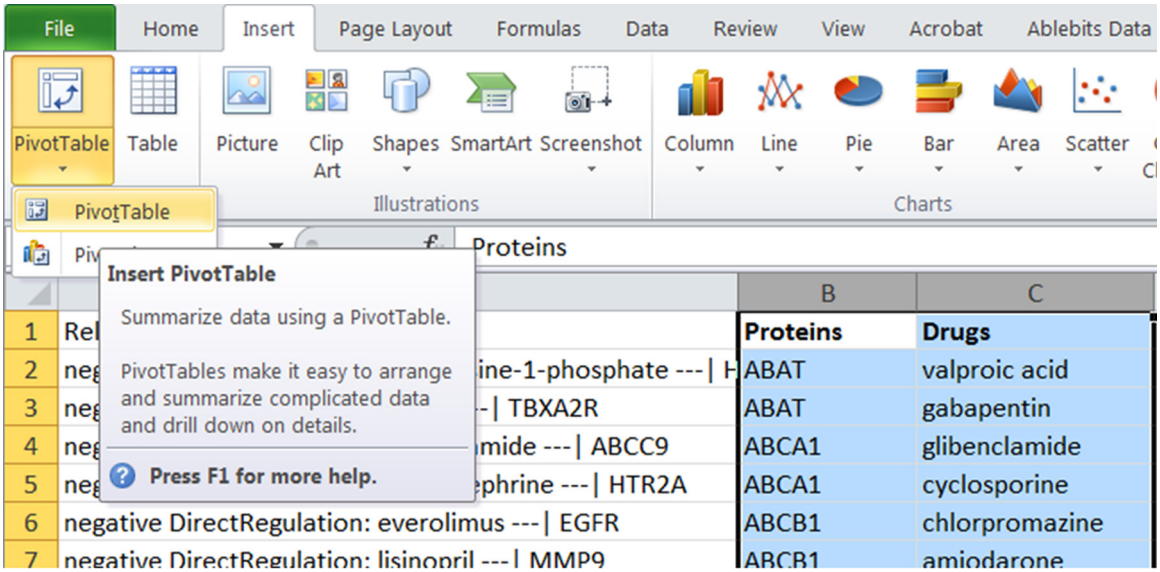
You can see right away that many of the proteins map to multiple drugs, and that is precisely the information we set out to capture. But, how do you quantify them? You don't want to count all 722-832 entities!

You don't have to. Excel will do it for you!

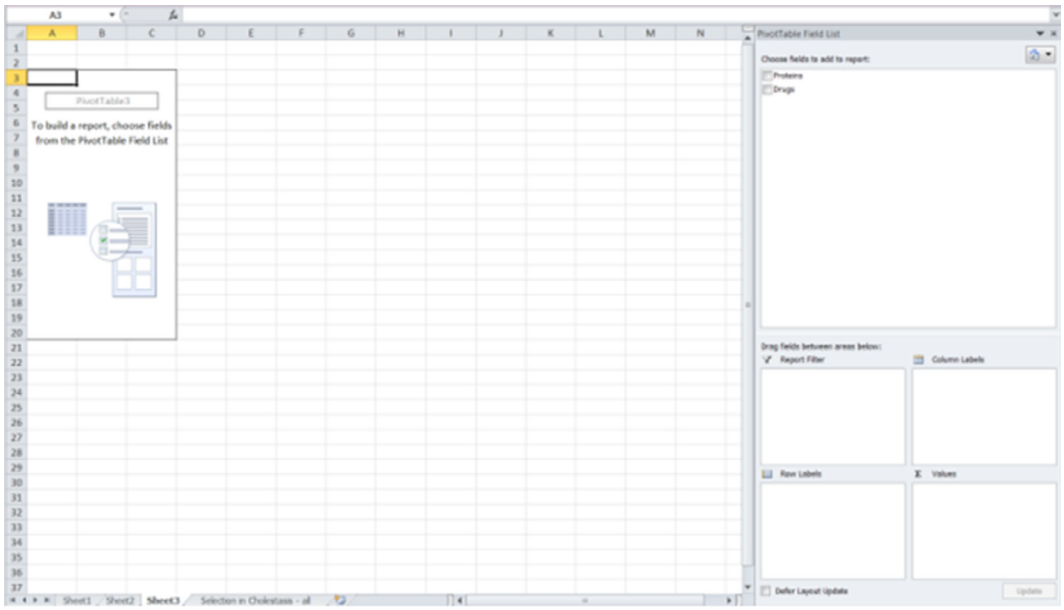
There are several different ways to accomplish your goal in Excel, one of which is to use the subtotal function. But, an even faster (and more efficient way) is to use something called a "pivot table," which is designed in Excel to answer questions just like this.

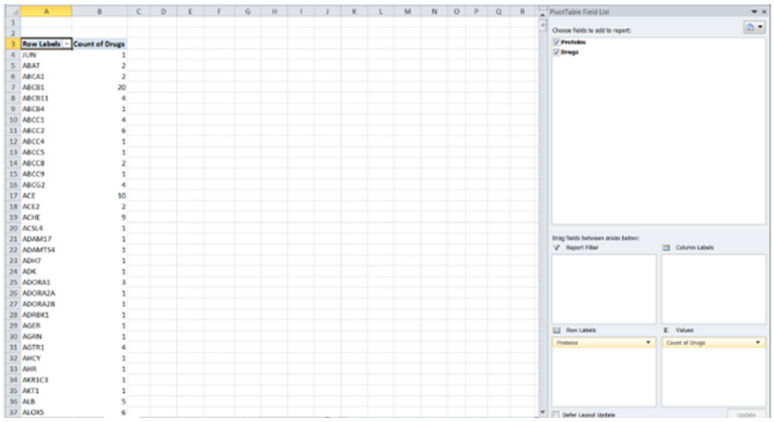
IT'S A
GOOD
TIME TO
START
SOMETHING
NEW.

So, here you go! First, select columns B and C, and then go to Insert, Pivot Table, and click OK when the window opens.



This should open up a new worksheet in Excel that looks something like this:





Check both proteins and drugs in "PivotTable Field List." Drag Proteins to "Row Labels" and Drugs to "Values" (Hint: Make sure values are set to Count, which is controlled in dropdown box by "Value Field Settings.").

Copy the data from columns A and B into a new worksheet and sort by Column B (largest to smallest).

Note the presence of five CYP proteins in the top 10 genes/proteins inhibited by the largest number of cholestasis-inducing drugs. Take a closer look at the top-ranked CYP3A4 gene/protein.

	A	B
1	Protein Name	Drug #
2	PTGS2	26
3	CYP3A4	23
4	ABCB1	21
5	TNF	17
6	ACE	11
7	CYP19A1	11
8	CYP2D6	11
9	ACHE	10
10	CYP2C9	10
11	DRD2	10
12	CYP2C19	9

LATEST NEWS & CME

CYP3A4 therapeutic implication for cholestasis

1 Jan 2014

CYP3A4 activity may be useful in treating cholestasis, this review stated. CYP3A4 is a major cytochrome P450. It catalyses a broad range of substrates including xenobiotics and endogenous compounds. Cholestasis is a condition where there is an excessively high concentration of bile which is very toxic. CYP3A4 detoxifies bile acids, a function that could be used for treating cholestasis. CYP3A4 is responsively up regulated in cholestasis as a defence mechanism. However, the regulation of CYP3A4 is complicated by estrogen, and its activity is inhibited by the detergent effect of the accumulated bile. It was concluded from the study that reducing the factors that inhibit CYP3A4 and employing CYP3A4 activators to maximize its activity for detoxification of bile acids could be an effective approach for the treatment of cholestasis.

Chen, et al. The role of CYP3A4 in the biotransformation of bile acids and therapeutic implication for cholestasis. Ann Transl Med. 2014 Jan;2(1):7. [Link]

The role of CYP3A4 in the biotransformation of bile acids and therapeutic implication for cholestasis.

Abstract

CYP3A4 is a major cytochrome P450. It catalyzes a broad range of substrates including xenobiotics such as clinically used drugs and endogenous compounds, such as bile acids. Its function to detoxify bile acids could be used for treating cholestasis, which is a condition characterized by accumulation of bile acids. Although bile acids have important physiological functions, they are very toxic when their concentrations are excessively high. The accumulated bile acids in cholestasis can cause liver and other tissue injuries. Thus, control of the concentrations of bile acids is critical for treatment of cholestasis. CYP3A4 is responsively unregulated in cholestasis mediated by the nuclear receptors farnesol X receptor (FXR) and pregnane X receptor (PXR) as a defense mechanism. However, the regulation of CYP3A4 is complicated by estrogen, which is increased in cholestasis and down regulates CYP3A4 expression. The activity of CYP3A4 is also inhibited by accumulated bile acids due to their property of detergent effect. In some cholestasis cases, genetic polymorphisms of the CYP3A4 and PXR genes may interfere with the adaptive response. Further stimulation of CYP3A4 activity in cholestasis could be an effective approach for treatment of the disease. In this review, we summarize recent progress about the roles of CYP3A4 in the metabolism of bile acids, its regulation and possible implication in the treatment of cholestasis.

Many thanks to Anton Yuryev, Pat Morgan, and Nikolai Daraselia whose previous work with modeling Drug-Induced Cholestasis using Pathway Studio inspired this workflow.

- Without prior knowledge you've identified a promising new target (CYP3A4) for treating cholestasis, a very common, induced, adverse drug effect.
- You've learned a lot of Pathway Studio functionalities along the way including :
    - Systematic network expansion—first drugs, then proteins.
    - Filtering using the "Interactive Network Filter" for surfacing the most important information.
    - Some PS tricks like "Invert Selection," and removing "Unconnected Entities."
- And, finally, you've learned some useful things in Excel, such as sorting, filtering, replacing, and don't forget the PivotTable (very useful)!

And that's enough for one day!

Study Questions 4

1. What's the number of small molecules that induce cirrhosis?

2. Identify the top ten small molecules (by # of references) that induce cirrhosis.

3. Add Reaxys ID to all small molecules from Question 1, not just those sorted by number of references (Hint: Customize Table). Select and remove those that do not have a Reaxys ID. Give the number of small molecules with and without the Reaxys ID.

4. Identify the number of proteins inhibited (negatively regulated) by the top ten small molecules from Question 2.

5. Identify proteins that inhibit cirrhosis.

6. What is the number of overlapping proteins from Question 3 and Question 4? List them here.

For Protein Modification(s), How Do I...

Exercise 4.1: Find protein(s) that acetylate/deacetylatea protein?

Identifies proteins involved in acetylation/deacetylation or target protein(s)

Step 1: Create New Pathway or within Pathway, select a protein(s)

Step 2: Select Add- > Network Builder

Step 3: Select Expand Pathway, Advanced Expand Pathway Tool

Step 4: Select Direction: Upstream

Step 5: Entities: Select "Protein" Relations: Select "ProtModification"

Add condition: Mechanism "is equal to" acetylation or deacetylation

Continue to Network Builder Step 5 and click Launch Interactive Network Builder or Finish

Exercise 4.2: Find protein(s) that cleave a protein?

Identifies proteins involved in the proteolytic cleavage of target protein(s).

Step 1: Create New Pathway or within Pathway, select a protein(s)

Step 2: Select Add- > Network Builder

Step 3: Select Expand Pathway, Advanced Expand Pathway Tool

Step 4: Select Direction: Upstream

Step 5: Entities: "Protein" Relations: Select "ProtModification"

Add condition: mechanism "is equal to" cleavage

Continue to Network Builder Step 5 and click Launch Interactive Network Builder or Finish

Exercise 4.3: Find proteins(s) that methylate/demethylate a protein?

Identifies proteins involved in the methylation/demethylation of target protein(s).

Step 1: Create New Pathway or within Pathway, select a protein(s)

Step 2: Select Add- > Network Builder

Step 3: Select Expand Pathway, Advanced Expand Pathway Tool

Step 4: Select Direction: Upstream

Step 5: Entities: Select "Protein" Relations: Select "ProtModification"

Add condition: mechanism "is equal to" methylation/demethylation

Continue to Network Builder Step 5 and click Launch Interactive Network Builder or Finish

Exercise 4.4: Find protein(s) that phosphorylate/dephosphorylate a protein?

Identifies protein(s) involved in the phosphorylation/dephosphorylation of target protein(s).

Step 1: Create New Pathway or within Pathway, select a protein(s)

Step 2: Select Add- > Network Builder

Step 3: Select Expand Pathway, Advanced Expand Pathway Tool

Step 4: Select Direction: Upstream

Step 5: Entities: Select "Protein" Relations: Select "ProtModification"

Add condition: mechanism "is equal to" phosphorylation/dephosphorylation

Continue to Network Builder Step 5 and click Launch Interactive Network Builder or Finish

Exercise 4.5: Find protein(s) that ubiquitinate a protein?

Identifies protein(s) involved in the ubiquitination of target protein(s).

Step 1: Create New Pathway or within Pathway, select a protein(s)

Step 2: Select Add- > Network Builder

Step 3: Select Expand Pathway, Advanced Expand Pathway Tool

Step 4: Select Direction: Upstream

Step 5: Entities: Select “Protein” Relations: Select “ProtModification”

Add condition: mechanism “is equal to” ubiquitination

Continue to Network Builder Step 5 and click Launch Interactive Network Builder or Finish

Module 5

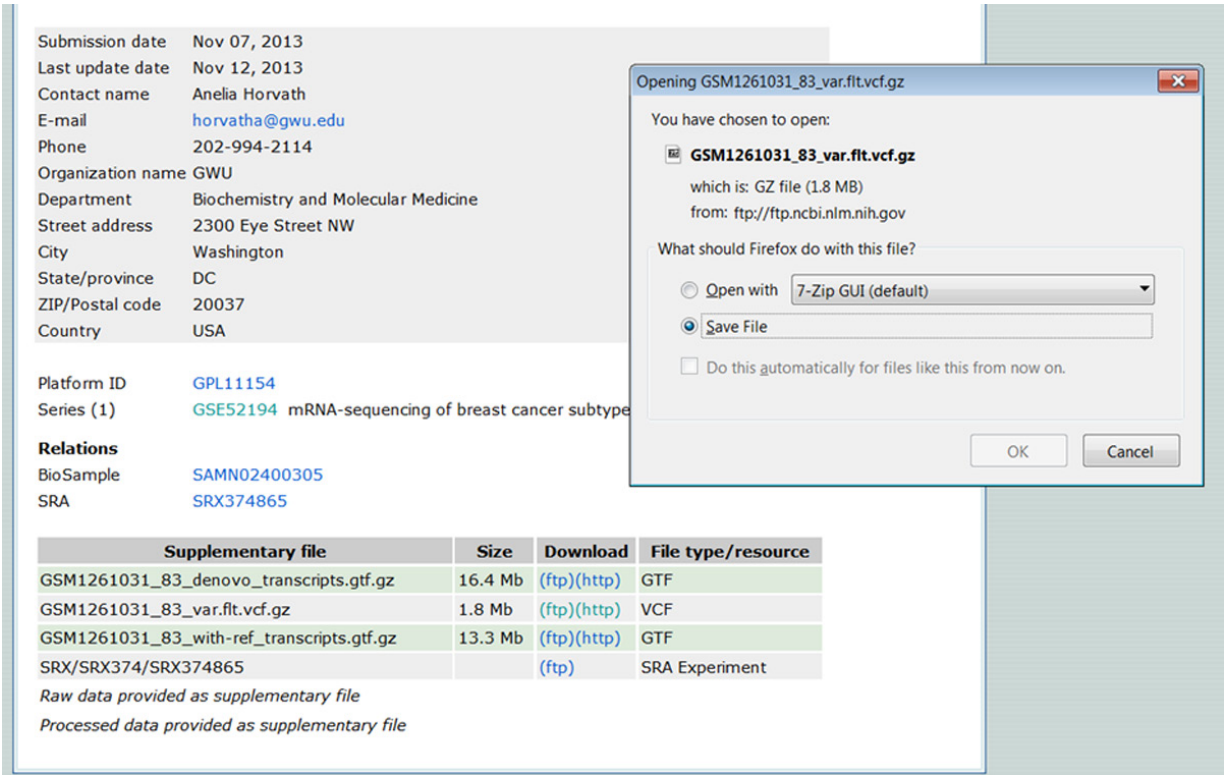
Variant Analysis II

Contents

Variant Analysis of an Individual Genome

5.1 Import Variant Data (.vcf files)

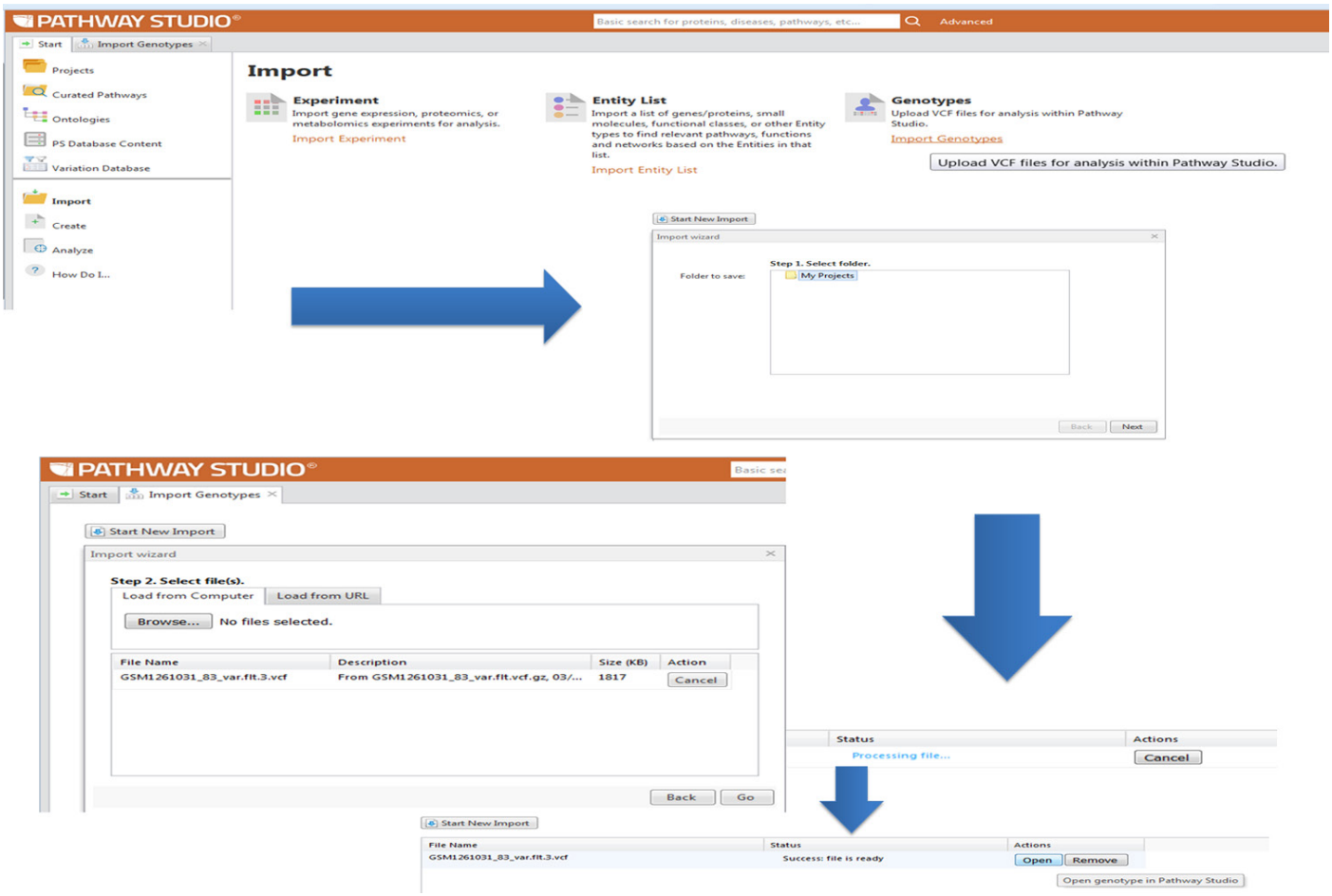
The first step in analysis of genomic sequence data is to import the data into Pathway Studio. The format required for both genomic or exome data is .vcf files. (Note: The time it takes to upload a file is dependent on the file size and number of files.)



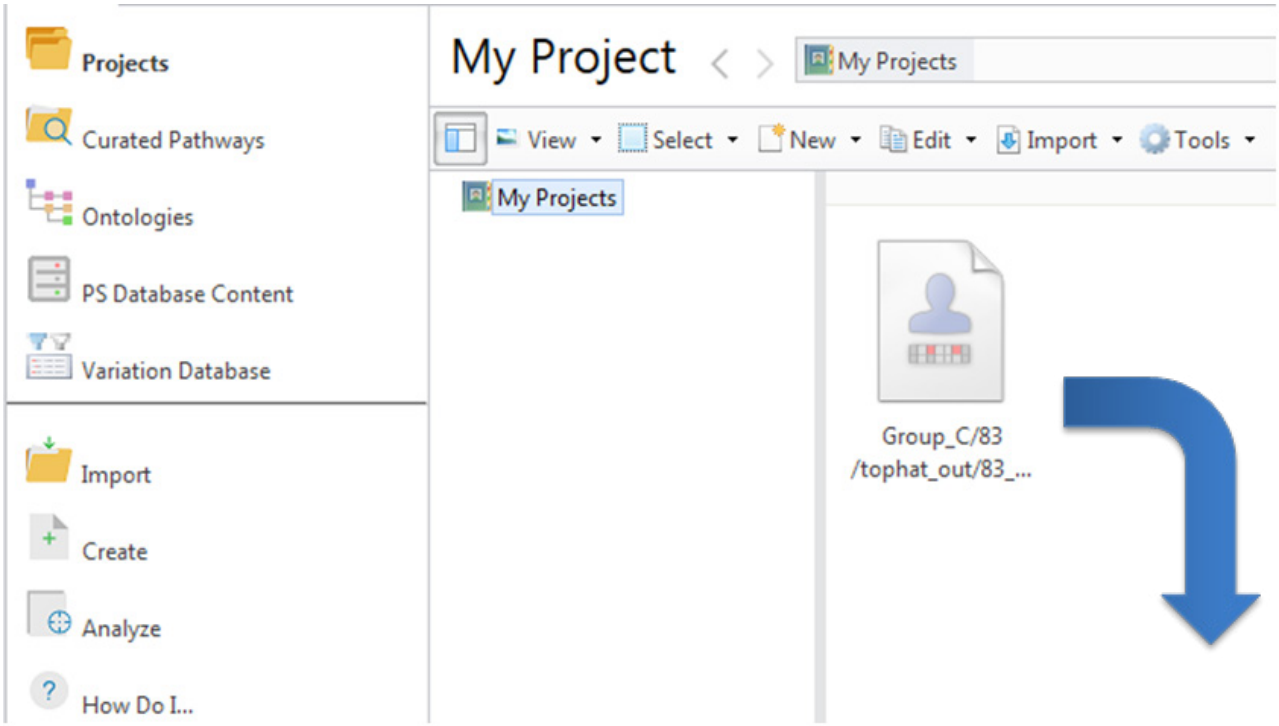
Start by analyzing a single exome file from a HER2 positive breast tumor biopsy sample. The file: GSM1261031_84_var.flt.vcf.gz can be found here: GSM1261031 or from Dropbox (data for Module 5)

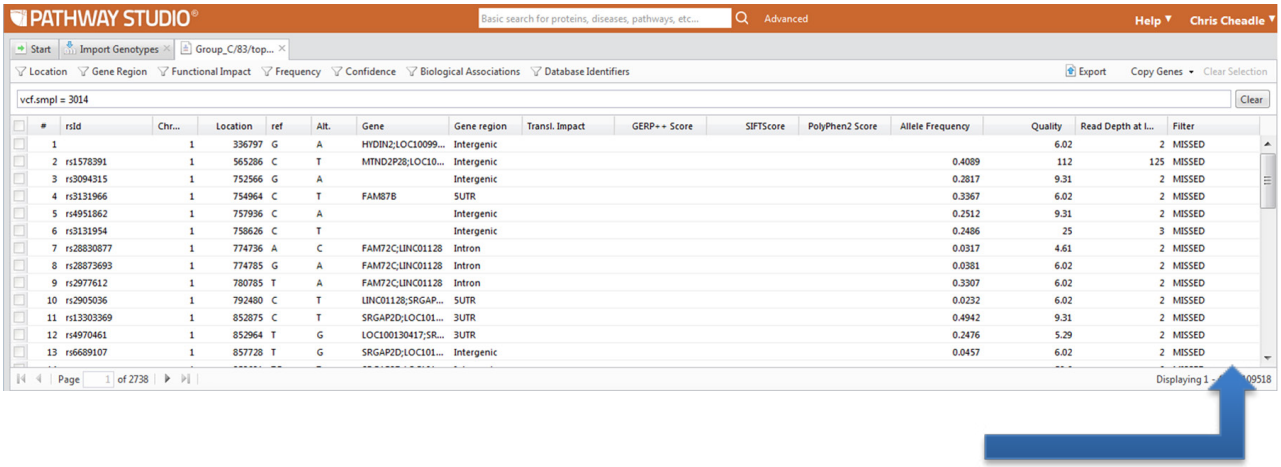
Download and save the file somewhere you can find it!

To upload a single file, go to the Start tab and select Import>Genotype:



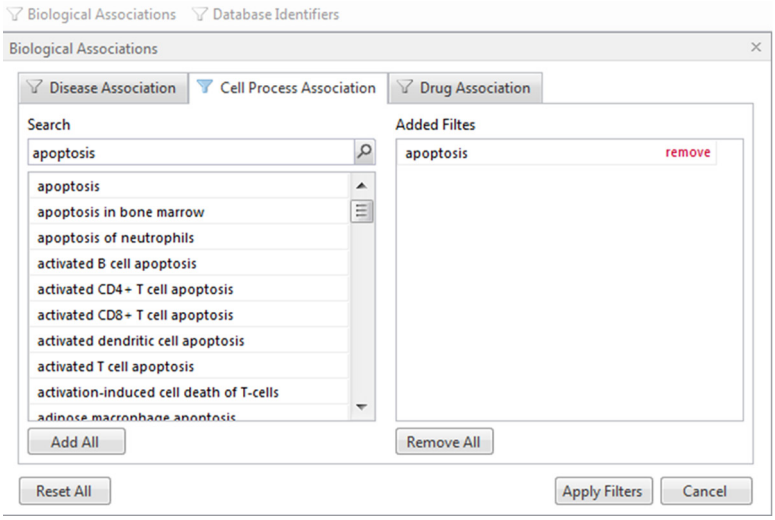
Double-click on the imported .vcf file now located in your My Projects folder. This will open up the data in the Variant Analysis window.





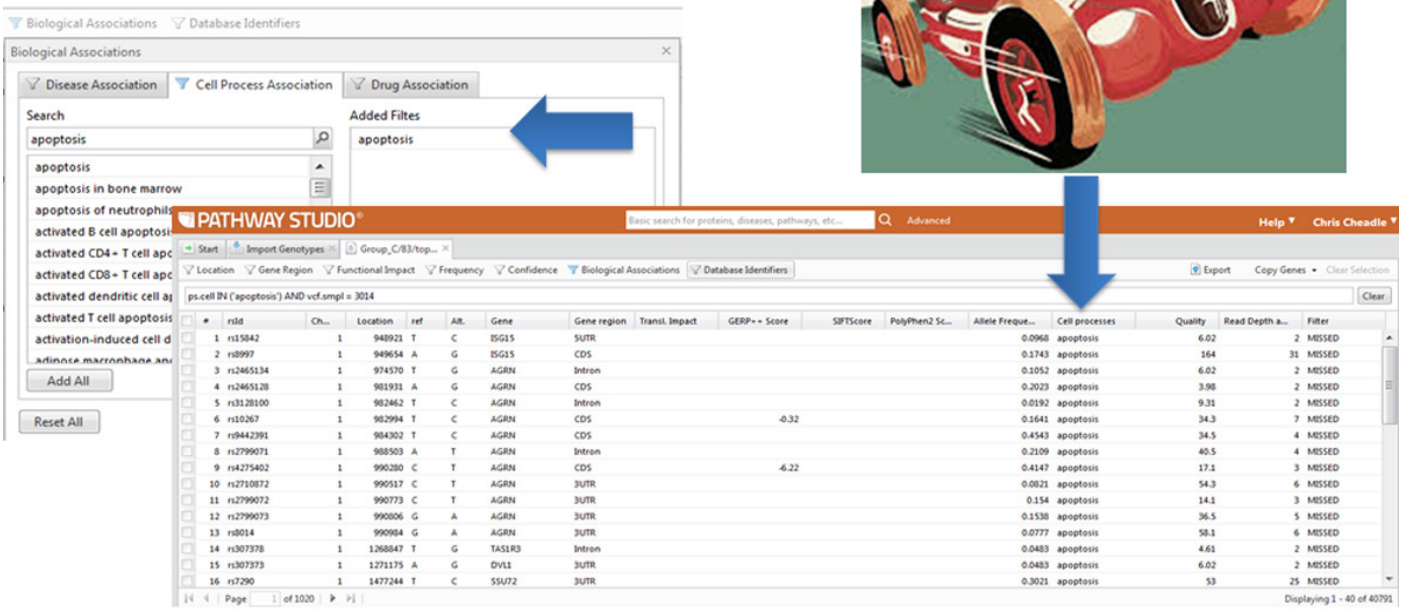
When the file is open, you will see that it has more than 100,000 variant SNPs.

Question: Are there any novel (not yet reported), potentially damaging variants in this data set in genes known to be associated with the cellular process of apoptosis?

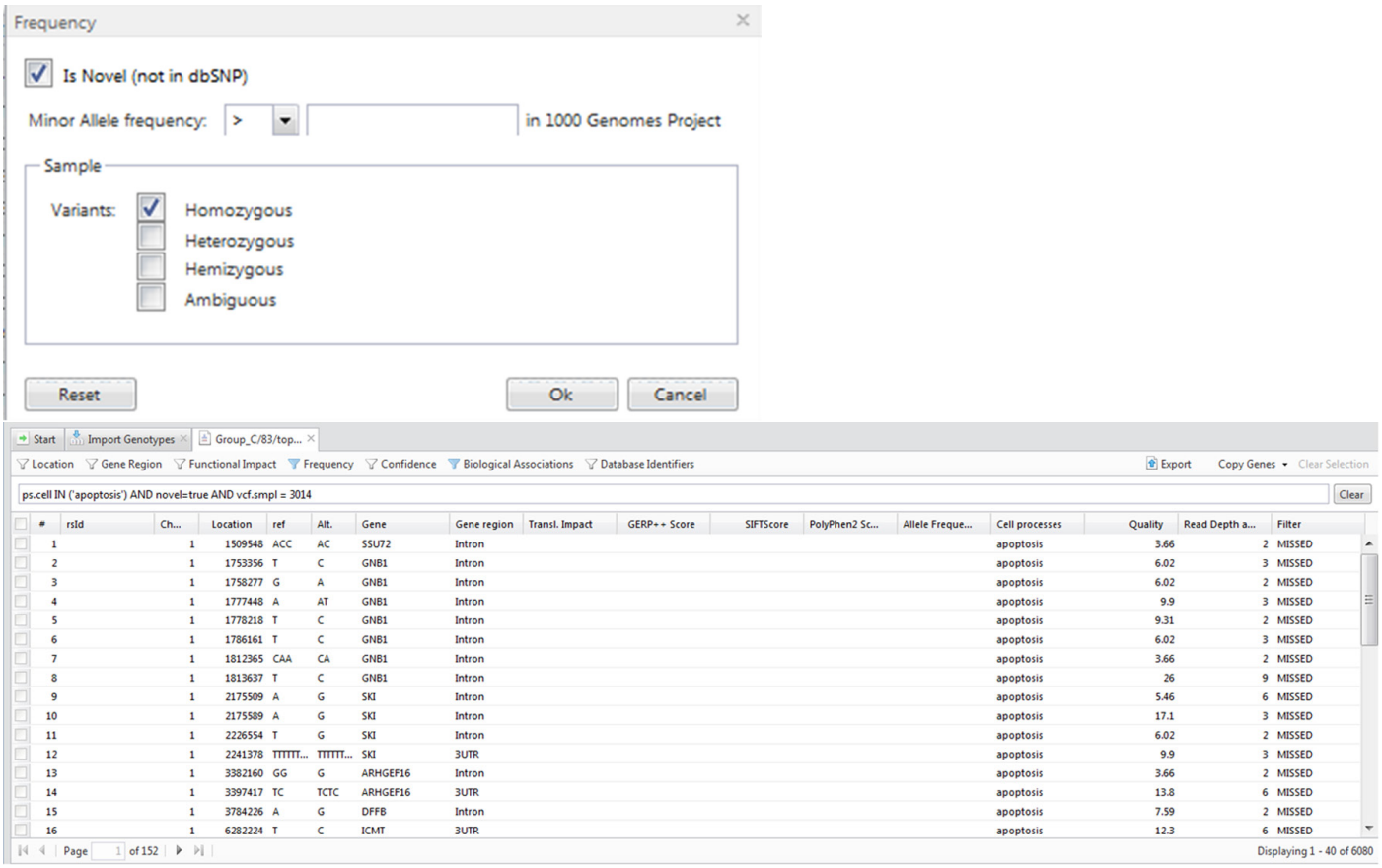


"Sounds like a job for Filterman!"

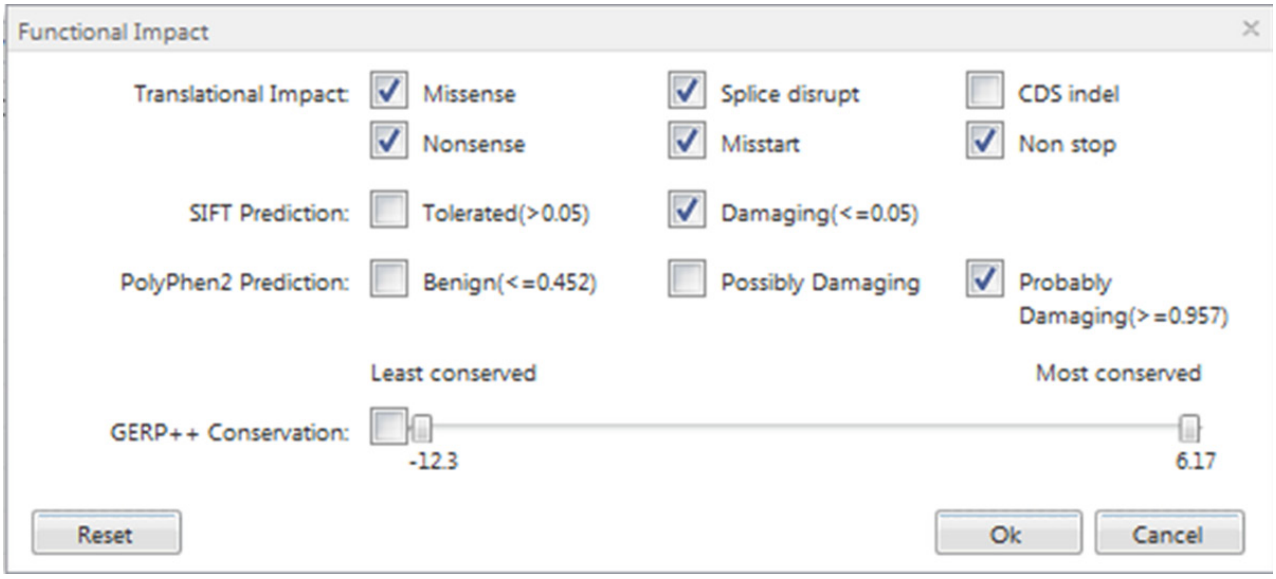
How does this work?



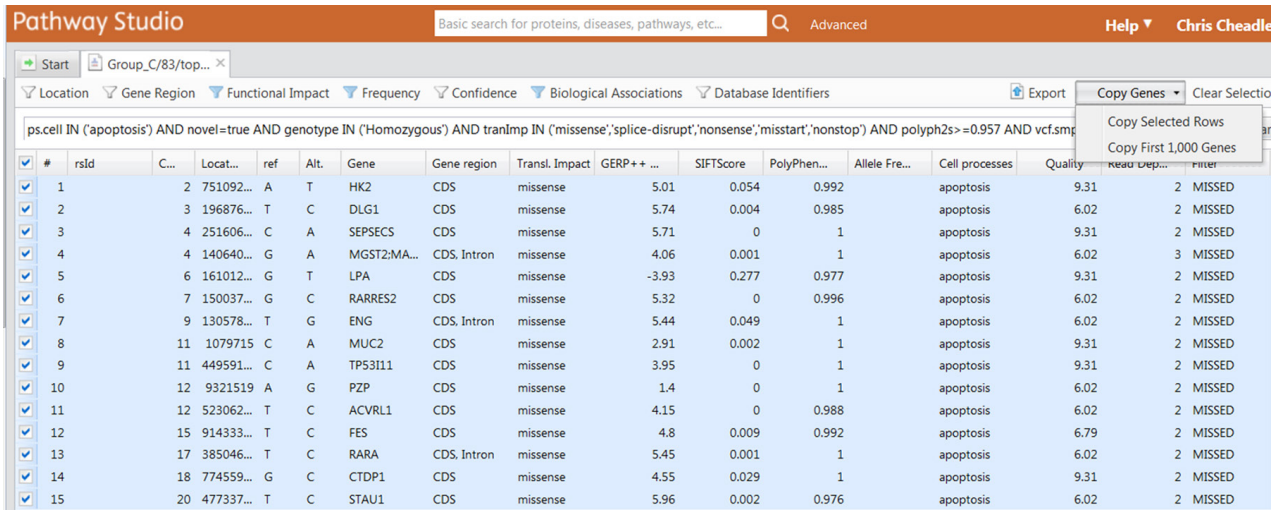
But wait we've forgot something! We said we wanted novel variants. How do we do that? (see Variant analysis I) OK, maybe you don't remember this but when we want something novel we can simply change the setting under Frequency to "Novel" (i.e., a variant not currently found in dbSNP) select "Homozygous" for maximum genetic impact.



And they should also be deleterious mutants: Functional Impact



Copy these genes over into Pathway Studio:

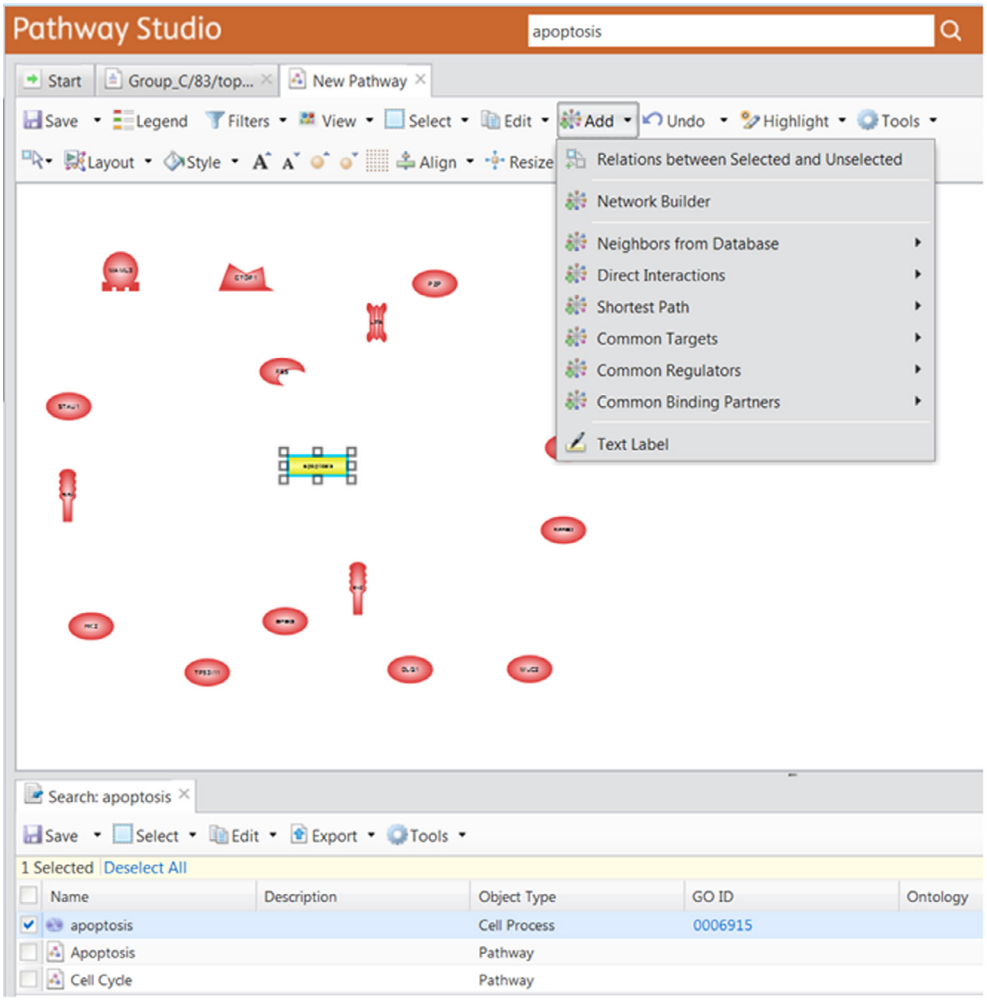


Create a New Pathway and paste in the filtered VCF genes:

Search on the term "apoptosis."

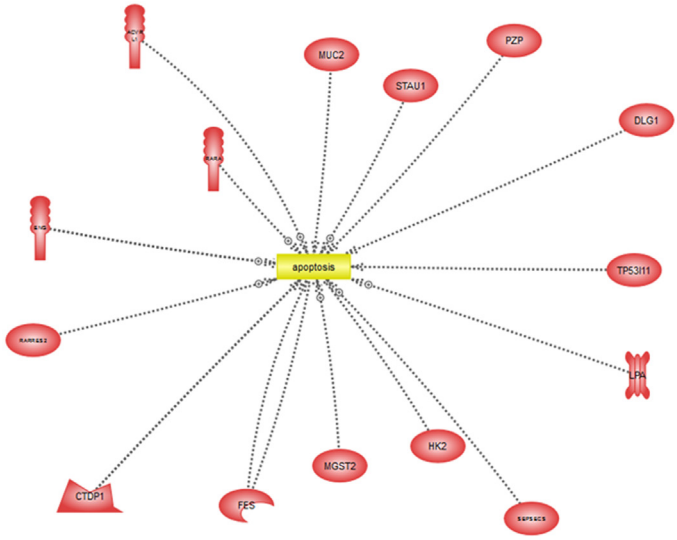
Add Relations between Selected (apoptosis) and Unselected (VCF genes).

What information is known about the association of each of these proteins with apoptosis? (Hint: Select apoptosis and add Relations between Selected and Unselected.)

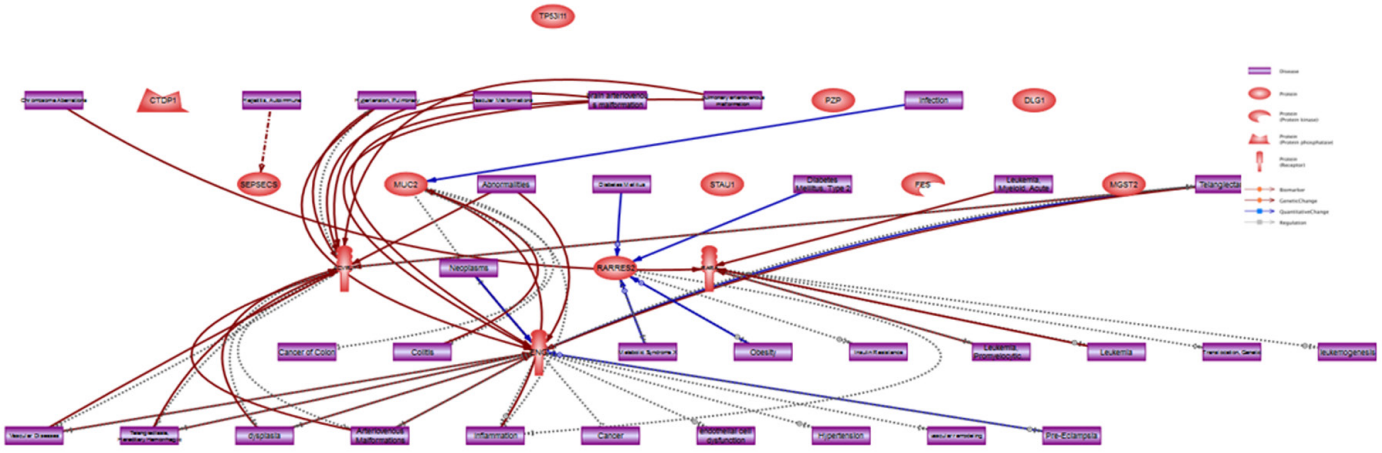


5.3 Exploring the Biology of Apoptosis-related Genes

Relation	Object Type	Effect	# of References
STAU1 --+> apoptosis	Regulation	positive	1
RARA --+> apoptosis	Regulation	positive	51
ACVRL1 --+> apoptosis	Regulation	positive	1
SEPSECS --+> apoptosis	Regulation	positive	1
MGST2 --+> apoptosis	Regulation	positive	1
RARRES2 --+> apoptosis	Regulation	positive	7
ENG --+> apoptosis	Regulation	positive	1
RARA ---| apoptosis	Regulation	negative	1
ENG ---| apoptosis	Regulation	negative	23
CTDP1 ---| apoptosis	Regulation	negative	1
PZP ---> apoptosis	Regulation		1
DLG1 ---> apoptosis	Regulation		3
FES ---> apoptosis	Regulation		2
RARA ---> apoptosis	Regulation		1
TP53I11 ---> apoptosis	Regulation		7
CTDP1 ---> apoptosis	Regulation		2
MUC2 ---> apoptosis	Regulation		2

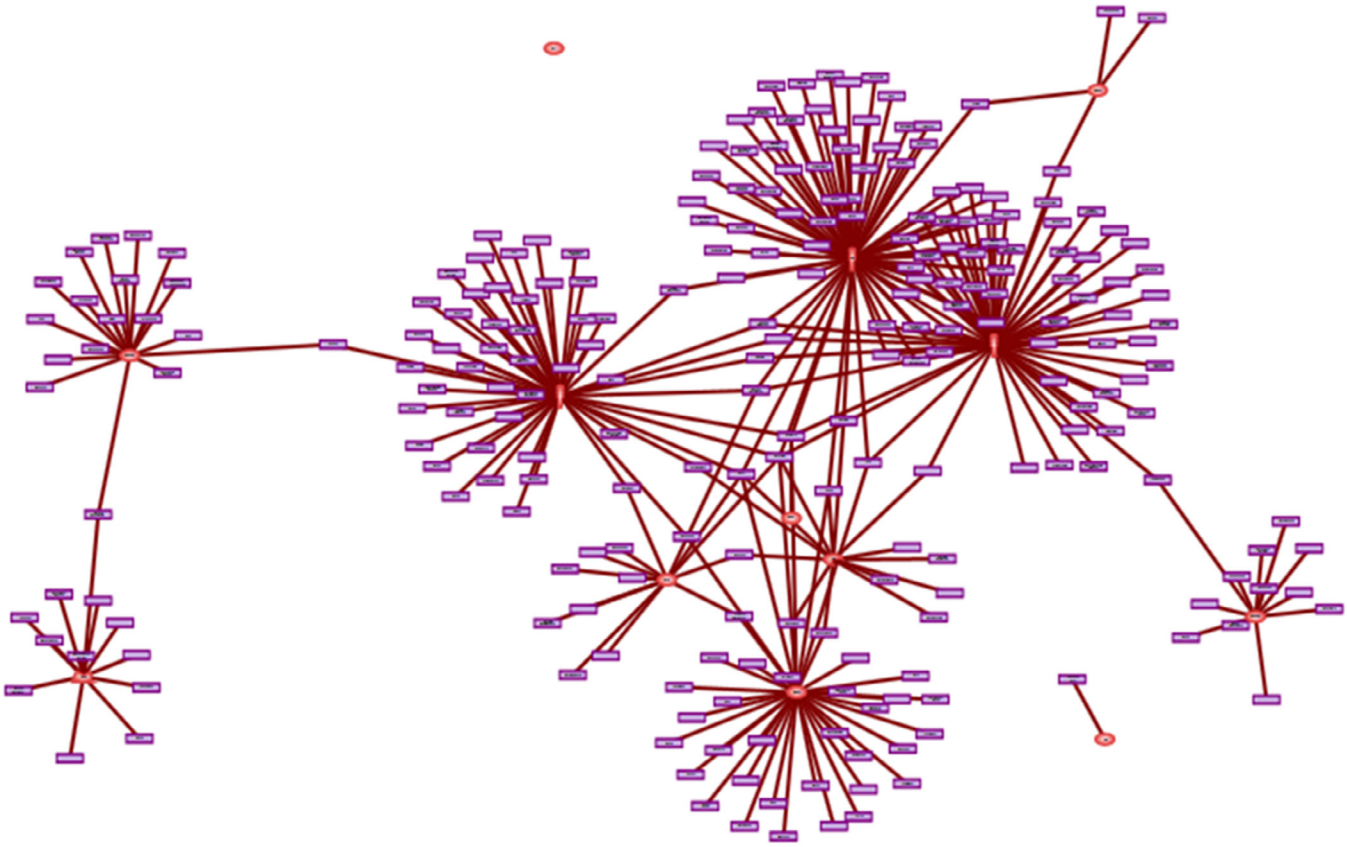


What are other cellular processes associated with these proteins? (Hint: Select all proteins; add "Network Builder," "Advanced Expand Pathway Tool," and "Cell Process with Regulation"(?); and filter for 10 references or more.)



Have mutations in any of these proteins been implicated in any diseases?

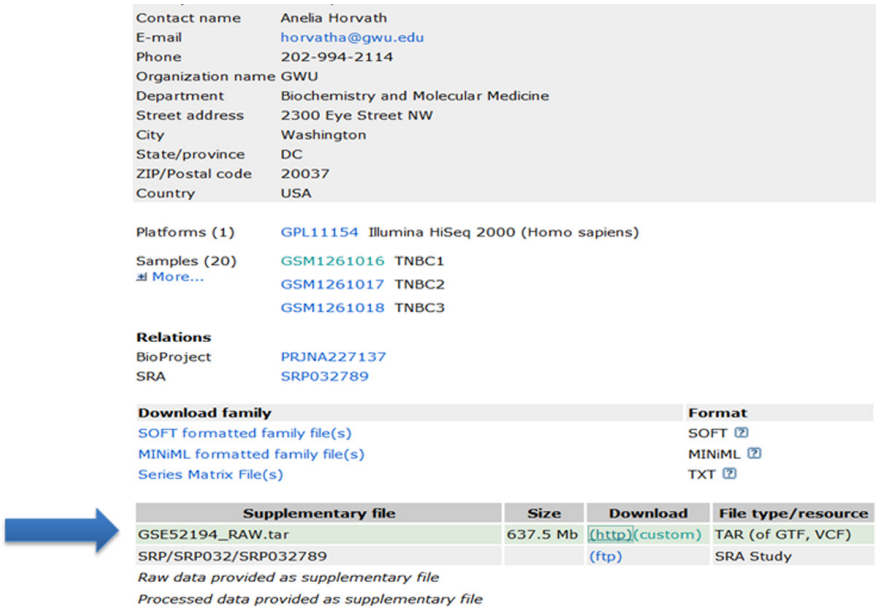
(HINT: Select all proteins; add "Network Builder," "Advanced Expand Pathway Tool," Disease with "GeneticChange," and ChangeType=Mutation; and set "Layout" to "Direct Force.")



5.4 Variant Analysis of Multiple Genomes

Upload multiple exome sequence files from NCBI's Geo site (GSE52194).

The study includes exome data from three breast cancer groups. You will use the HER2 positive breast group, and normal breast tissue (NBS) is included as a control.



Contact name	Anelia Horvath
E-mail	horvatha@gwu.edu
Phone	202-994-2114
Organization name	GWU
Department	Biochemistry and Molecular Medicine
Street address	2300 Eye Street NW
City	Washington
State/province	DC
ZIP/Postal code	20037
Country	USA

Platforms (1) GPL11154 Illumina HiSeq 2000 (Homo sapiens)

Samples (20) GSM1261016 TNBC1
GSM1261017 TNBC2
GSM1261018 TNBC3
More...

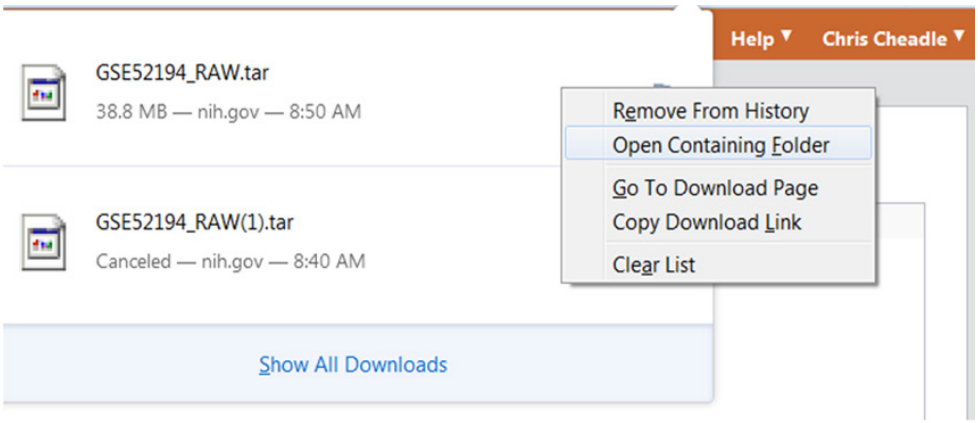
Relations

BioProject PRJNA227137
SRA SRP032789

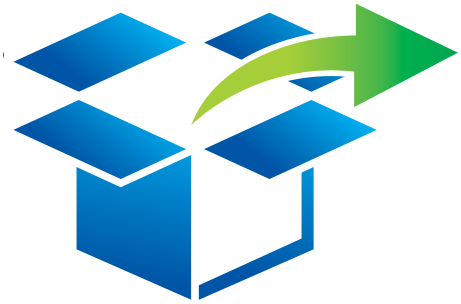
Download family	Format
SOFT formatted family file(s)	SOFT
MINiML formatted family file(s)	MINiML
Series Matrix File(s)	TXT

Supplementary file	Size	Download	File type/resource
GSE52194_RAW.tar	637.5 Mb	(http)(custom)	TAR (of GTF, VCF)
SRP/SRP032/SRP032789		(ftp)	SRA Study

Raw data provided as supplementary file
Processed data provided as supplementary file



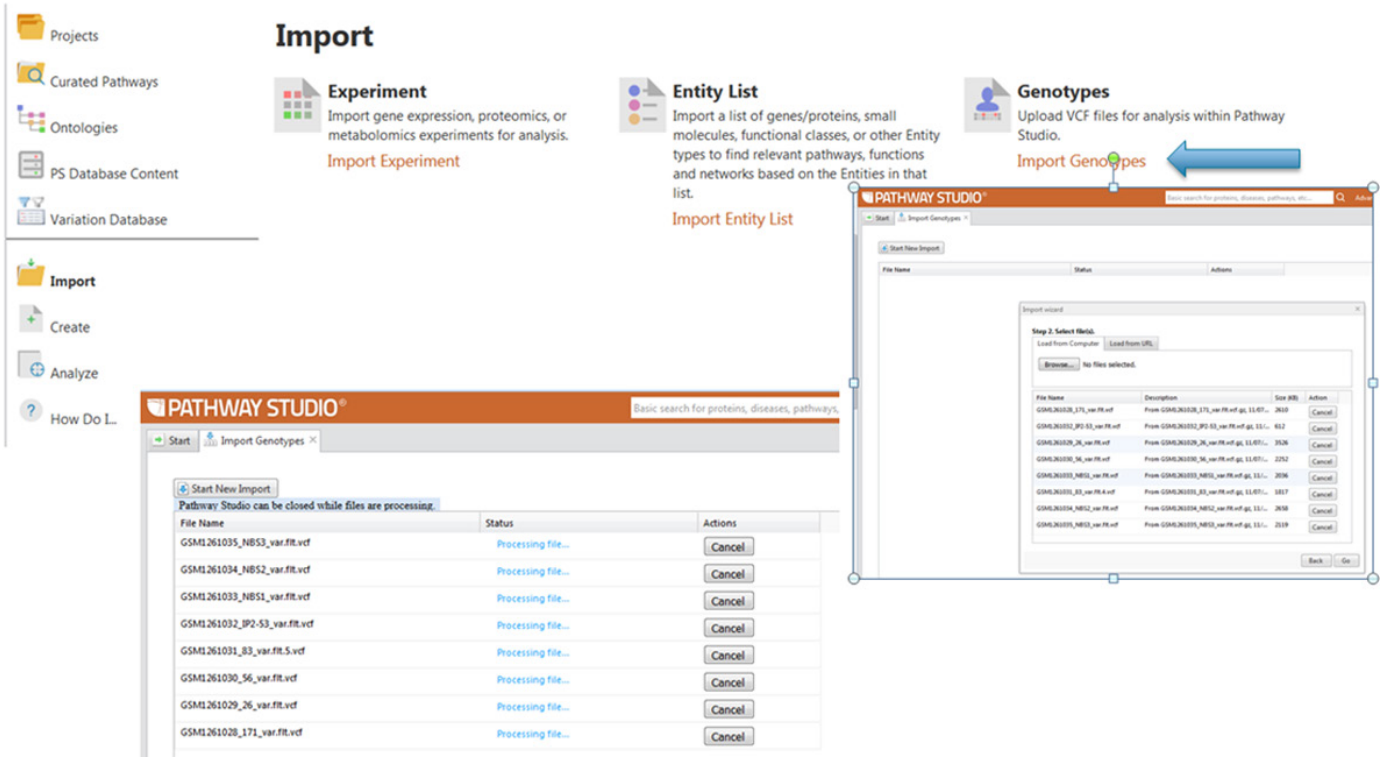
Extract all files and move files containing the file extension _var.flt.vcf.gz into a new folder (which you will be able to find later!).



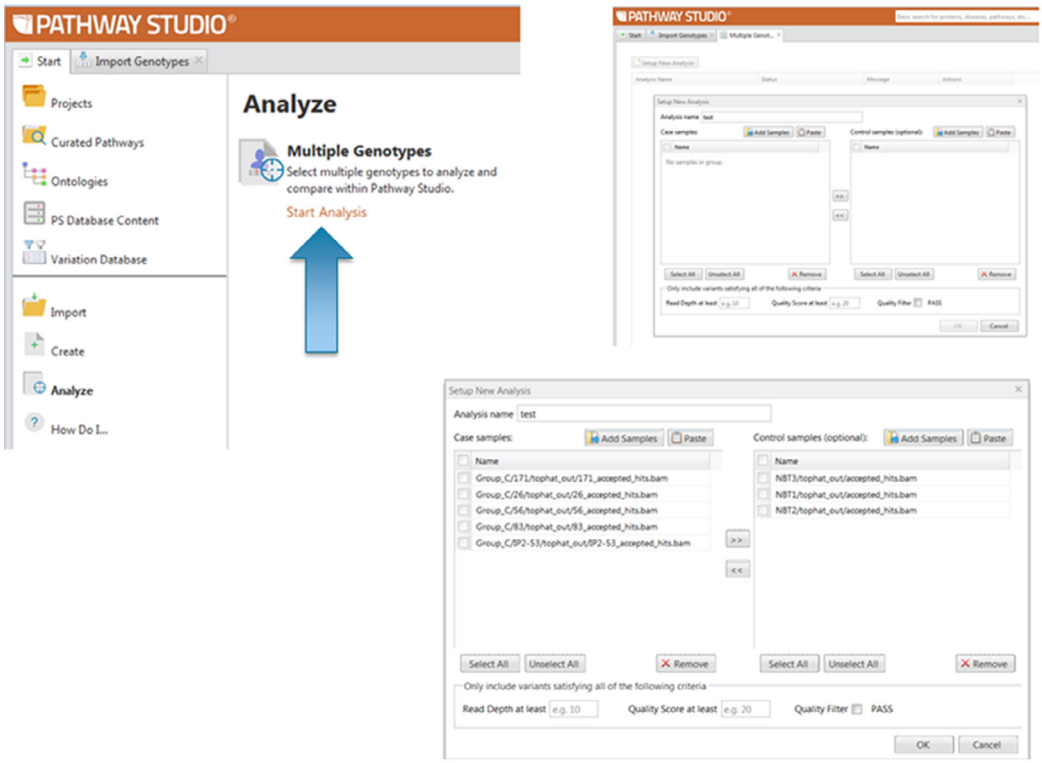
GSM1261028	HER2-1
GSM1261029	HER2-2
GSM1261030	HER2-3
GSM1261031	HER2-4
GSM1261032	HER2-5
GSM1261033	NBS1
GSM1261034	NBS2
GSM1261035	NBS3

5.5 Import Genotypes and Analyze Case versus Controls

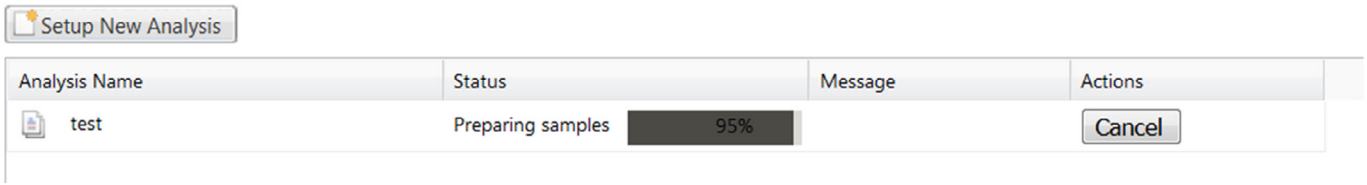
Import Genotypes.



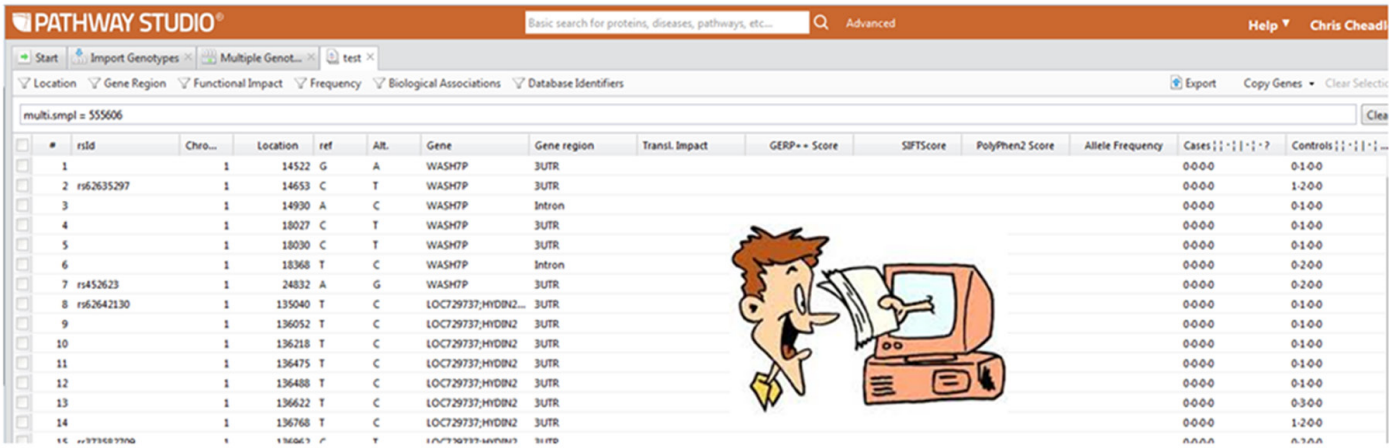
Add samples.



Turn the Crank!



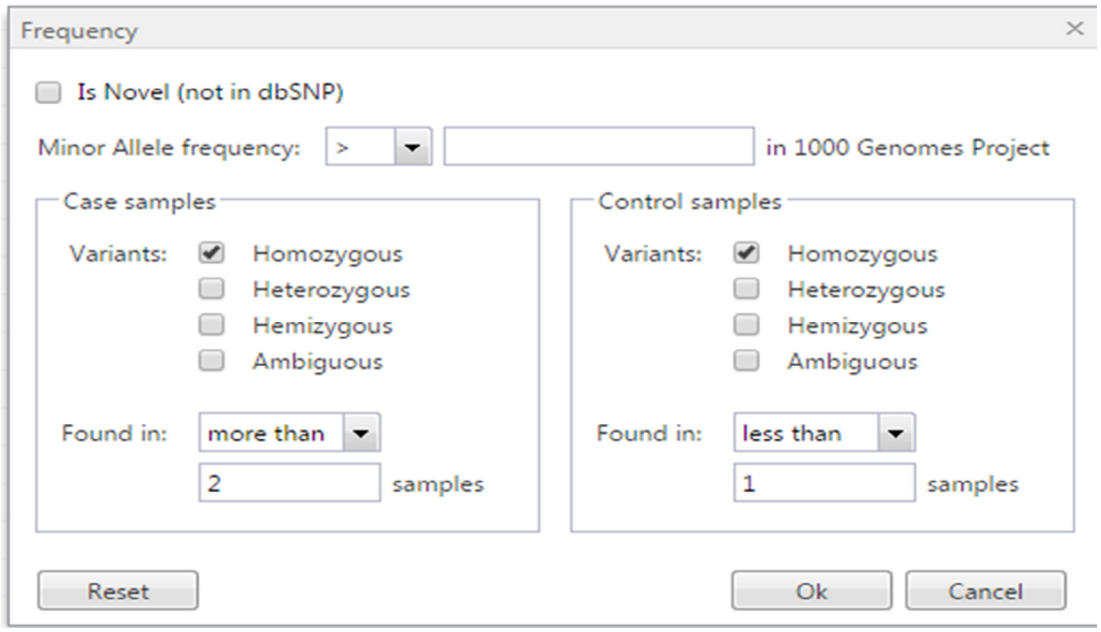
Output!

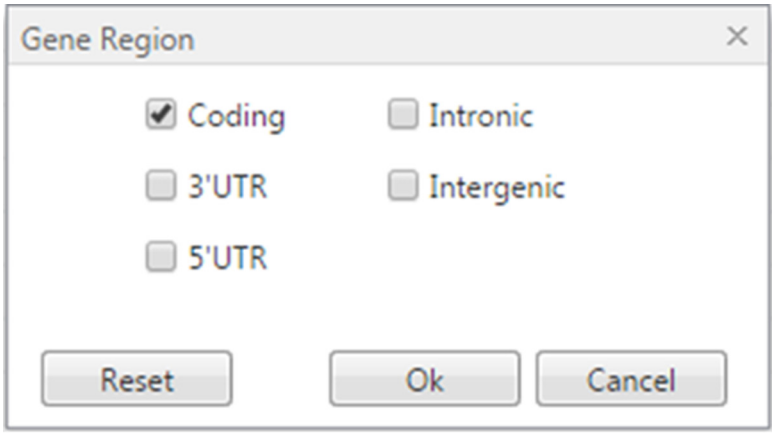


Now, we can start some data analysis!

Find homozygous variants that are present in at least three of the case samples. However, these variants cannot be in any of the control samples that are also in the coding region and are known to be associated with breast cancer.

In the Frequency filter, select for homozygous mutations that are present in at least three case samples (greater than 2) and not present in the control samples (less than 1).

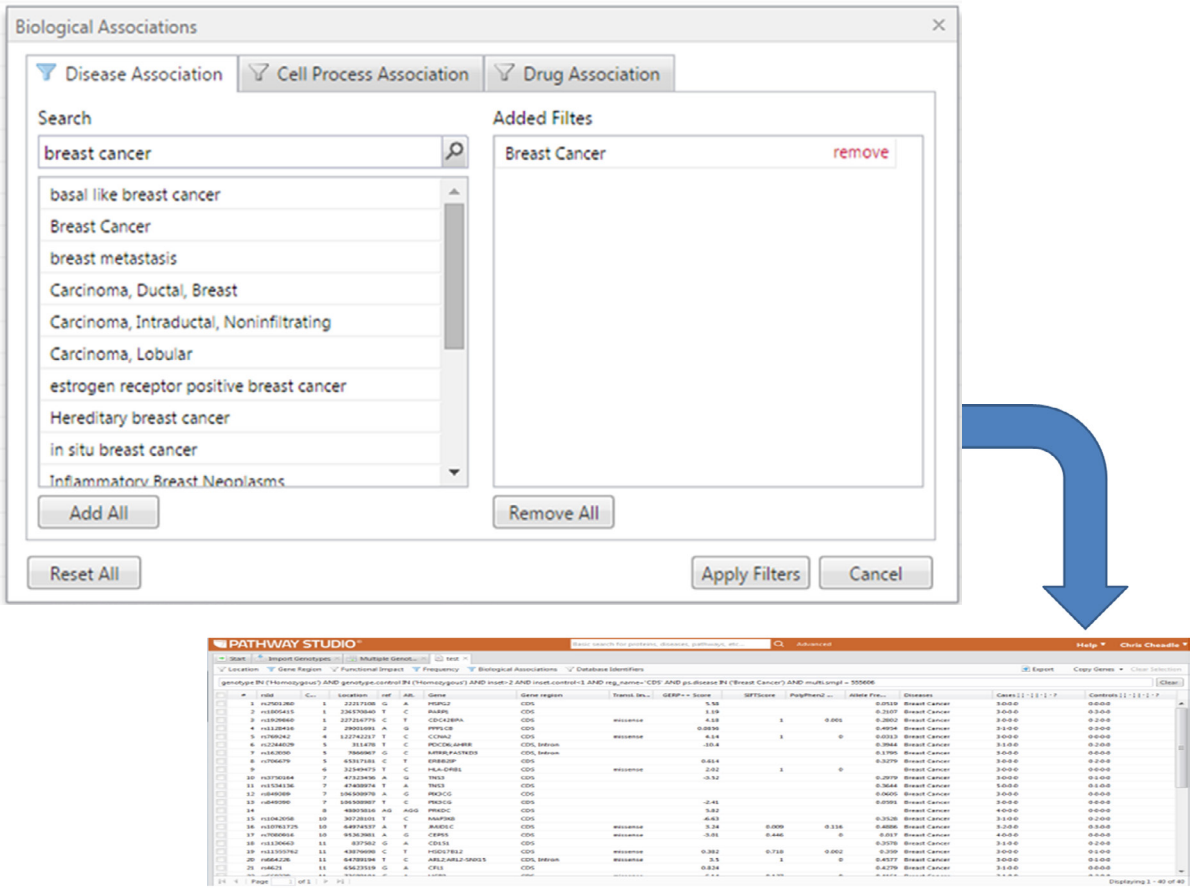




In the Gene Region filter, select for coding mutations.

5.6 Investigate the Biological Functions for Selected Genes

In the Biological Associations filter, select Disease Association>Breast Cancer.

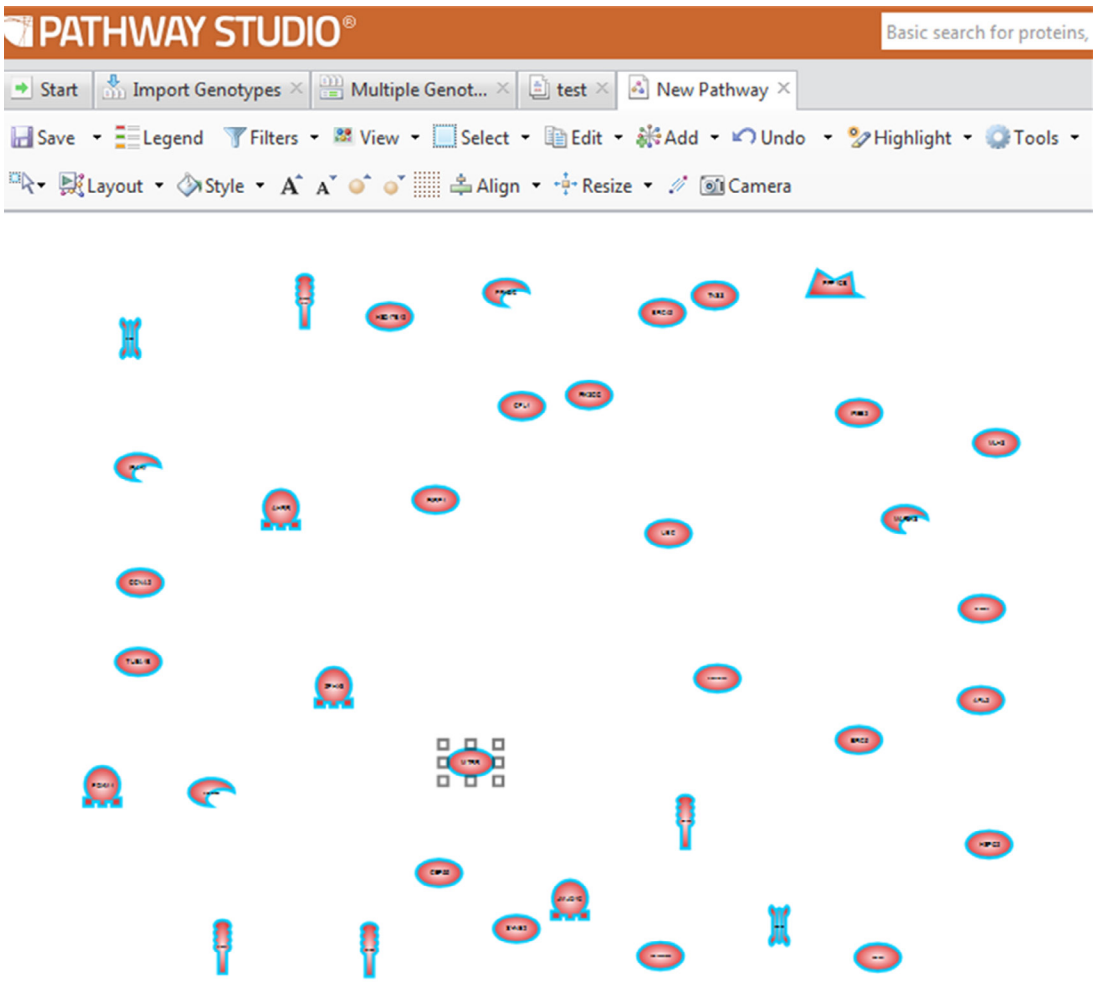


Note: In the Case and Control columns, insert the number of times a specific variant was identified is represented by four numbers in this order:

- homozygous
- heterozygous
- hemizygous
- ambiguous

Cases	Controls
3·0·0·0	0·0·0·0
3·0·0·0	0·3·0·0
3·0·0·0	0·2·0·0
3·1·0·0	0·3·0·0
3·0·0·0	0·0·0·0
3·1·0·0	0·2·0·0
3·0·0·0	0·0·0·0
3·0·0·0	0·2·0·0
3·0·0·0	0·0·0·0
3·0·0·0	0·1·0·0

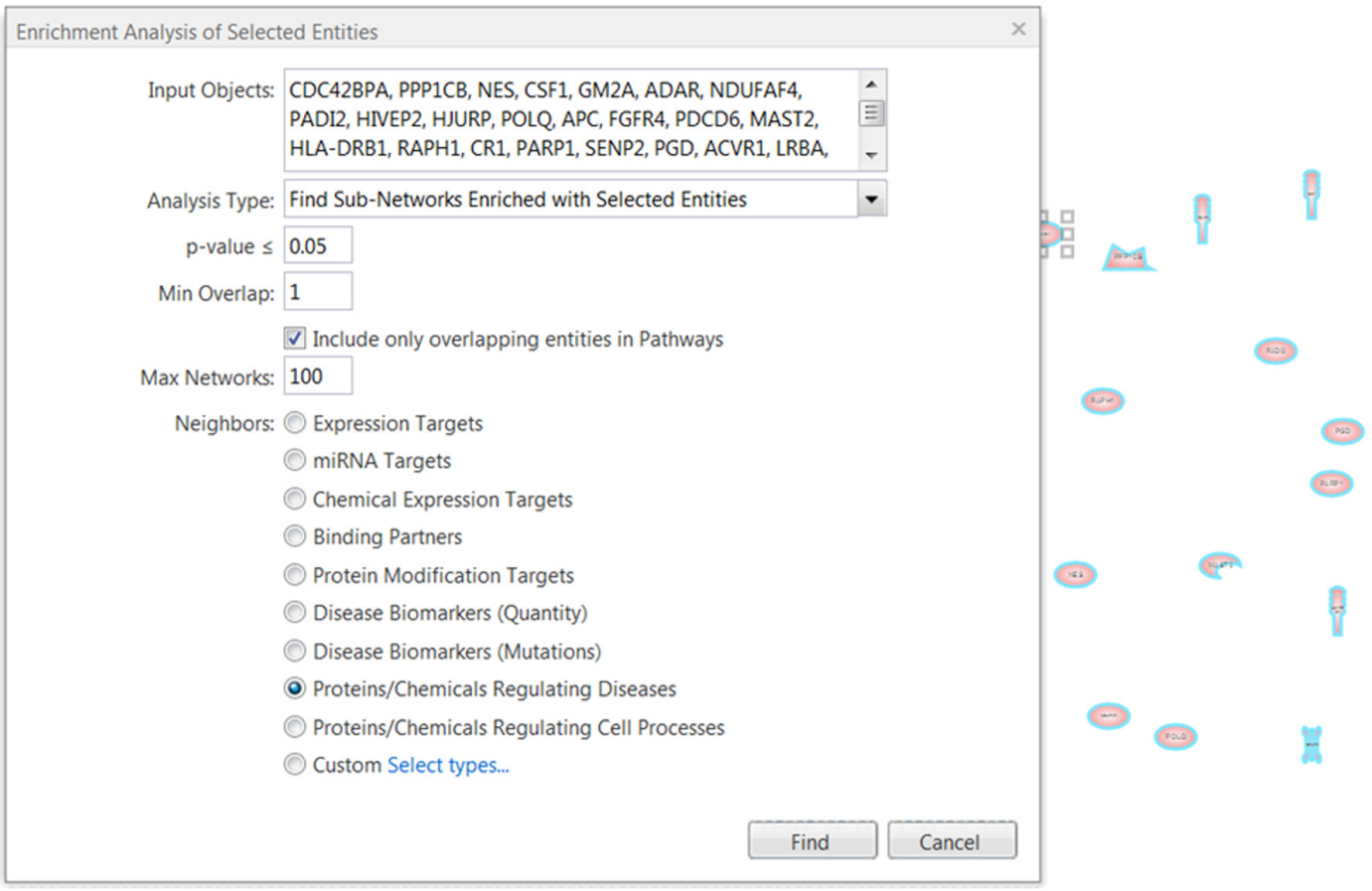
Copy the remaining genes after filtering into Pathway Studio:

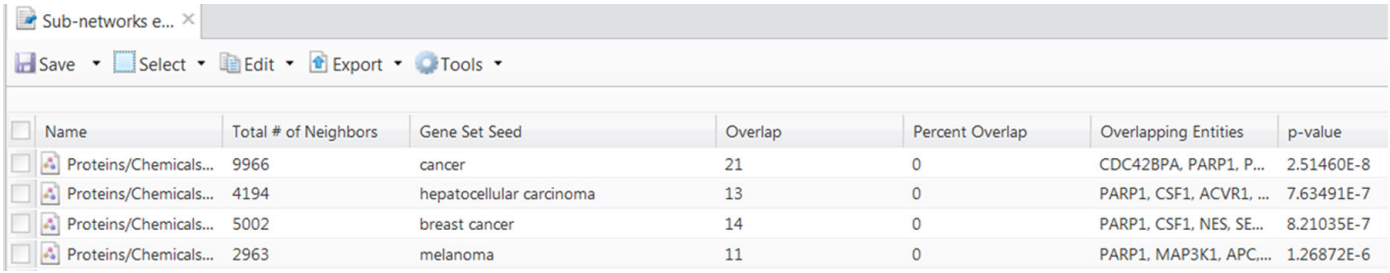


From this point, the list can be further investigated using the tools in Pathway Studio to answers questions such as:

- What information is known about the association of each of these proteins with breast cancer?
- What other cellular processes or diseases are associated with these proteins?
- Have mutations in any of these proteins been implicated in any other diseases?
- Are any of these proteins included in Elsevier's curated pathways?
- What proteins do these proteins regulate that a mutation might impact that regulation?

Finish up by performing a Sub-Network Enrichment Analysis (SNEA) on this group of genes/proteins for disease association.





Name	Total # of Neighbors	Gene Set Seed	Overlap	Percent Overlap	Overlapping Entities	p-value
Proteins/Chemicals...	9966	cancer	21	0	CDC42BPA, PARP1, P...	2.51460E-8
Proteins/Chemicals...	4194	hepatocellular carcinoma	13	0	PARP1, CSF1, ACVR1, ...	7.63491E-7
Proteins/Chemicals...	5002	breast cancer	14	0	PARP1, CSF1, NES, SE...	8.21035E-7
Proteins/Chemicals...	2963	melanoma	11	0	PARP1, MAP3K1, APC,...	1.26872E-6

We can see that the top 4 enriched diseases for the Breast Cancer DNA variants involve mutations in genes previously associated with cancer including, specifically, breast cancer.

But wait, before we go let's do a quick review of what we learned today.

- We learned how to identify, download, and analyze variant data.
- We learned how to identify genes known to be associated with apoptosis damaging variants using a single sample data set.
- We learned how to identify damaging variants in a case versus control multiple sample data set.
- We performed multiple drills for investigating biological functions for our network of selected genes.
- All in all, not a bad day's work!

And now we're done, see you next time!

Study Questions 5

1. For the single exome data used in this module (GSM1261041_83), identify the number of variants that are novel? (Hint: Suggested by the data, but not in the dbSNP database.) Identify the number of variants that are related to cell cycle.

2. After filtering for homozygous variants, how many SNPs are left?

3. After filtering for Functional Impact (as in Module 5), how many SNPs are left? How many genes are left?

4. Using the multiple exome data in Module 5, how many homozygous variants are present in at least three of the case samples but not in any of the control samples?

    a. After filtering for Coding Region only?

    b. After filtering for "cell proliferation" under Biological Associations?

5. Which gene(s) have variant SNPs for all five breast cancer patients and none in controls? (Hint: Export table to Excel for faster inspection.)

6. How many references connect these genes with cell proliferation?

For Relations Localized in a Tissue/ Organ/Cell Type/Cell Line, How Do I...

Exercise 5.1: Find relations that are known to be found in a specific organ?

Step 1: Create New Pathway or within Pathway, select a protein(s)

Step 2: Select Add- > Network Builder

Step 3: Select Expand Pathway, Advanced Expand Pathway Tool

Step 4: Select Direction: All

Step 5: Entities: Select "Protein" Relations: Select your specific interest

Add condition: Organ "is equal to" your specific interest

Continue to Network Builder Step 5 and click Launch Interactive Network Builder or Finish

Alternatively: After pathway is built, in the Relation Table View, add column for "organ."

Exercise 5.2: Find relations that are known to be found in a specific tissue?

Step 1: Create New Pathway or within Pathway, select a protein(s)

Step 2: Select Add- > Network Builder

Step 3: Select Expand Pathway, Advanced Expand Pathway Tool

Step 4: Select Direction: All

Step 5: Entities: Select "Protein" Relations: Select your specific interest

Add condition: Tissue "is equal to" your specific interest

Continue to Network Builder Step 5 and click Launch Interactive Network Builder or Finish

Alternatively: After pathway is built, in the Relation Table View, add column for "tissue."

Exercise 5.3: Find relations that are known to be found in a specific cell type?

Step 1: Create New Pathway or within Pathway, select a protein(s)

Step 2: Select Add- > Network Builder

Step 3: Select Expand Pathway, Advanced Expand Pathway Tool

Step 4: Select Direction: All

Step 5: Entities: Select "Protein" Relations: Select your specific interest

Add condition: CellType "is equal to" your specific interest

Continue to Network Builder Step 5 and click Launch Interactive Network Builder or Finish

Alternatively: After pathway is built, in the Relation Table View, add column for "celltype."

Exercise 5.4: Find relations that are known to be found in a specific cell line name?

Step 1: Create New Pathway or within Pathway, select a protein(s)

Step 2: Select Add- > Network Builder

Step 3: Select Expand Pathway, Advanced Expand Pathway Tool

Step 4: Select Direction: All

Step 5: Entities: Select "Protein" Relations: Select your specific interest

Add condition: CellLinename "is equal to" your specific interest

Continue to Network Builder Step 5 and click Launch Interactive Network Builder or Finish

Alternatively: After pathway is built, in the Relation Table View, add column for "CellLineName."

Module 6

Gene Expression Data Analysis

Contents

Study Questions 6

6.1 Import Pre-eclampsia Experimental Data

RNA Data Analysis

Data Import*

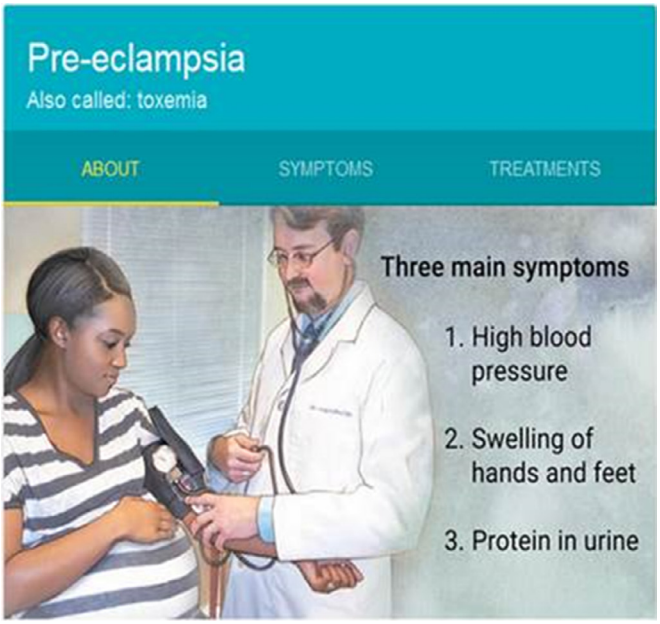
- Microarray
- RNASeq
- Semi-automatic versus manual

*Data Import protocols from GEO, from tab-delimited files, full data sets and differential calculations only are available in Appendices 1.A through 1.C.

Data Analysis

- Experiment Summary Analysis
- **S**ub-**N**etwork **E**nrichment **A**nalysis (**SNEA**) – (unique to Pathway Studio)

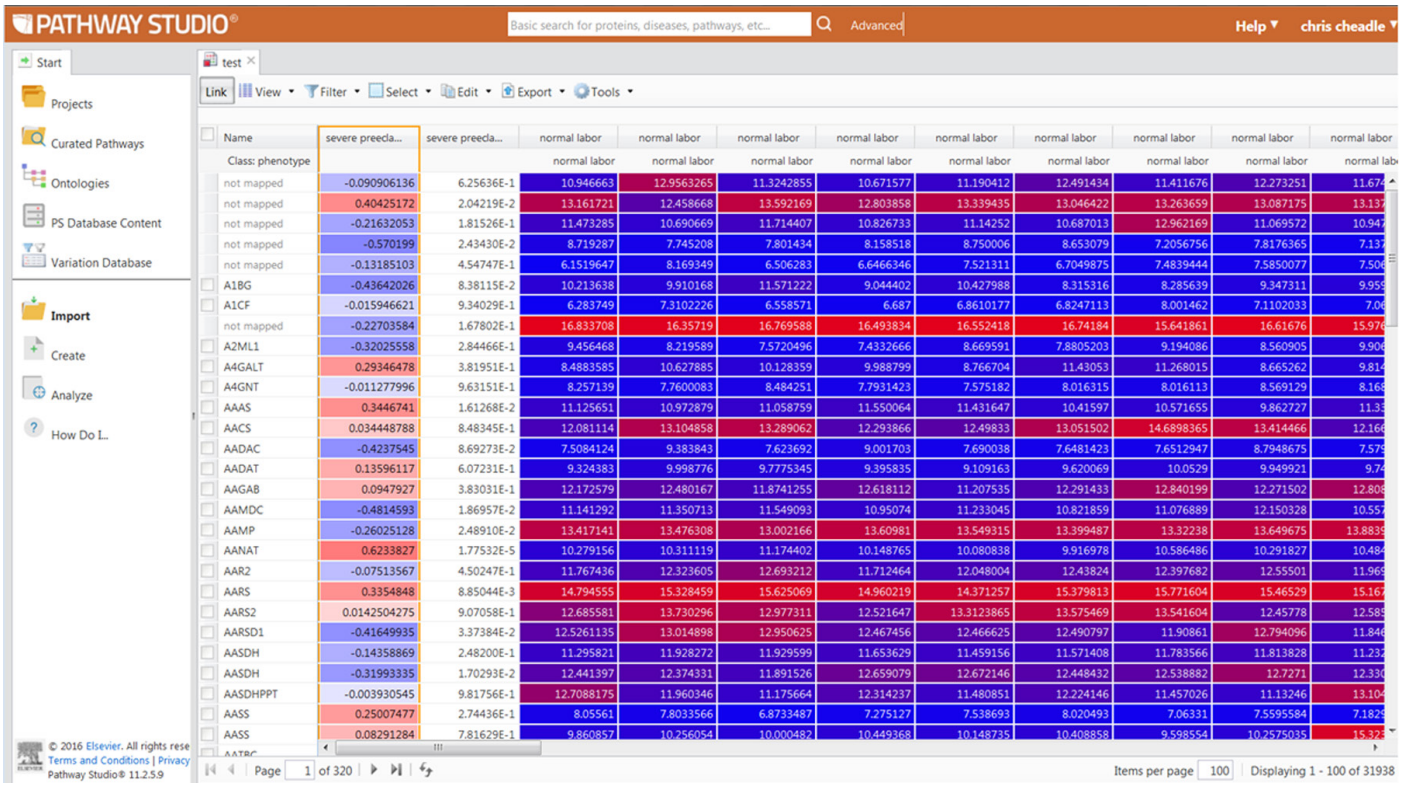
Pre-eclampsia or preeclampsia (PE) is a disorder of pregnancy characterized by high blood pressure and a large amount of protein in the urine. The disorder usually occurs in the third trimester of pregnancy and gets worse over time. In severe disease there may be red blood cell breakdown, a low blood platelet count, impaired liver function, kidney dysfunction, swelling, shortness of breath due to fluid in the lungs, or visual disturbances. Preeclampsia increases the risk of poor outcomes for both the mother and the baby. If left untreated, it may result in seizures at which point it is known as eclampsia.



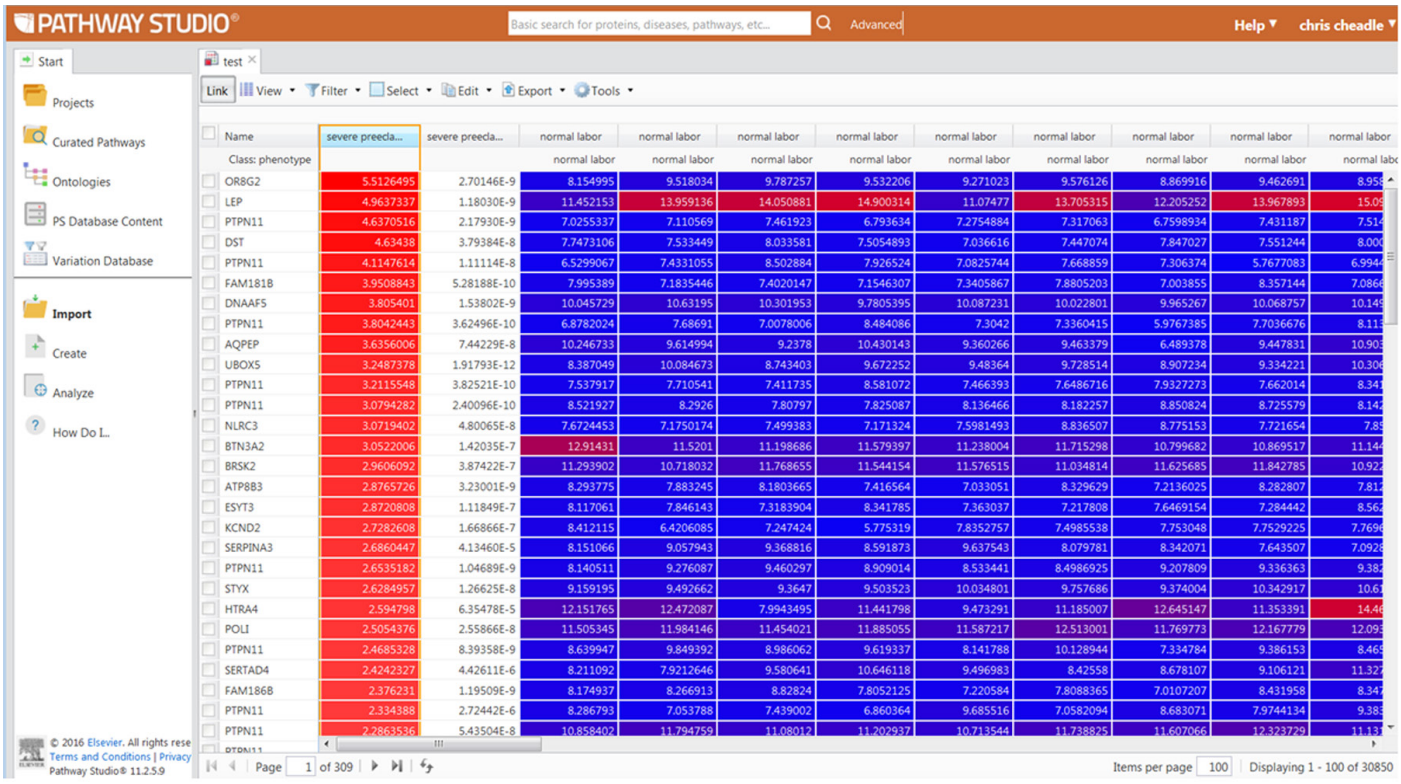
Data Analysis: Pre-eclampsia example.

Note: Data is derived from GEO data set, GSE10588 "Placental gene expression in severe preeclampsia." (please note that the platform used for this microarray study is the now discontinued ABI Human Genome Survey Microarray Version 2, preventing direct import from GEO. A properly mapped data file ready for import into Pathway Studio is available from Dropbox (GSE10588_Pre-eclampsia_dataset_10-07-16).

First Step: Import your data (Hint: Make sure during the import you indicate that the data sample type = "Log-intensity.")

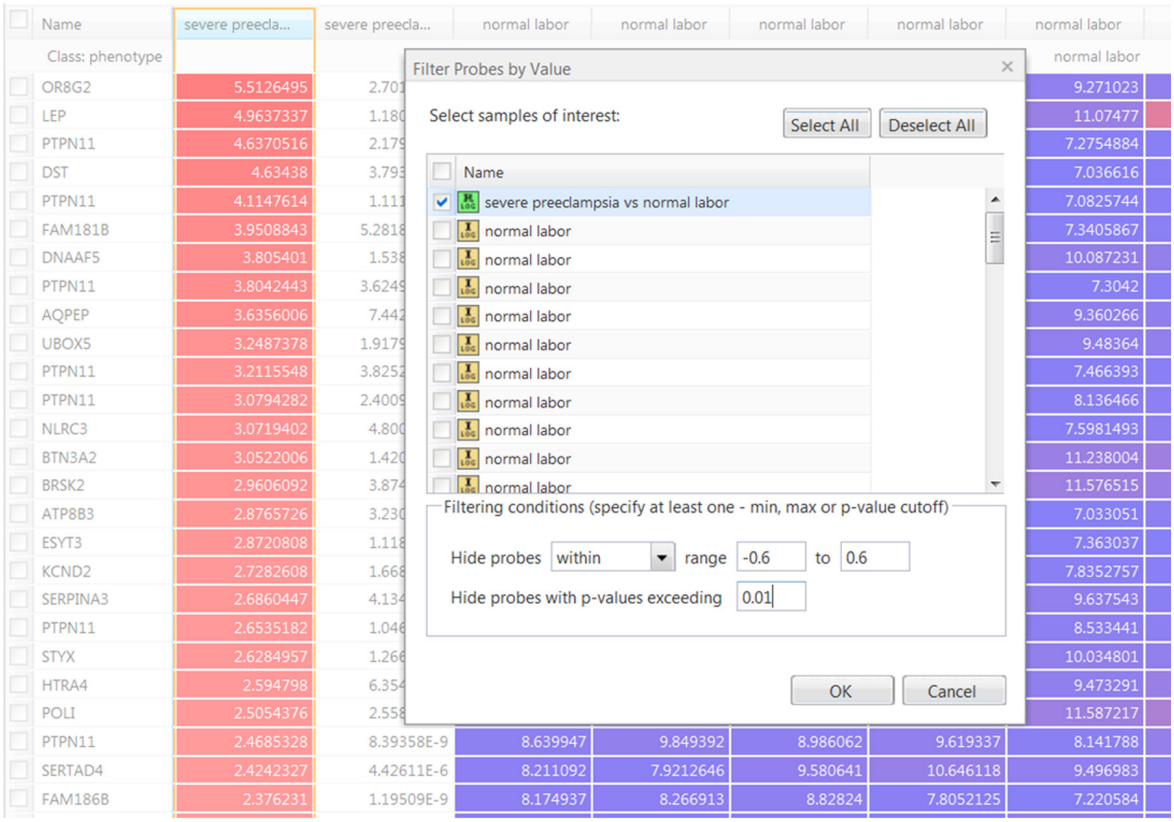


Next Step: View, hide Unmapped Probes, and sort by Fold Changes (Descending):

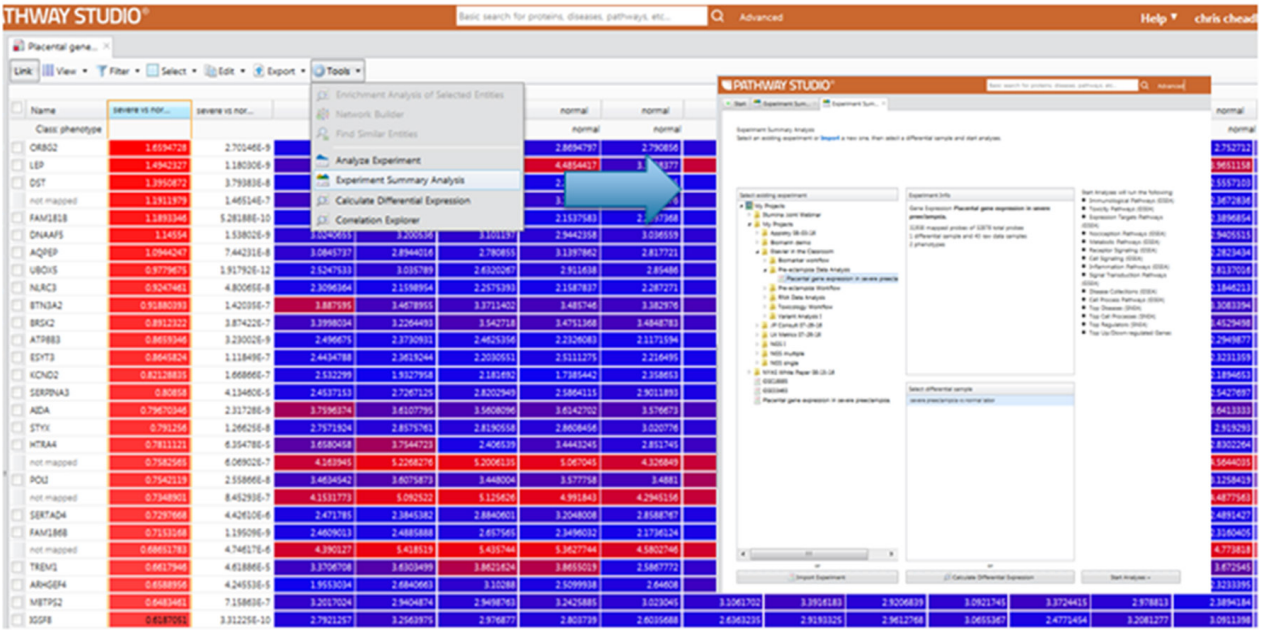


6.2 Filter Probes by Value

Filter probes by value (FC ± 0.6, p-val <= 0.01):



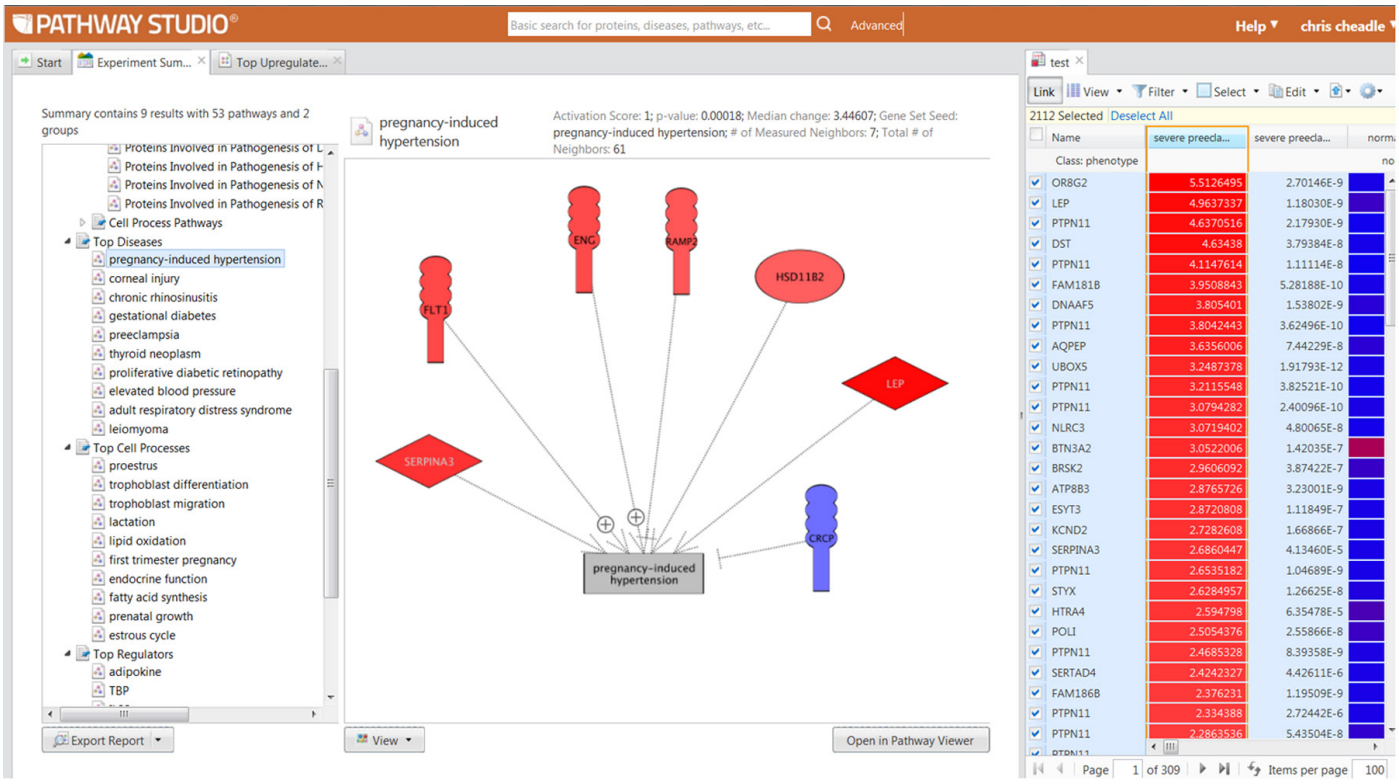
Select "Probes Remaining After Filtering" and run Experiment Summary Analysis:



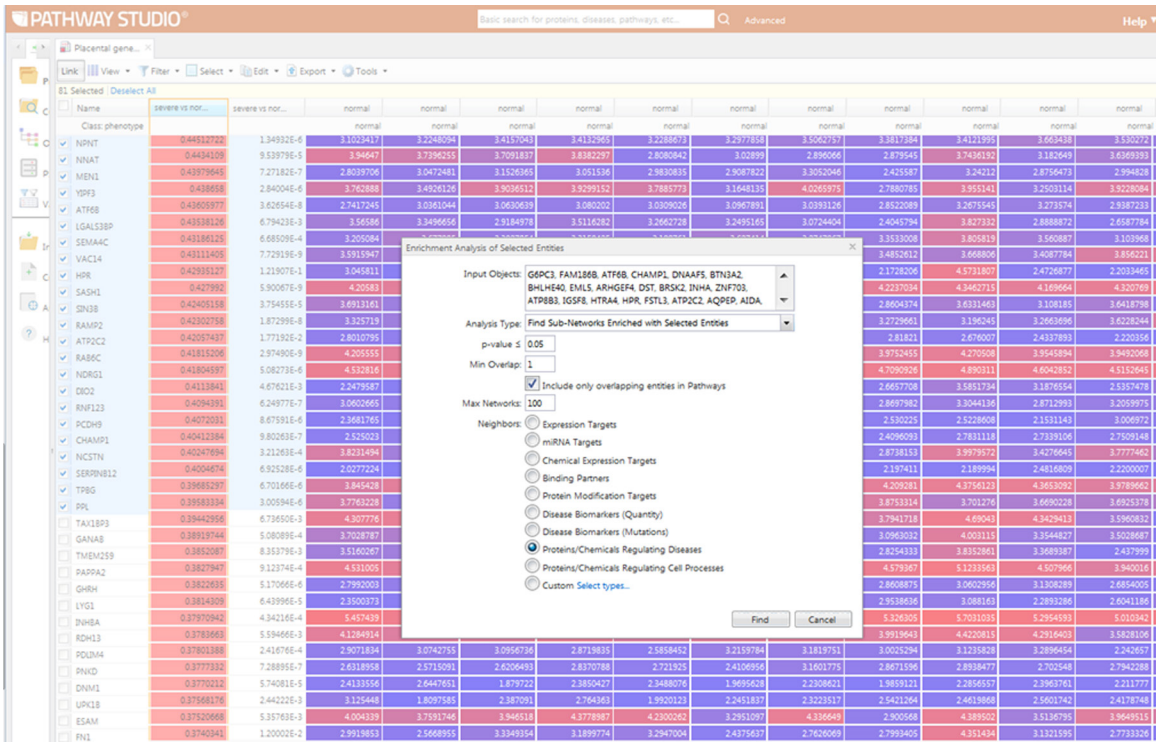
Select the proper class comparison and then "Start Analyses."

6.3 Explore Enriched Pathways

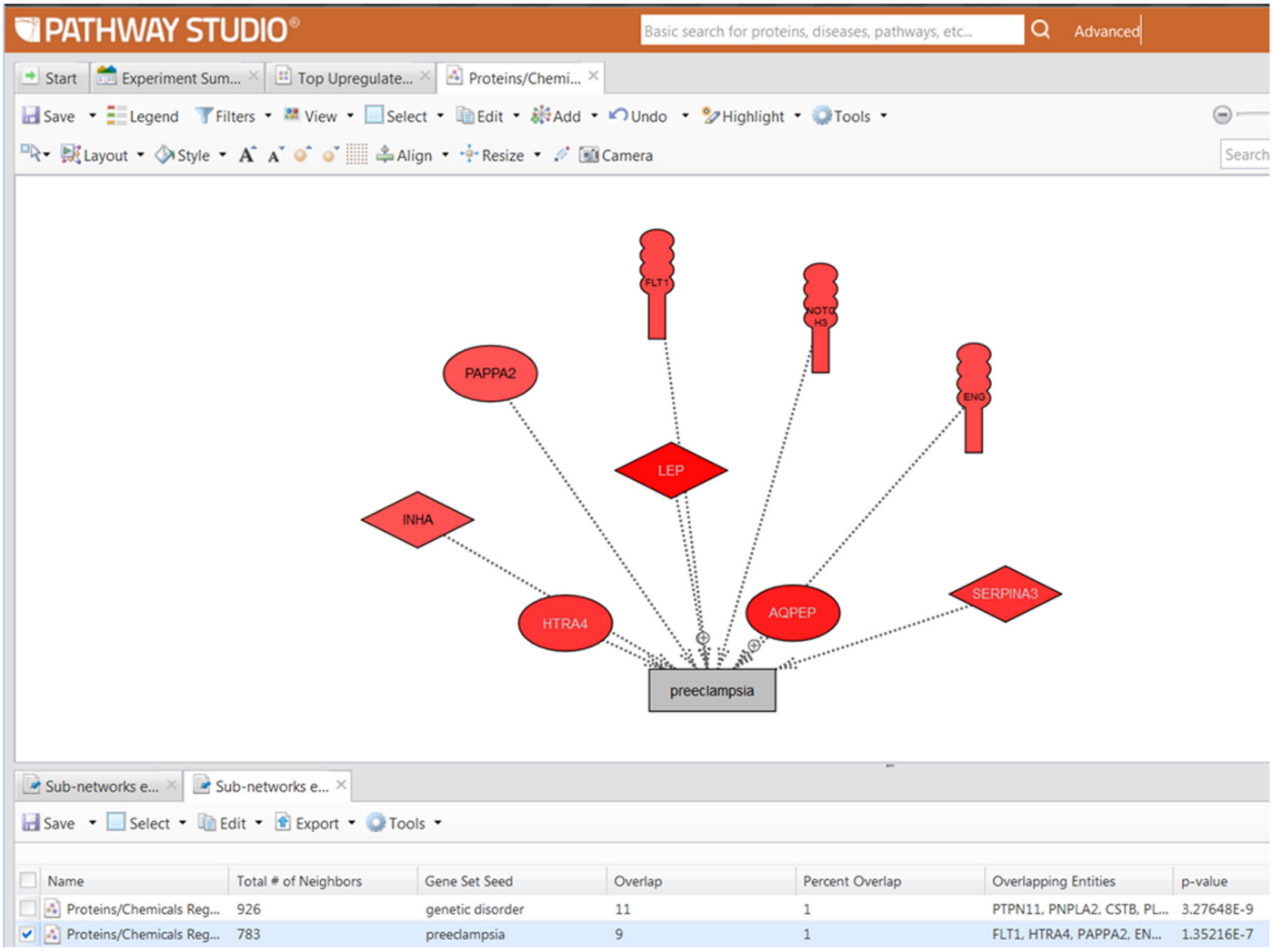
Featured here are genes involved with pregnancy-induced hypertension that is up-regulated in placental tissue from patients with severe pre-eclampsia.



Sub-Network Enrichment Analysis (SNEA) of top 100 upregulated genes.

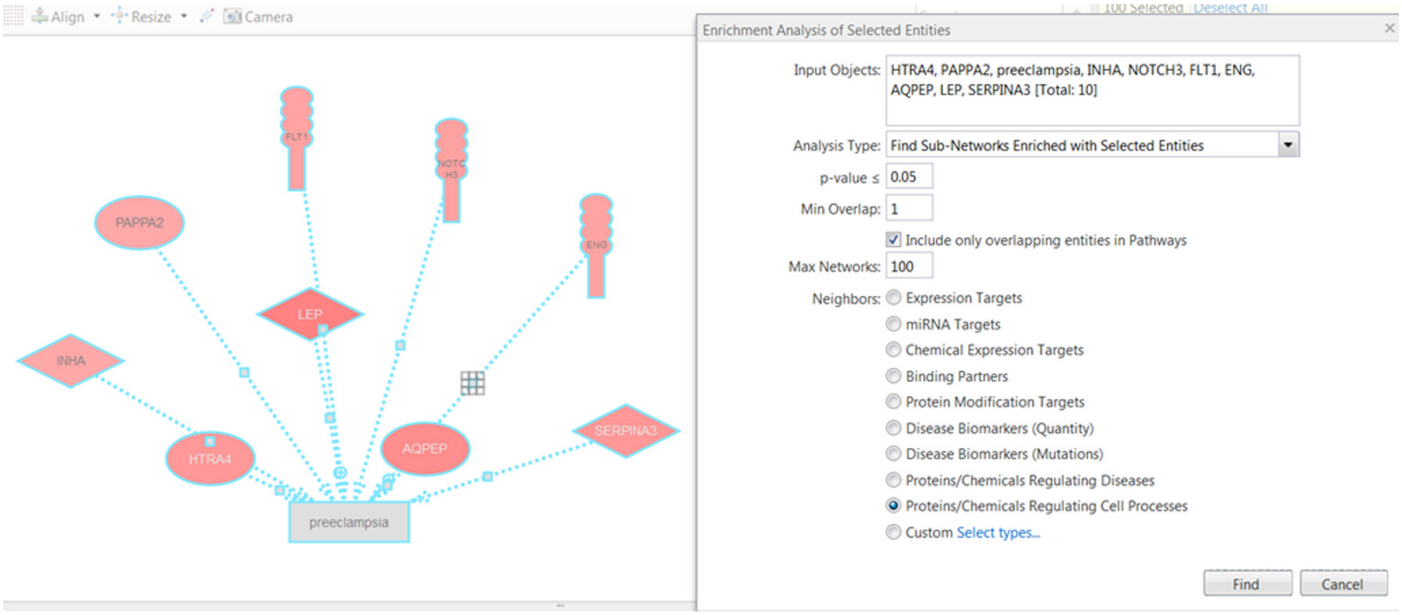


SNEA: diseases and genes/proteins leads to...



...identification of genes directly associated with pre-eclampsia as reflected in the scientific literature.

Question: What is the functional significance of this group of genes?

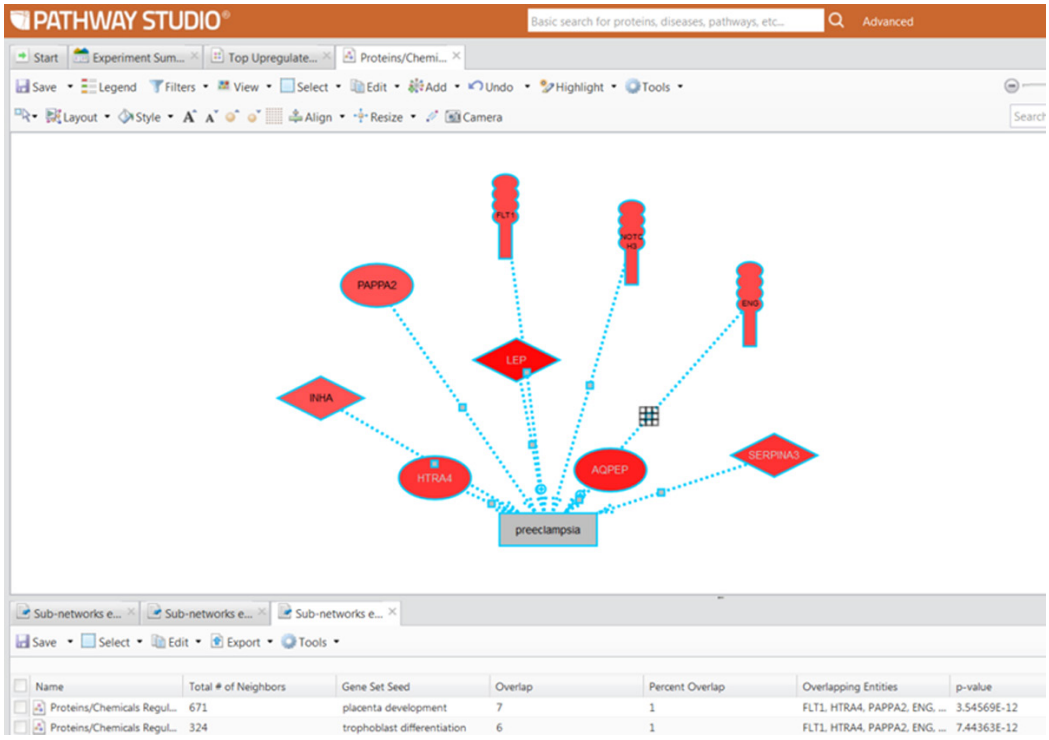


SNEA: cell process and genes/proteins

6.4 What Exactly is SNEA Doing, Anyway?

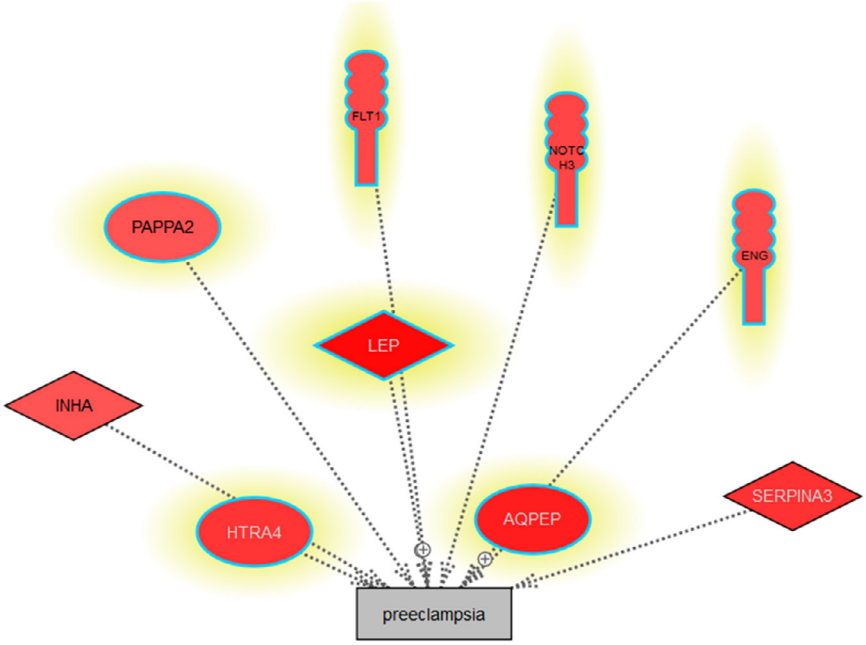
What exactly is SNEA doing, anyway?

When you perform an SNEA (for diseases and genes/proteins, for example), the Pathway Studio program first identifies all relations between diseases and genes/proteins (>400K) in the entire database. This becomes the basis (the sub-network) for enrichment analysis for the input genes/proteins in your list. The power of this approach is that the sub-network, which is generated "on the fly" every time an analysis is performed, always reflects the latest updates to the database (performed weekly). Another BIG advantage is that any combination of entities and relations as supported in the database overall can be precisely specified by the user. This allows for highly customized analyses that are either broad-based or fine-tuned. So, try it, you'll like it!

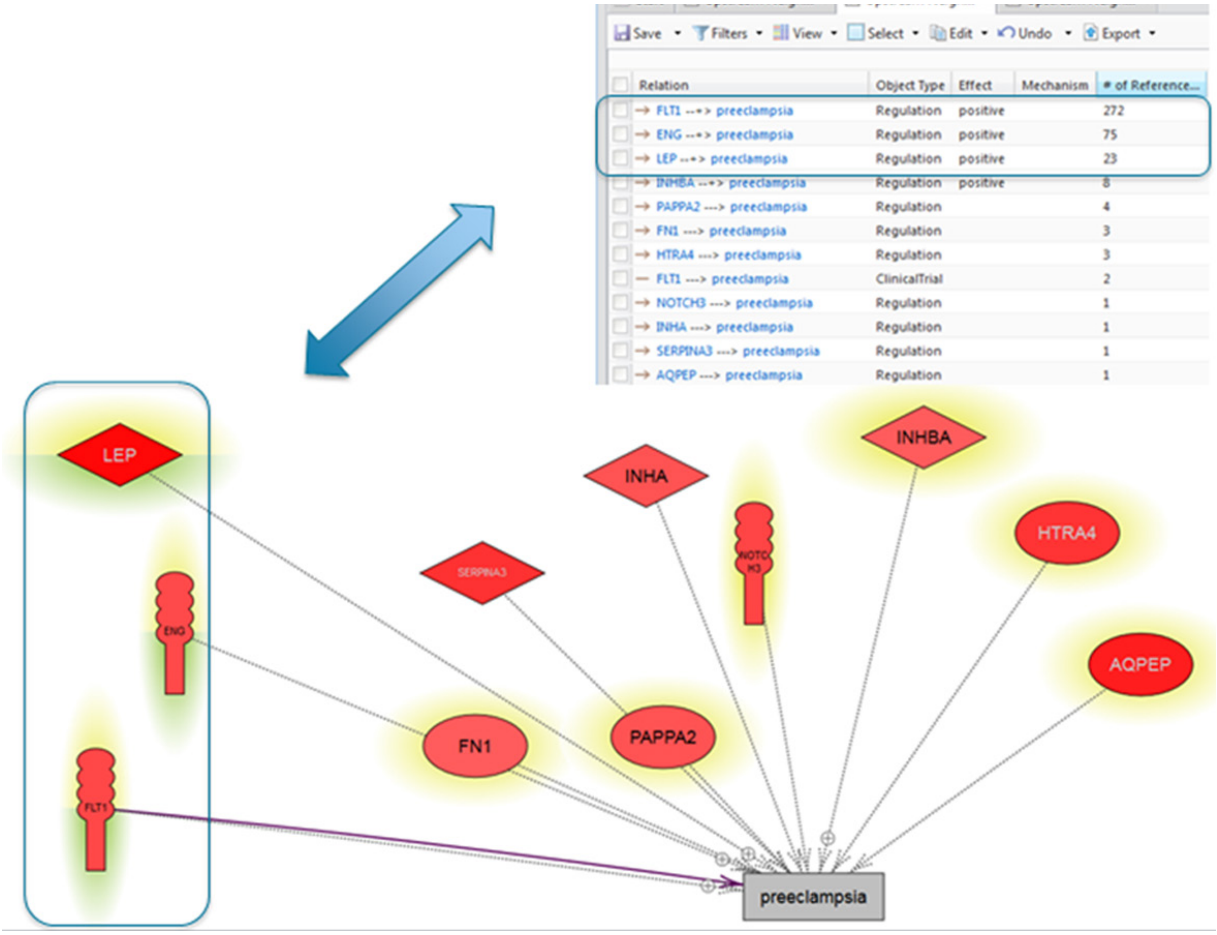


SNEA: cell process and genes/proteins

Placental development genes (yellow) overlaid on pre-eclampsia associated genes are clearly a major biological theme for this group.

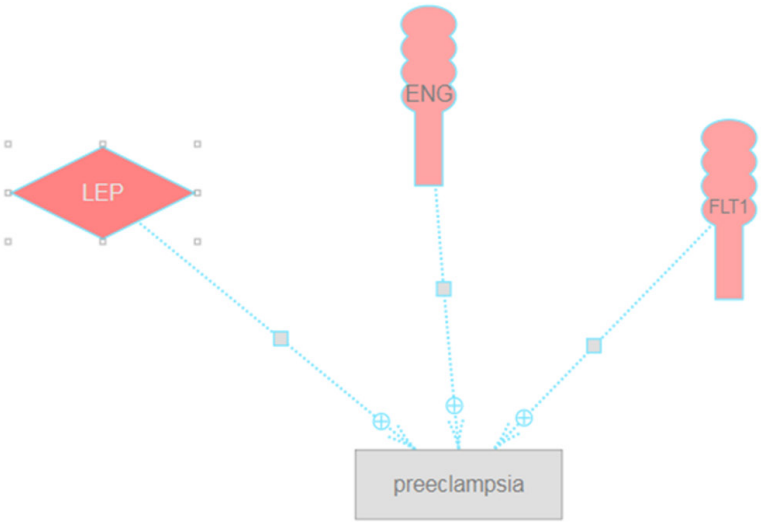


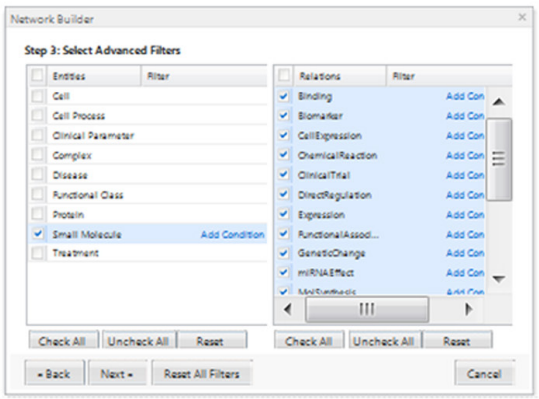
FLT1, ENG, & LEP are selected for further study on the basis of high literature metrics.



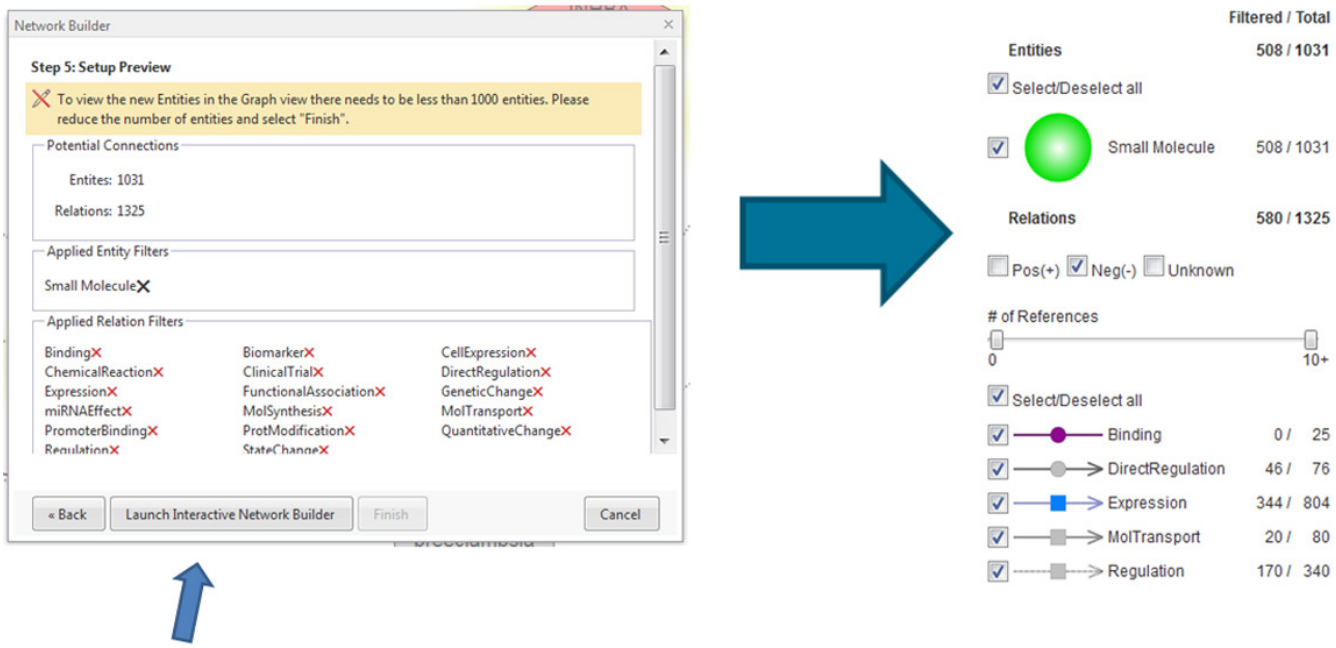
6.5 Small Molecule Inhibitors of Diseases and Genes

Identification of small molecule inhibitors of LEP, ENG, and FLT1.

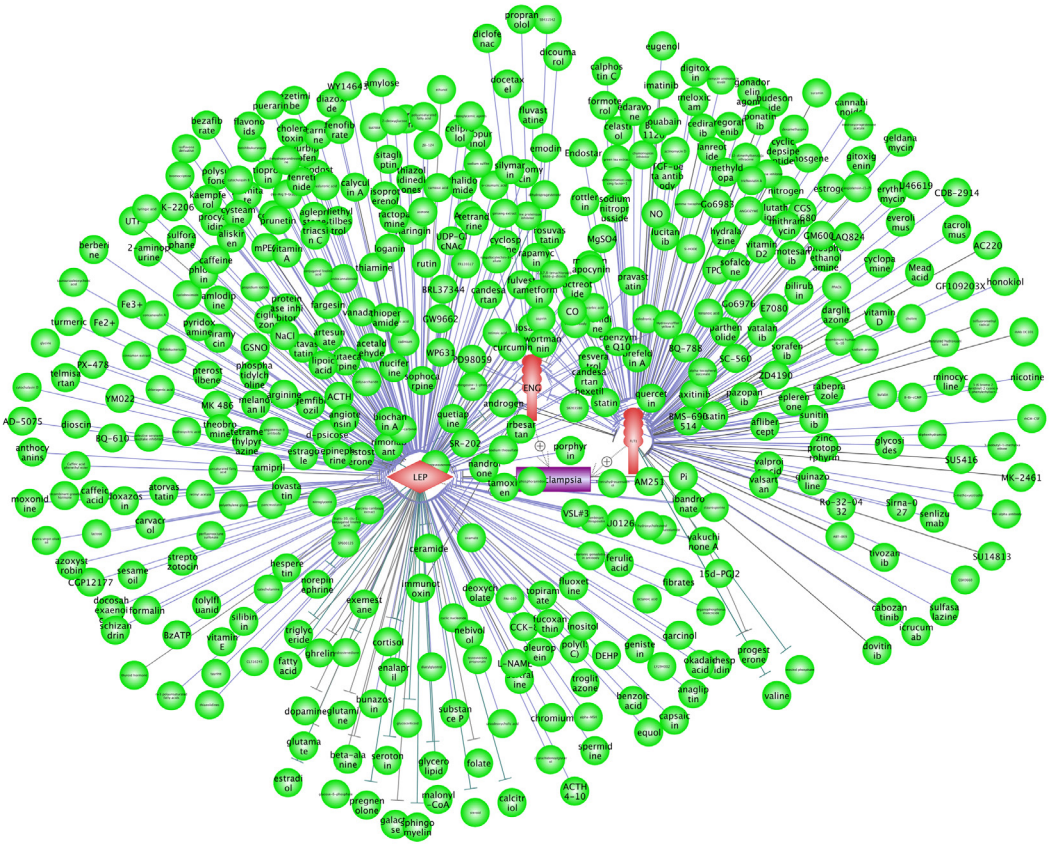




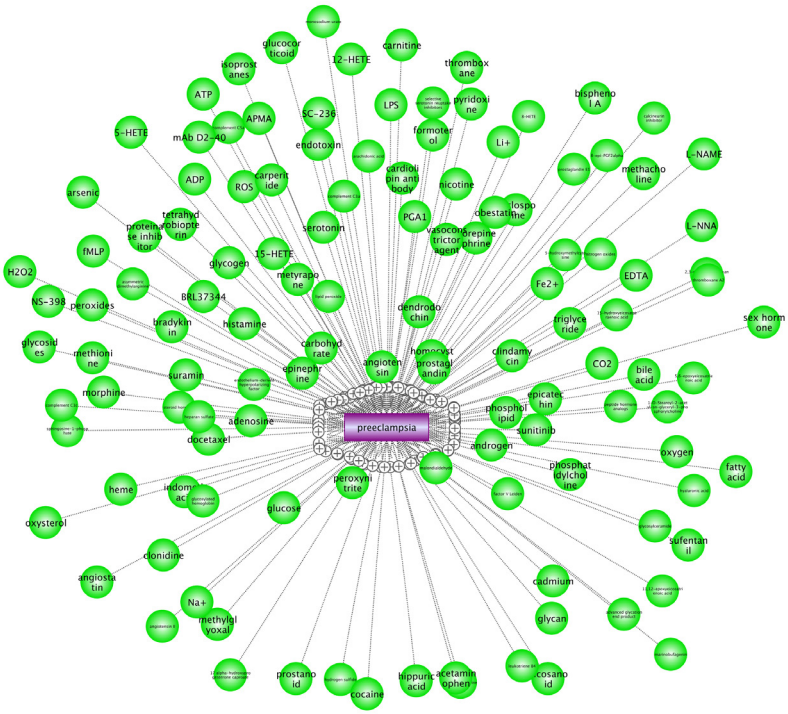
Use Network Builder: Advanced Expand Pathway Tool.



Launch Interactive Network Builder to filter for inhibitors only.



Small molecule inhibitors of pre-eclampsia target gene/proteins.

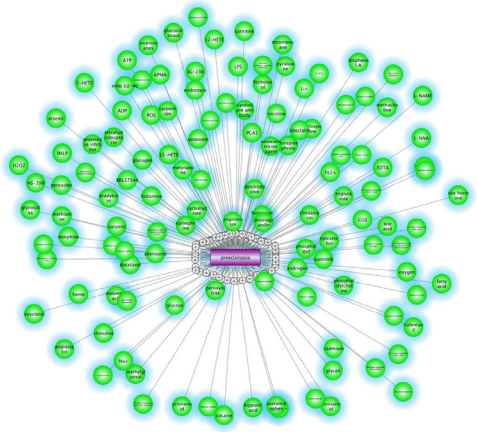


Small molecule inhibitors of pre-eclampsia disease as reported in the literature. (Hint: Use the same protocol you just used for small molecule inhibitors of ENG, LEP, and FLT1, and simply exchange those targets for the disease entity “pre-eclampsia.”)

Now find small molecule inhibitors of both pre-eclampsia and target genes/proteins (FLT1, LEP, and ENG). (Hint: This is essentially a combination of the two previous networks.)

There are several ways to combine and manipulate networks in Pathway Studio, let’s look at one of the simplest, shall we?

Step One: Select All and edit/copy the small molecule inhibitors of pre-eclampsia.

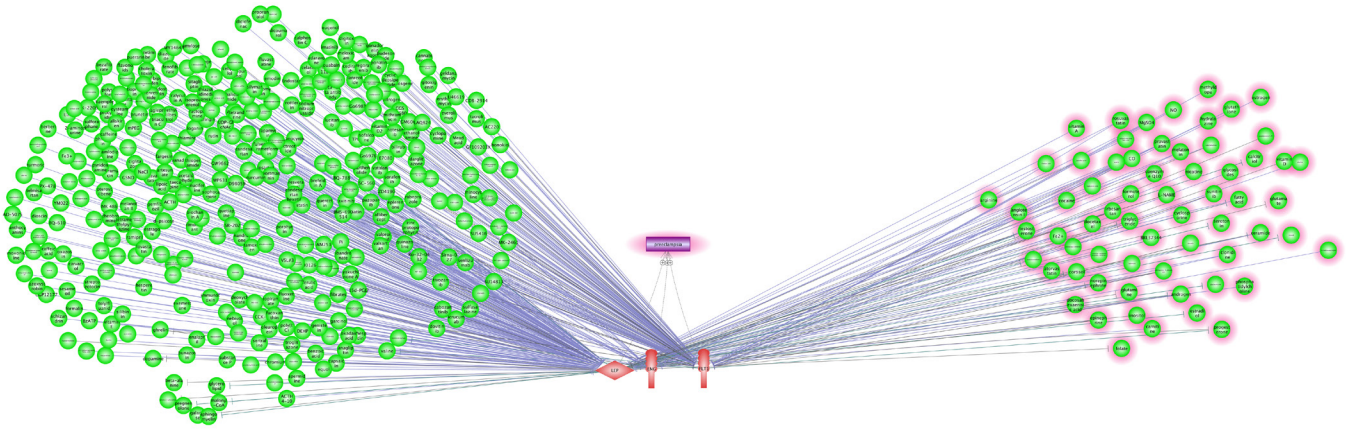


Small molecule inhibitors of pre-eclampsia.

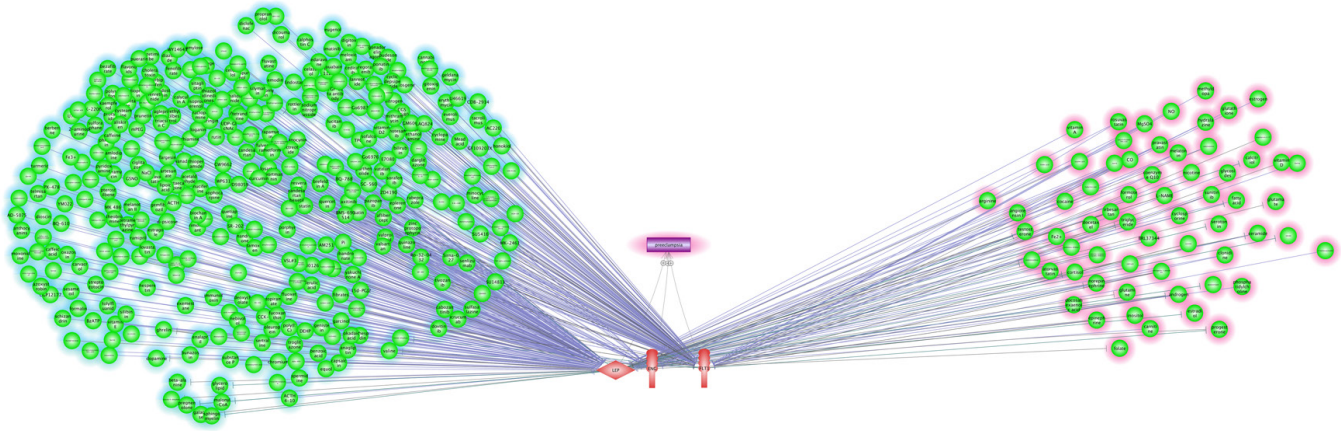
Small molecule inhibitors of pre-eclampsia target gene/proteins.

Step Two: Go to the small molecule inhibitors of FLT1, ENG, and LEP network, select "Clipboard Content," and highlight selections in red.

Step Three: Now just pull the selected objects as a group away from the unselected objects. (Hint: First make the whole display smaller using the slider in the upper right corner of the dashboard. Also, after you move the selection over then immediately move FLT1, ENG, and LEP to the right- hand side as well – this way they won't get lost in the final step below!)

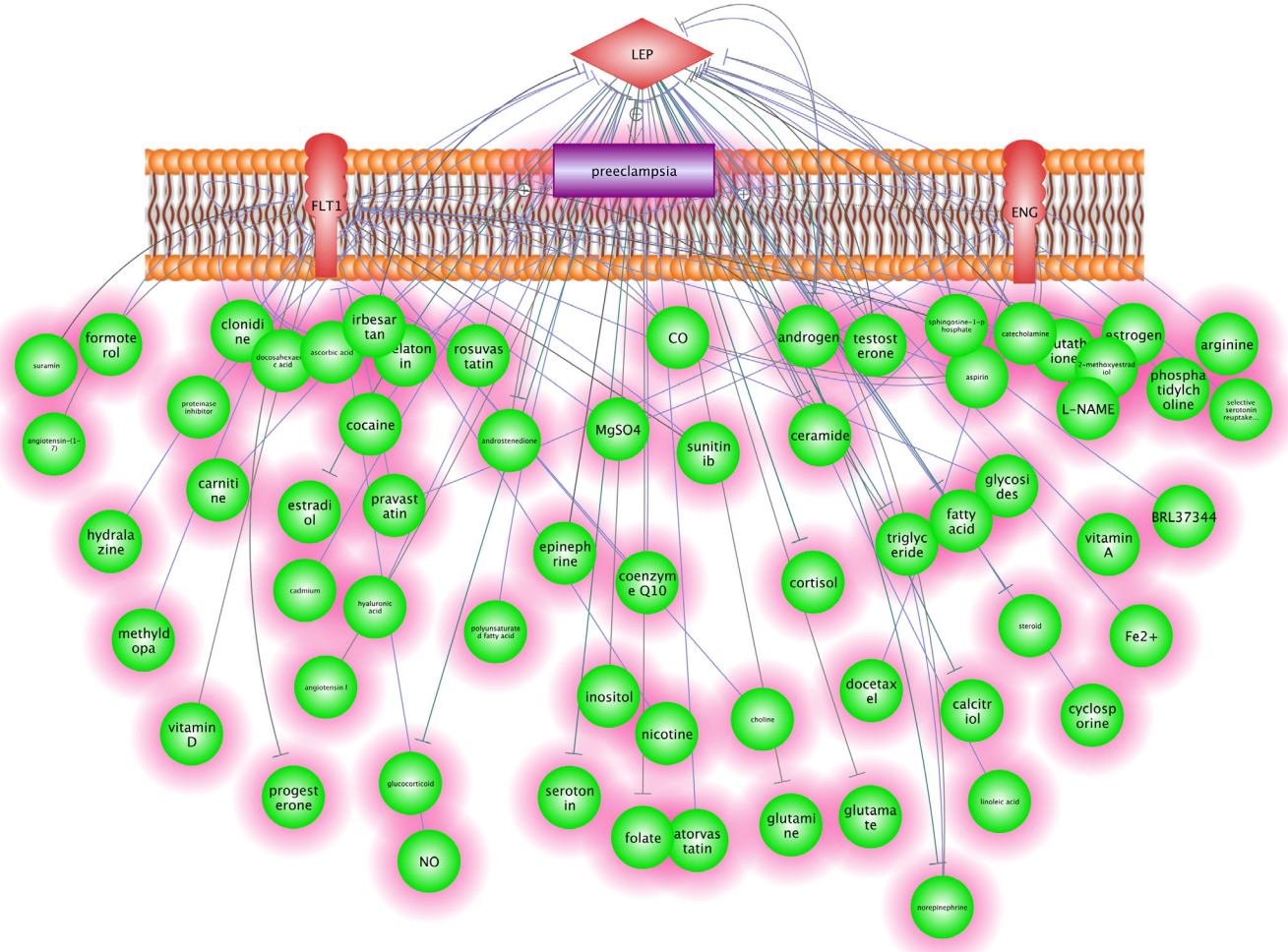


Step Four: Box the left-hand side, which will select everything on that side.



Step Five: Edit/remove selection.

Step Six: Rearrange the graphic to your taste. (Hint: This view was prepared by using "Layout by Localization, Plain Membrane" followed by some rearrangements formatted by hand to make a more pleasing visual! of ?)



Be patient with yourself!

It may take several attempts to get it right!

OK, are you ready for a little more?

6.6 Small Molecules in Clinical Trials

Just one last question!

How many of the small molecule inhibitors of FLT1, ENG, LEP, and pre-eclampsia are already in clinical trials?

Good question, right?

How to go about answering that, I wonder.

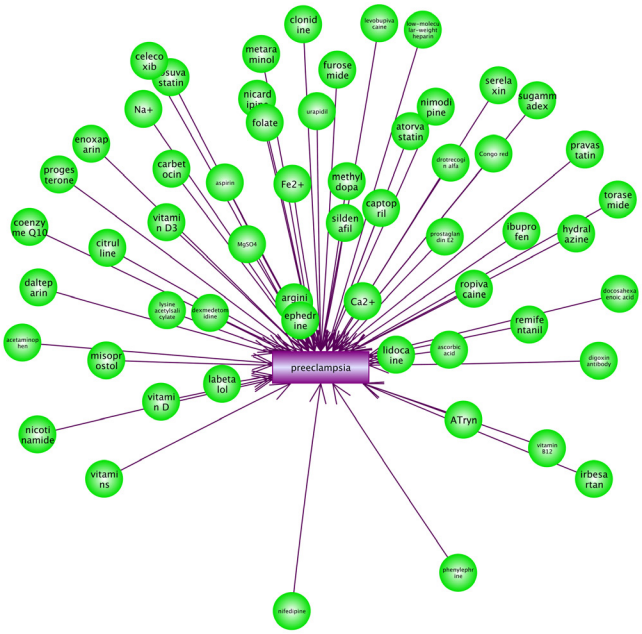
Hey, c'mon this should be easy by now!

All you need to do is map the small molecules in clinical trials to the disease of pre-eclampsia…

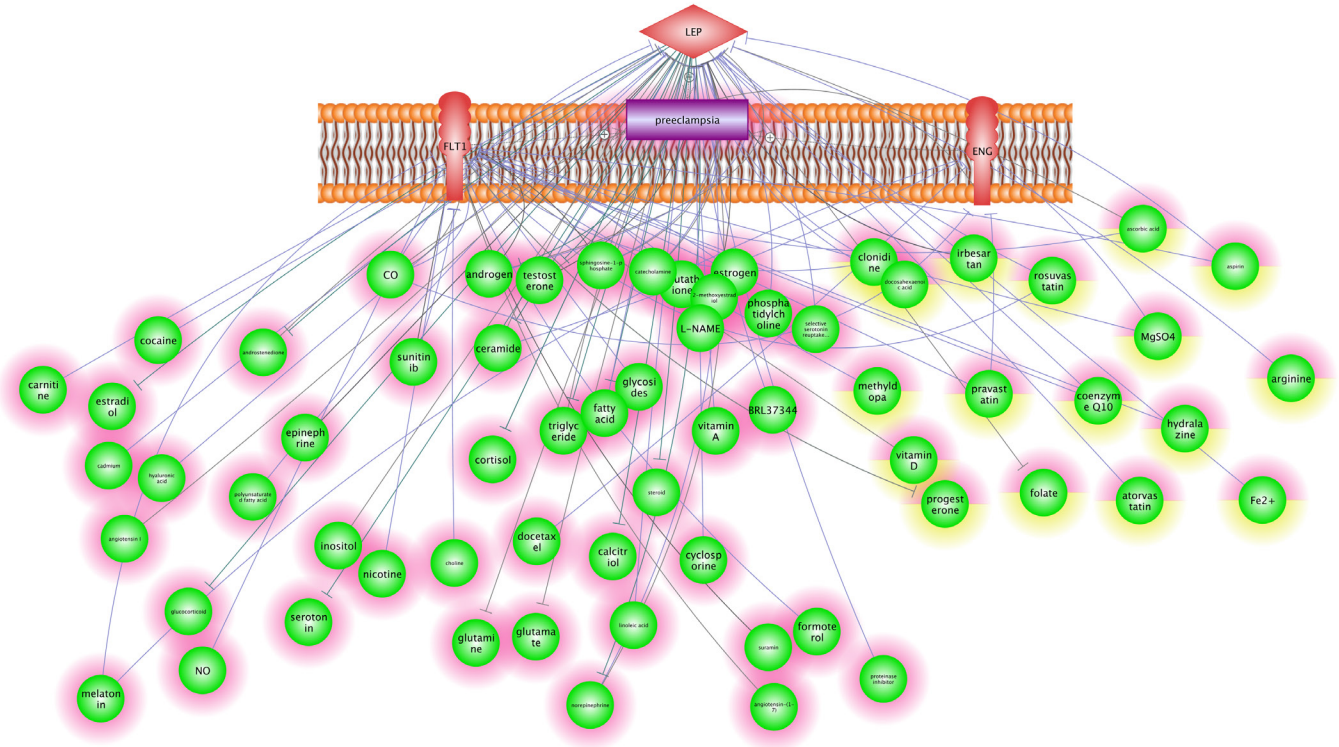
First, create a new pathway with the entity of pre-eclampsia disease.

Using the network builder, add small molecules as the entity and clinical trials as the relation, and you will get something that looks like this.

Now, select all, edit/copy, and then go to the pre-ecalmpsia network last constructed, and select "Clipboard Content."



Highlight selection by "Mix-in" Yellow (Note: This will in essence do a double highlight in yellow of small molecule inhibitors of FLT1, ENG, LEP, and pre-eclampsia already highlighted in red, which also are in clinical trials, and make it very easy to see these small molecules against the general background).



So before we go home...

What did we learn today?

- You learned data analysis of differential gene expression values.
- You learned filter probes by value (e.g. fold change and p-value).
- You ran an "Experiment Summary Analysis."
- You learned how to run an SNEA and understand what it is doing.
- You learned identification of small molecule inhibitors for selected target genes/proteins.
- You learned identification of small molecules involved in clinical trials for a specific disease.

Study Questions 6

Before answering the questions, read in data provided, use Name+Alias for probe ID, and use Benjamini-Hochberg for multiple comparisons correction:

1. Find the top 10 upregulated genes (by fold change) for the expression data in the module. (Hint: Count repeating genes one time only.)
2. Run an Experimental Summary and an SNEA (diseases) on the top upregulated genes. List the top five diseases (by p-value).
3. Run an SNEA (cell processes) on the top upregulated genes. List the top five cell processes (by p-value).
4. The "endocardial cushion" from Question 3 is involved in what major organ?
5. List the top five small molecule inhibitors of LEP (by # of references).
6. Run an SNEA (diseases) on all small molecule inhibitors of LEP. List the top five diseases (by p-value).
7. List the top five small molecule inhibitors of LEP in clinical trials for diabetes mellitus (by # of trials). (Hint: Click on SNEA subnetwork from Question 6, select diabetes mellitus in Pathway Viewer and "Add Relations between Selected and Unselected," and filter for clinical trials.)

For Connections Between Entities Not Directly Connected, How Do I...

Exercise 6.1: Can two proteins not directly connected, be connected through protein/small molecule (s)?

Identifies shortest path of molecular connection (physical interactions) between two proteins in the database through proteins/small molecule intermediates through selected relations type. (Can be extended to non-physical interactions by selecting additional relation types) Note: Regulation is the least specific relation type and should be excluded unless more specific relations do not produce results.

Step 1: Create New Pathway or within Pathway, select a protein(s)

Step 2: Select Add- > Network Builder

Step 3: Select Shortest Path for Pair of Entities. Advanced Shortest Path for Pair of Entities Tool

Step 4: Select Direction: All

Step 5: Entities: Select "Protein and Small Molecules" Relations: Select "Binding, ChemicalReaction, DirectrRgulation, Expression, miRNAEffect, MolSynthesis, MolTransport, PromoterBinding"

Continue to Network Builder Step 5 and click Launch Interactive Network Builder or Finish

Exercise 6.2: Can two proteins not directly connected, be connected through association to diseases/cell(s)?

Step 1: Create New Pathway or within Pathway, select a protein(s)

Step 2: Select Add- > Network Builder

Step 3: Select Shortest Path for Pair of Entities, Advanced Shortest Path for Pair of Entities Tool

Step 4: Select Direction: All

Step 5: Entities: Select "Cell Process + Disease" Relations: Select "Regulation"

Continue to Network Builder Step 5 and click Launch Interactive Network Builder or Finish

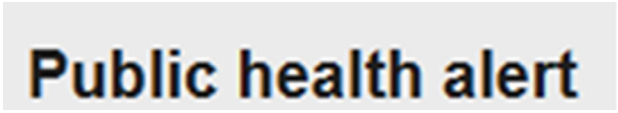
Module 7

Biomarker ID and Drug Repurposing

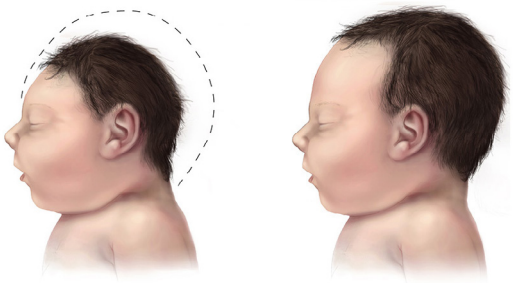
Contents

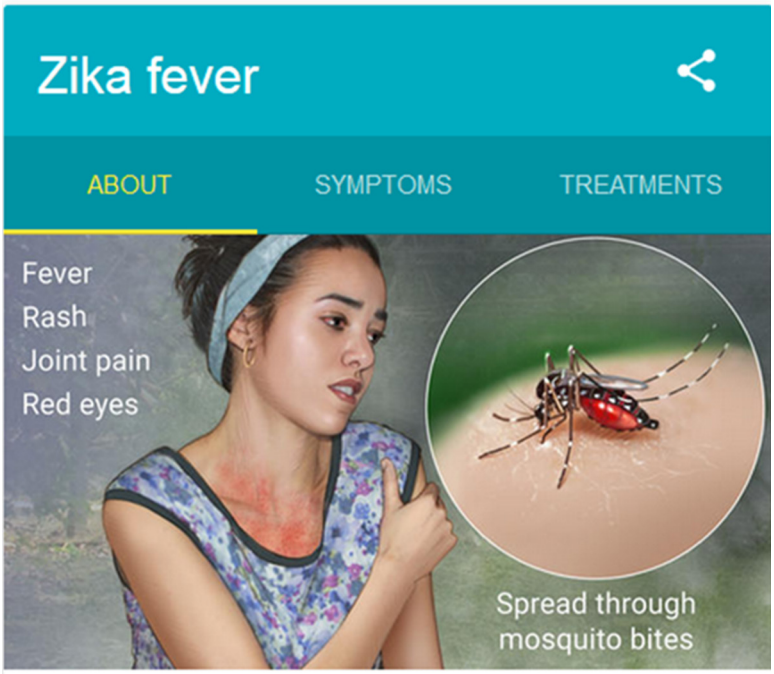
7.1 Introduction

Zika Virus Infection and Microcephaly



The CDC has issued an alert for travel to areas where Zika virus is spreading, which now includes Miami-Dade County, Florida, Puerto Rico, American Samoa, and the US Virgin Islands. Travelers who are pregnant or considering pregnancy should consult a doctor.





A disease caused by Zika virus that's spread through mosquito bites.

Zika Virus Infection and Microcephaly
What causes congenital microcephaly?

Causes of congenital microcephaly may include genetic conditions (e.g., chromosomal abnormalities), craniosynostosis, cerebral anoxia, or maternal exposures (e.g., alcohol, mercury, radiation, or severe malnutrition) during pregnancy. Maternal infections that have been associated with microcephaly include cytomegalovirus (CMV), herpes simplex virus, rubella virus, lymphocytic choriomeningitis virus (LCMV), varicella, Treponema pallidum (i.e., syphilis), and Toxoplasma gondii. Additional information about microcephaly is available on CDC's Microcephaly website.

What is the link between Zika virus and microcephaly?

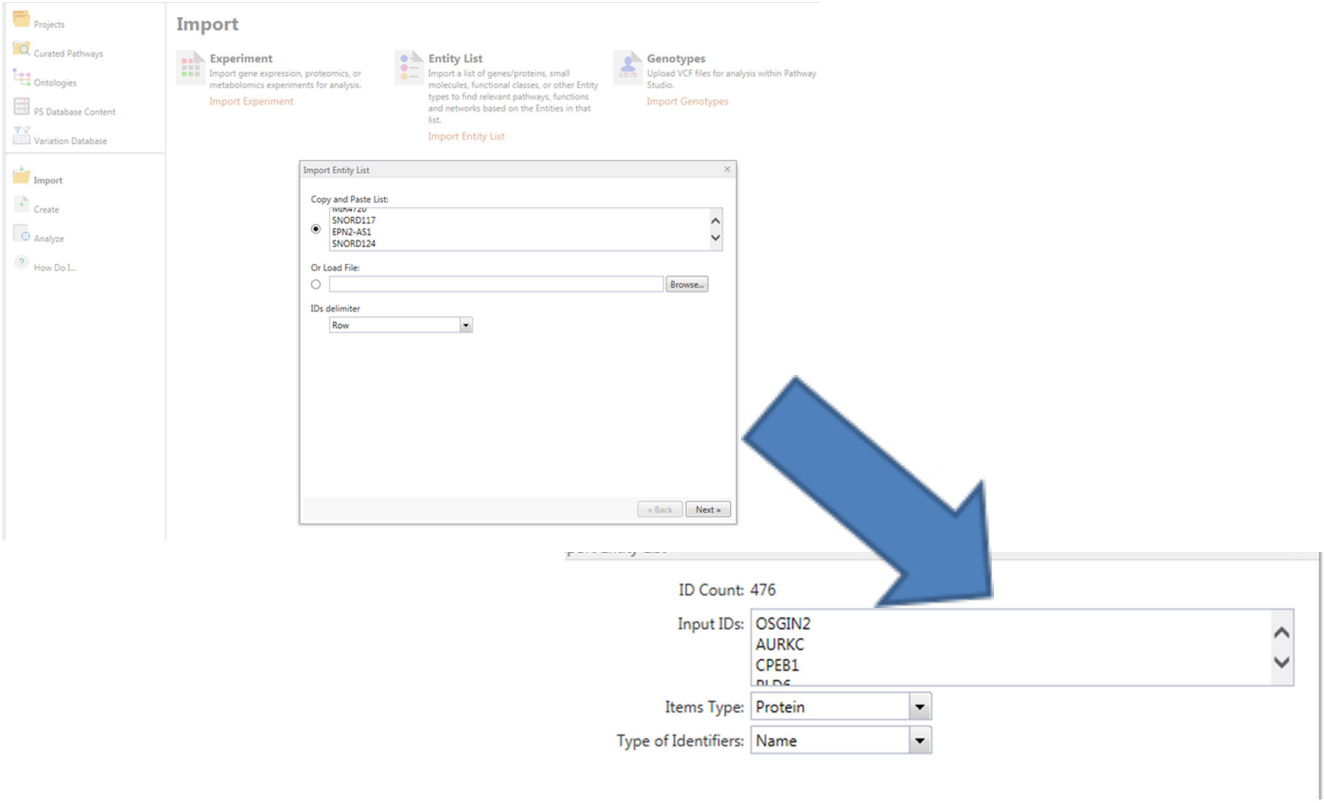
There is now scientific consensus that Zika virus is a cause of microcephaly - a congenital malformation with smaller than normal head size for age and sex. It has also been associated with other birth defects and neurologic conditions in children and adults.

7.2 Zika Virus Use Case Workflow

- Import genes up-regulated by Zika virus infection in neural progenitor cells.
- Explore potential disease associations of Zika virus up-regulated genes using the tools of causal reasoning in Pathway Studio.
- Map major regulator genes included in the Zika up-regulated genes, which are common to multiple disorders of glucose metabolism.
- Select model target gene/protein (EDN1) as a potential point of intervention.
- Map small molecule inhibitors of the model target gene/protein (EDN1) already in clinical trials for the treatment of insulin resistance.

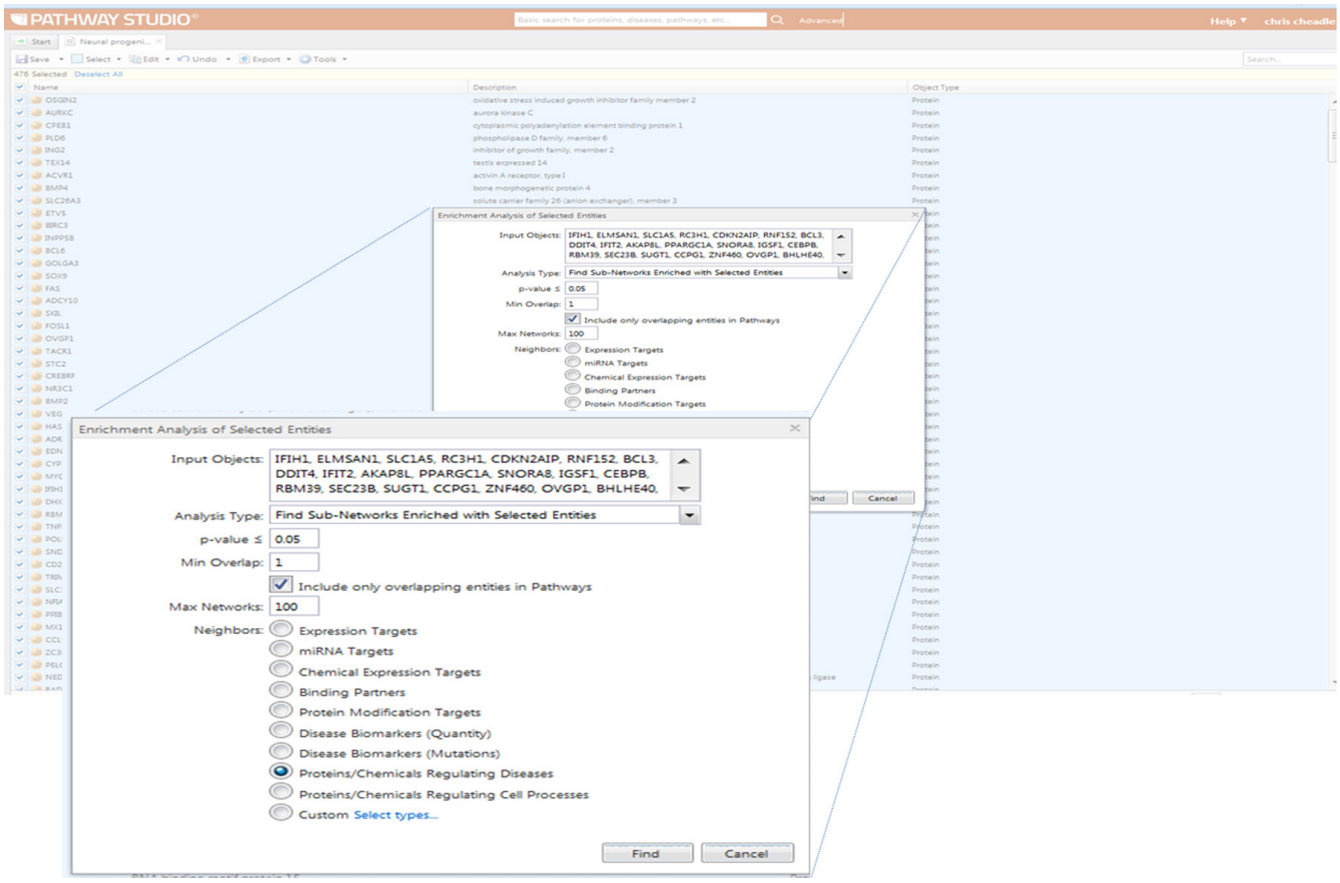
Dropbox link for Zika virus up-regulated gene list

Copy and paste gene list into Pathway Studio (Import Entity List).



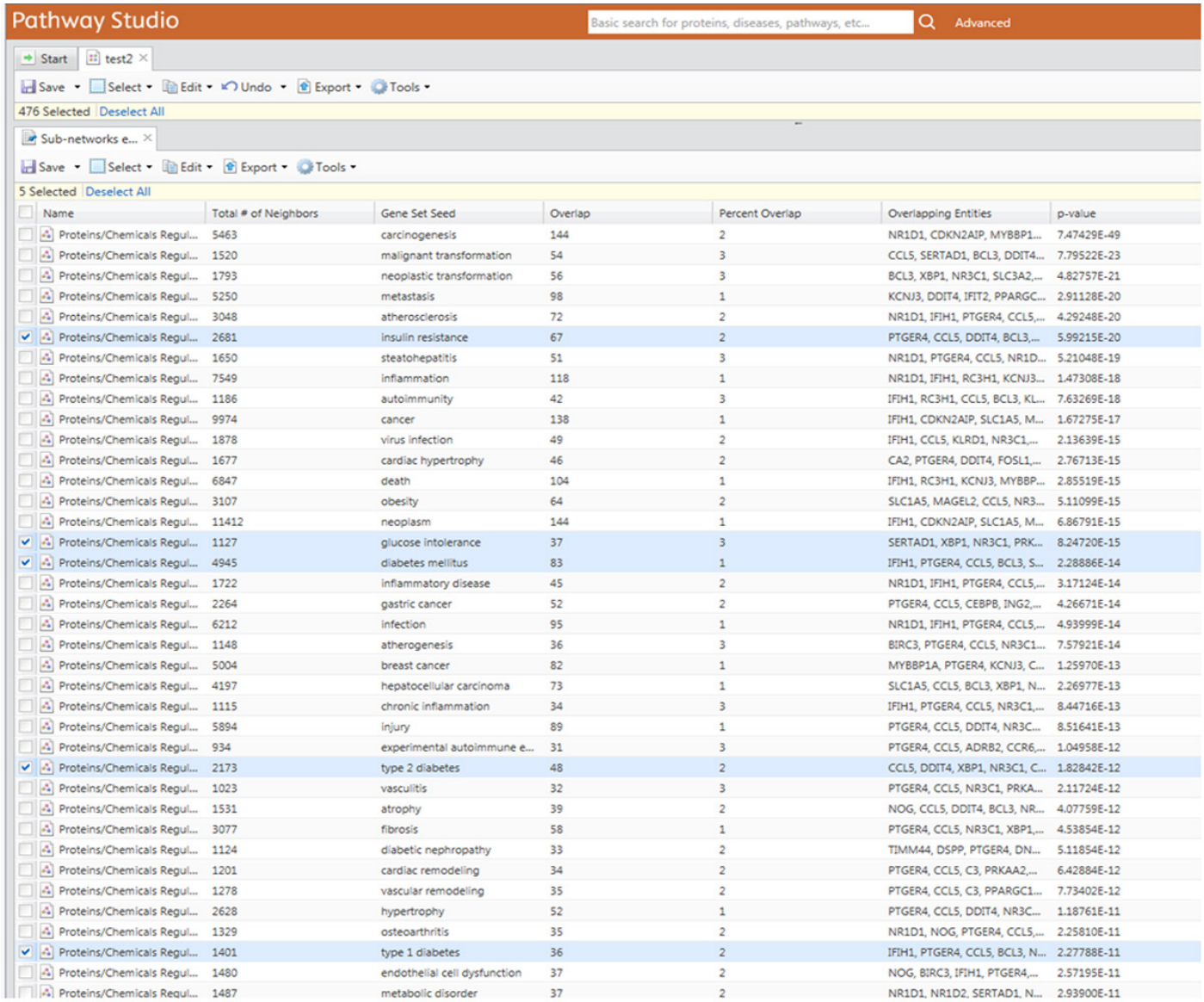
7.3 SNEA on Zika Virus Up-regulated Genes versus Diseases

Perform a Sub-Network Enrichment Analysis (SNEA) on Zika virus up-regulated genes versus diseases



7.4 Diseases Involving Disorders of Glucose Metabolism

Diseases involving disorders of glucose metabolism are enriched in Zika infection up-regulated gene



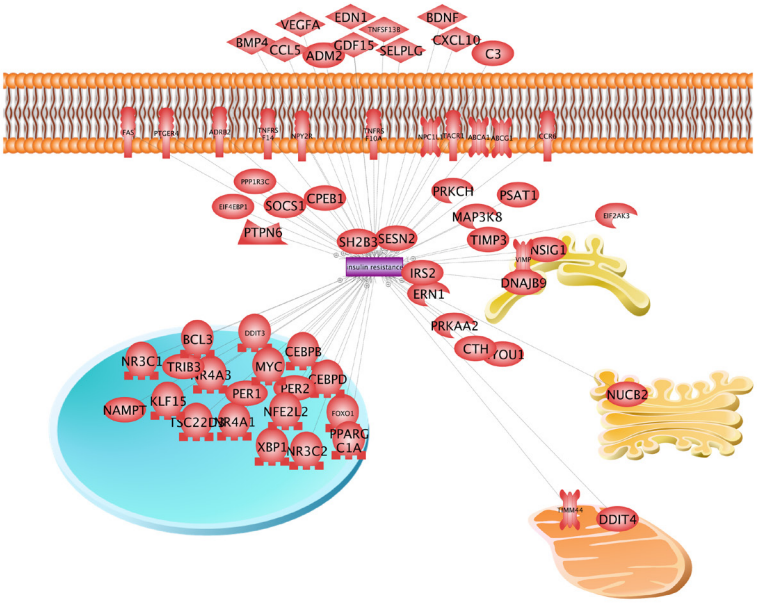
Name	Total # of Neighbors	Gene Set Seed	Overlap	Percent Overlap	Overlapping Entities	p-value
Proteins/Chemicals Regul...	5463	carcinogenesis	144	2	NR1D1, CDKN2AIP, MYBBP1...	7.47429E-49
Proteins/Chemicals Regul...	1520	malignant transformation	54	3	CCL5, SERTAD1, BCL3, DDIT4...	7.79522E-23
Proteins/Chemicals Regul...	1793	neoplastic transformation	56	3	BCL3, XBP1, NR3C1, SLC3A2,...	4.82757E-21
Proteins/Chemicals Regul...	5250	metastasis	98	1	KCNJ3, DDIT4, IFIT2, PPARGC...	2.91128E-20
Proteins/Chemicals Regul...	3048	atherosclerosis	72	2	NR1D1, IFIH1, PTGER4, CCL5,...	4.29248E-20
Proteins/Chemicals Regul...	2681	insulin resistance	67	2	PTGER4, CCL5, DDIT4, BCL3,...	5.99215E-20
Proteins/Chemicals Regul...	1650	steatohepatitis	51	3	NR1D1, PTGER4, CCL5, NR1D...	5.21048E-19
Proteins/Chemicals Regul...	7549	inflammation	118	1	NR1D1, IFIH1, RC3H1, KCNJ3...	1.47308E-18
Proteins/Chemicals Regul...	1186	autoimmunity	42	3	IFIH1, RC3H1, CCL5, BCL3, KL...	7.63269E-18
Proteins/Chemicals Regul...	9974	cancer	138	1	IFIH1, CDKN2AIP, SLC1A5, M...	1.67275E-17
Proteins/Chemicals Regul...	1878	virus infection	49	2	IFIH1, CCL5, KLRD1, NR3C1,...	2.13639E-15
Proteins/Chemicals Regul...	1677	cardiac hypertrophy	46	2	CA2, PTGER4, DDIT4, FOSL1,...	2.76713E-15
Proteins/Chemicals Regul...	6847	death	104	1	IFIH1, RC3H1, KCNJ3, MYBBP...	2.85519E-15
Proteins/Chemicals Regul...	3107	obesity	64	2	SLC1A5, MAGEL2, CCL5, NR3...	5.11099E-15
Proteins/Chemicals Regul...	11412	neoplasm	144	1	IFIH1, CDKN2AIP, SLC1A5, M...	6.86791E-15
Proteins/Chemicals Regul...	1127	glucose intolerance	37	3	SERTAD1, XBP1, NR3C1, PRK...	8.24720E-15
Proteins/Chemicals Regul...	4945	diabetes mellitus	83	1	IFIH1, PTGER4, CCL5, BCL3, S...	2.28886E-14
Proteins/Chemicals Regul...	1722	inflammatory disease	45	2	NR1D1, IFIH1, PTGER4, CCL5,...	3.17124E-14
Proteins/Chemicals Regul...	2264	gastric cancer	52	2	PTGER4, CCL5, CEBPB, ING2,...	4.26671E-14
Proteins/Chemicals Regul...	6212	infection	95	1	NR1D1, IFIH1, PTGER4, CCL5,...	4.93999E-14
Proteins/Chemicals Regul...	1148	atherogenesis	36	3	BIRC3, PTGER4, CCL5, NR3C1...	7.57921E-14
Proteins/Chemicals Regul...	5004	breast cancer	82	1	MYBBP1A, PTGER4, KCNJ3, C...	1.25970E-13
Proteins/Chemicals Regul...	4197	hepatocellular carcinoma	73	1	SLC1A5, CCL5, BCL3, XBP1, N...	2.26977E-13
Proteins/Chemicals Regul...	1115	chronic inflammation	34	3	IFIH1, PTGER4, CCL5, NR3C1,...	8.44716E-13
Proteins/Chemicals Regul...	5894	injury	89	1	PTGER4, CCL5, DDIT4, NR3C...	8.51641E-13
Proteins/Chemicals Regul...	934	experimental autoimmune e...	31	3	PTGER4, CCL5, ADRB2, CCR6,...	1.04958E-12
Proteins/Chemicals Regul...	2173	type 2 diabetes	48	2	CCL5, DDIT4, XBP1, NR3C1, C...	1.82842E-12
Proteins/Chemicals Regul...	1023	vasculitis	32	3	PTGER4, CCL5, NR3C1, PRKA...	2.11724E-12
Proteins/Chemicals Regul...	1531	atrophy	39	2	NOG, CCL5, DDIT4, BCL3, NR...	4.07759E-12
Proteins/Chemicals Regul...	3077	fibrosis	58	1	PTGER4, CCL5, NR3C1, XBP1,...	4.53854E-12
Proteins/Chemicals Regul...	1124	diabetic nephropathy	33	2	TIMM44, DSPP, PTGER4, DN...	5.11854E-12
Proteins/Chemicals Regul...	1201	cardiac remodeling	34	2	PTGER4, CCL5, C3, PRKAA2,...	6.42884E-12
Proteins/Chemicals Regul...	1278	vascular remodeling	35	2	PTGER4, CCL5, C3, PPARGC1...	7.73402E-12
Proteins/Chemicals Regul...	2628	hypertrophy	52	1	PTGER4, CCL5, DDIT4, NR3C...	1.18761E-11
Proteins/Chemicals Regul...	1329	osteoarthritis	35	2	NR1D1, NOG, PTGER4, CCL5,...	2.25810E-11
Proteins/Chemicals Regul...	1401	type 1 diabetes	36	2	IFIH1, PTGER4, CCL5, BCL3, N...	2.27788E-11
Proteins/Chemicals Regul...	1480	endothelial cell dysfunction	37	2	NOG, BIRC3, IFIH1, PTGER4,...	2.57195E-11
Proteins/Chemicals Regul...	1487	metabolic disorder	37	2	NR1D1, NR1D2, SERTAD1, N...	2.93900E-11

Save all five gene networks to your project folder.

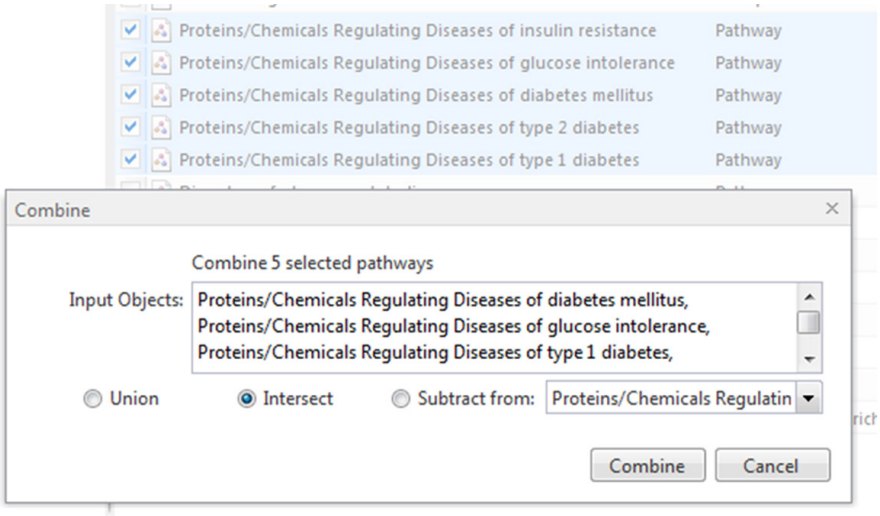
Proteins/Chemicals Regulating Diseases of insulin resistance	Pathway	68
Proteins/Chemicals Regulating Diseases of glucose intolerance	Pathway	38
Proteins/Chemicals Regulating Diseases of diabetes mellitus	Pathway	83
Proteins/Chemicals Regulating Diseases of type 2 diabetes	Pathway	49
Proteins/Chemicals Regulating Diseases of type 1 diabetes	Pathway	37

7.5 Identifying Major Regulators

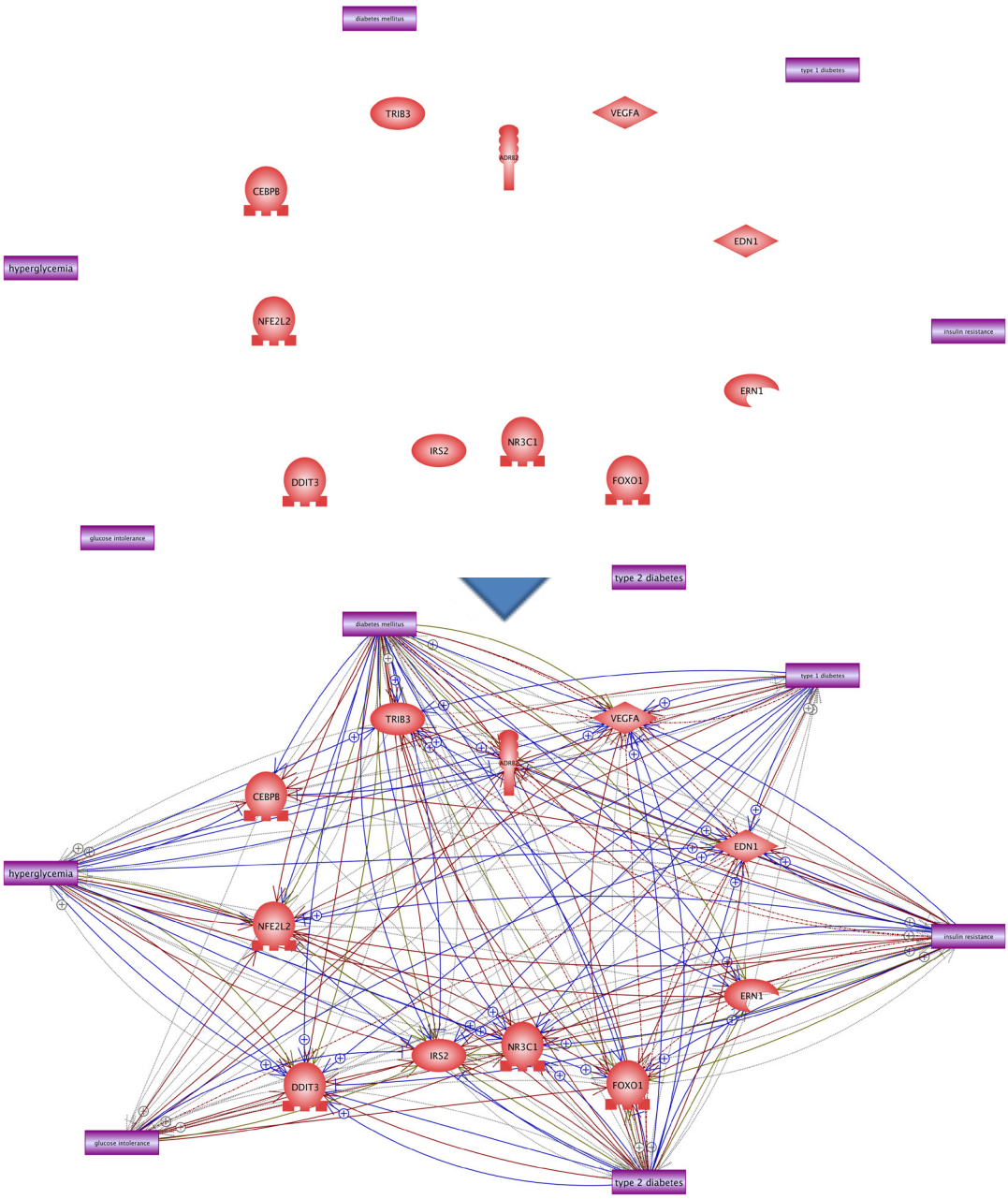
All Zika infection up-regulated genes/ proteins related to insulin resistance. (Hint: use "Layout by Localization, Plain Membrane")



Combine all five disorders of glucose metabolism and take their intersection:

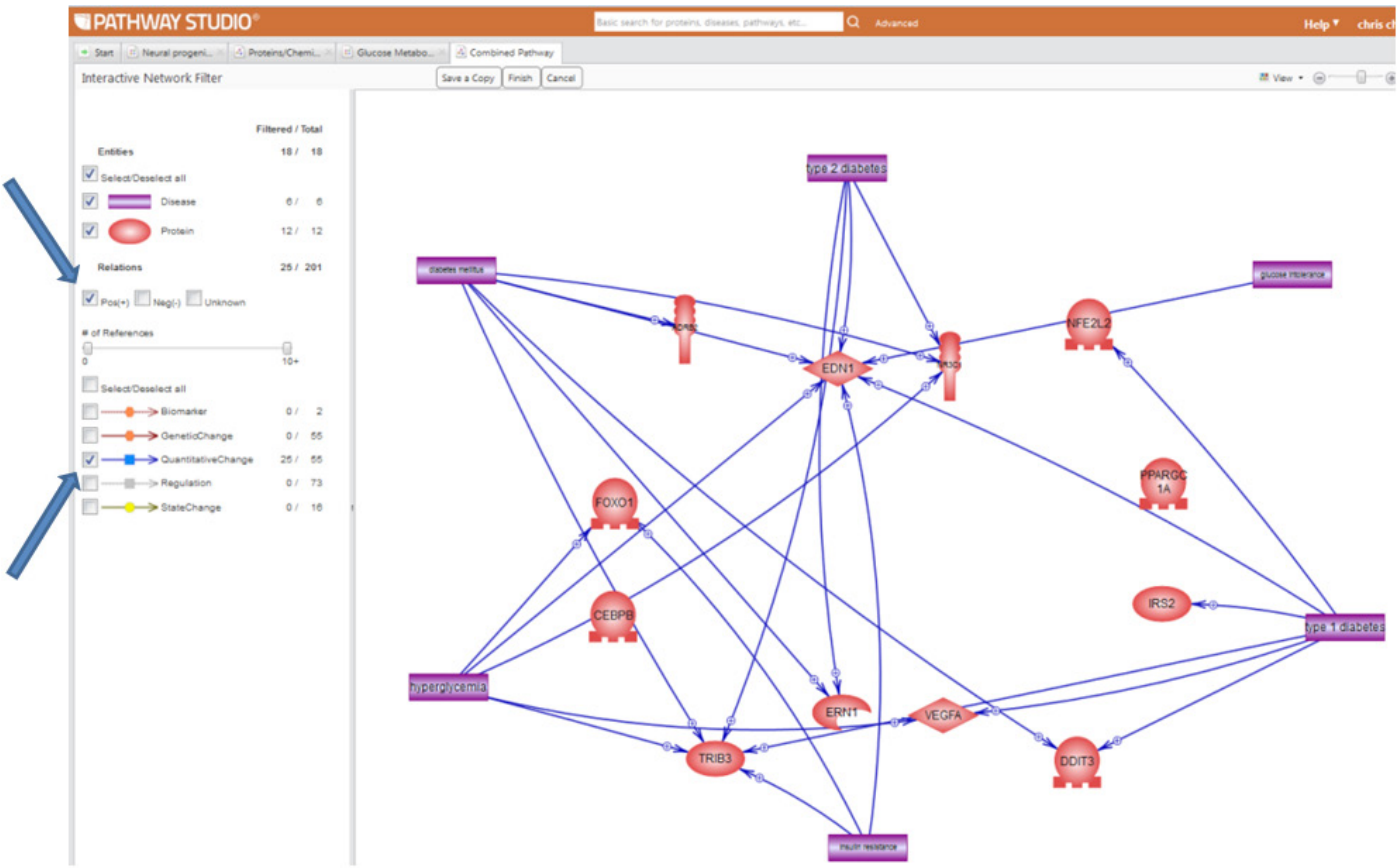


Copy and paste all five diseases of glucose metabolism back into the gene/ protein intersection and add "Relations between Selected and Unselected":



Zika up-regulated genes related to all six glucose metabolic disorders. Use the Interactive Network Filter and select gene/protein activators only.

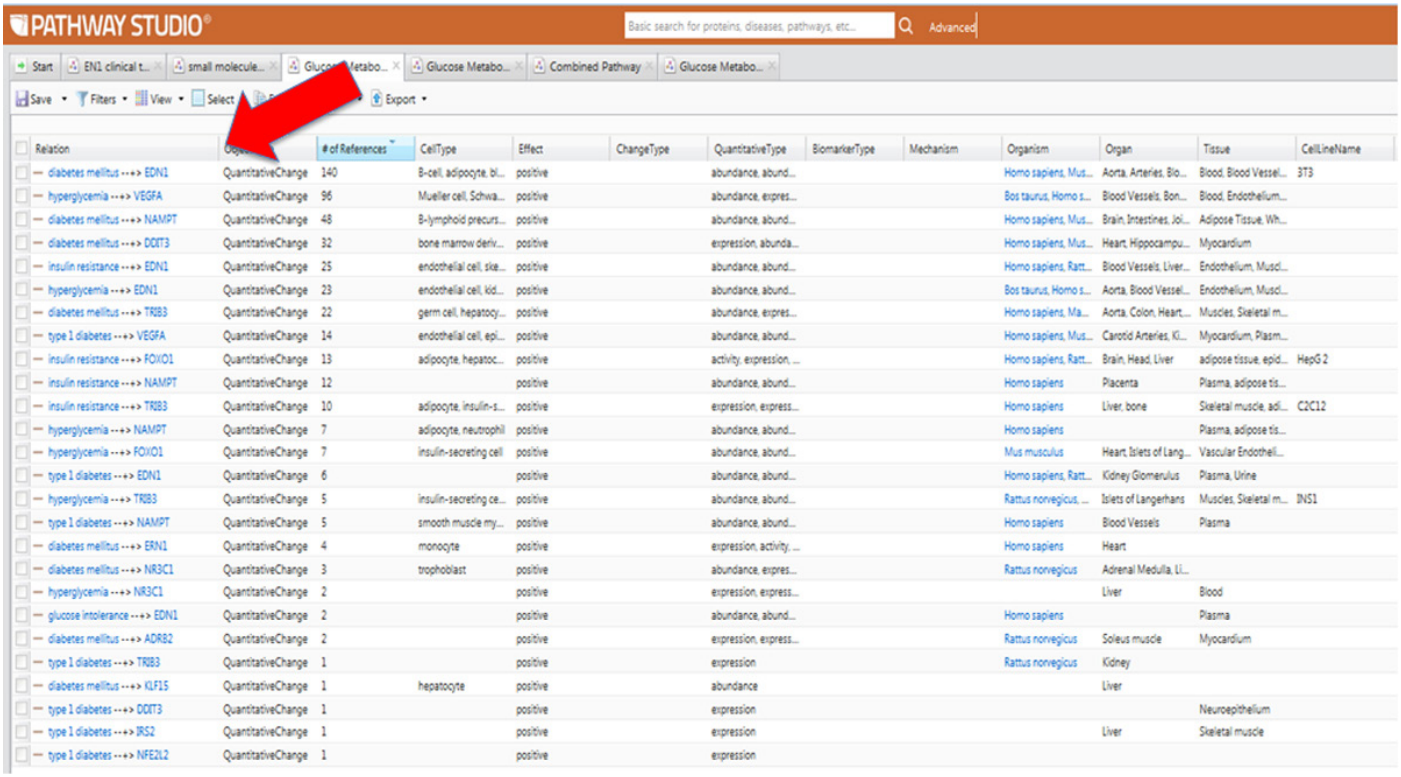
7.6 Using the Interactive Network Filter for Better Specificity



View the Relation Table View.

It's an easy bear!

Target identification: Endothelin 1 (EDN1)

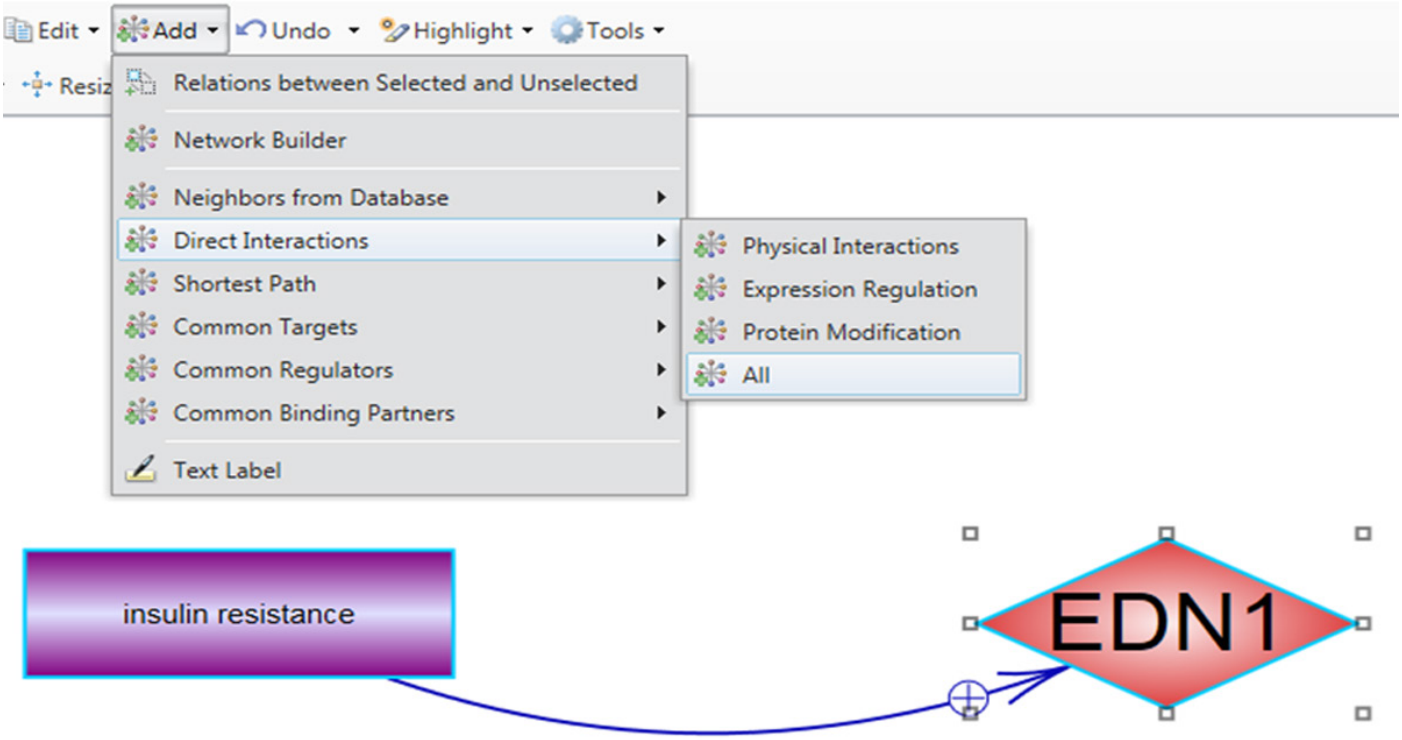


Relation	# of References	CellType	Effect	ChangeType	QuantitativeType	BiomarkerType	Mechanism	Organism	Organ	Tissue	CellLineName
diabetes mellitus ---> EDN1	140	B-cell, adipocyte, bl...	positive		abundance, abund...			Homo sapiens, Mus...	Aorta, Arteries, Blo...	Blood, Blood Vessel...	3T3
hyperglycemia ---> VEGFA	96	Mueller cell, Schwa...	positive		abundance, expres...			Bos taurus, Homo s...	Blood Vessels, Bon...	Blood, Endothelium...	
diabetes mellitus ---> NAMPT	48	B-lymphoid precurs...	positive		abundance, abund...			Homo sapiens, Mus...	Brain, Intestines, Joi...	Adipose Tissue, Wh...	
diabetes mellitus ---> DDIT3	32	bone marrow deriv...	positive		expression, abunda...			Homo sapiens, Mus...	Heart, Hippocampu...	Myocardium	
insulin resistance ---> EDN1	25	endothelial cell, ske...	positive		abundance, abund...			Homo sapiens, Ratt...	Blood Vessels, Liver...	Endothelium, Muscl...	
hyperglycemia ---> EDN1	23	endothelial cell, kid...	positive		abundance, abund...			Bos taurus, Homo s...	Aorta, Blood Vessel...	Endothelium, Muscl...	
diabetes mellitus ---> TRIB3	22	germ cell, hepatocy...	positive		abundance, expres...			Homo sapiens, Ma...	Aorta, Colon, Heart,...	Muscles, Skeletal m...	
type 1 diabetes ---> VEGFA	14	endothelial cell, epi...	positive		abundance, abund...			Homo sapiens, Mus...	Carotid Arteries, Ki...	Myocardium, Plasm...	
insulin resistance ---> FOXO1	13	adipocyte, hepatoc...	positive		activity, expression, ...			Homo sapiens, Ratt...	Brain, Head, Liver	adipose tissue, epid...	HepG 2
insulin resistance ---> NAMPT	12		positive		abundance, abund...			Homo sapiens	Placenta	Plasma, adipose tis...	
insulin resistance ---> TRIB3	10	adipocyte, insulin-s...	positive		expression, express...			Homo sapiens	Liver, bone	Skeletal muscle, adi...	C2C12
hyperglycemia ---> NAMPT	7	adipocyte, neutrophil	positive		abundance, abund...			Homo sapiens		Plasma, adipose tis...	
hyperglycemia ---> FOXO1	7	insulin-secreting cell	positive		abundance, abund...			Mus musculus	Heart, Islets of Lang...	Vascular Endothel...	
type 1 diabetes ---> EDN1	6		positive		abundance, abund...			Homo sapiens, Ratt...	Kidney Glomerulus	Plasma, Urine	
hyperglycemia ---> TRIB3	5	insulin-secreting ce...	positive		abundance, abund...			Rattus norvegicus, ...	Islets of Langerhans	Muscles, Skeletal m...	INS1
type 1 diabetes ---> NAMPT	5	smooth muscle my...	positive		abundance, abund...			Homo sapiens	Blood Vessels	Plasma	
diabetes mellitus ---> ERN1	4	monocyte	positive		expression, activity, ...			Homo sapiens	Heart		
diabetes mellitus ---> NR3C1	3	trophoblast	positive		abundance, expres...			Rattus norvegicus	Adrenal Medulla, Li...		
hyperglycemia ---> NR3C1	2		positive		expression, express...				Liver	Blood	
glucose intolerance ---> EDN1	2		positive		abundance, abund...			Homo sapiens		Plasma	
diabetes mellitus ---> ADRB2	2		positive		expression, express...			Rattus norvegicus	Soleus muscle	Myocardium	
type 1 diabetes ---> TRIB3	1		positive		expression			Rattus norvegicus	Kidney		
diabetes mellitus ---> KLF15	1	hepatocyte	positive		abundance				Liver		
type 1 diabetes ---> DDIT3	1		positive		expression					Neuroepithelium	
type 1 diabetes ---> IRS2	1		positive		expression				Liver	Skeletal muscle	
type 1 diabetes ---> NFE2L2	1		positive		expression						

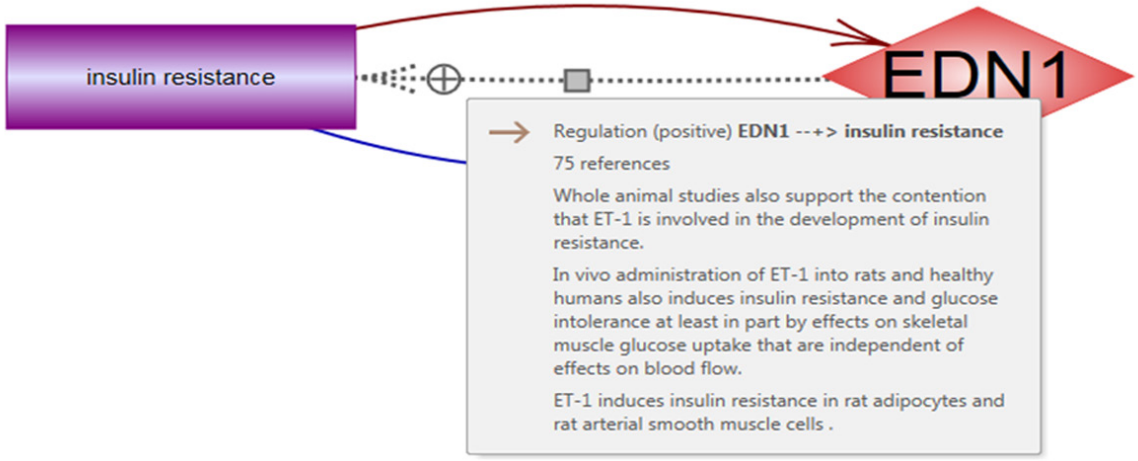
Now, focus on EDN1 and insulin resistance.

7.7 Identifying Major Therapeutic Point of Intervention

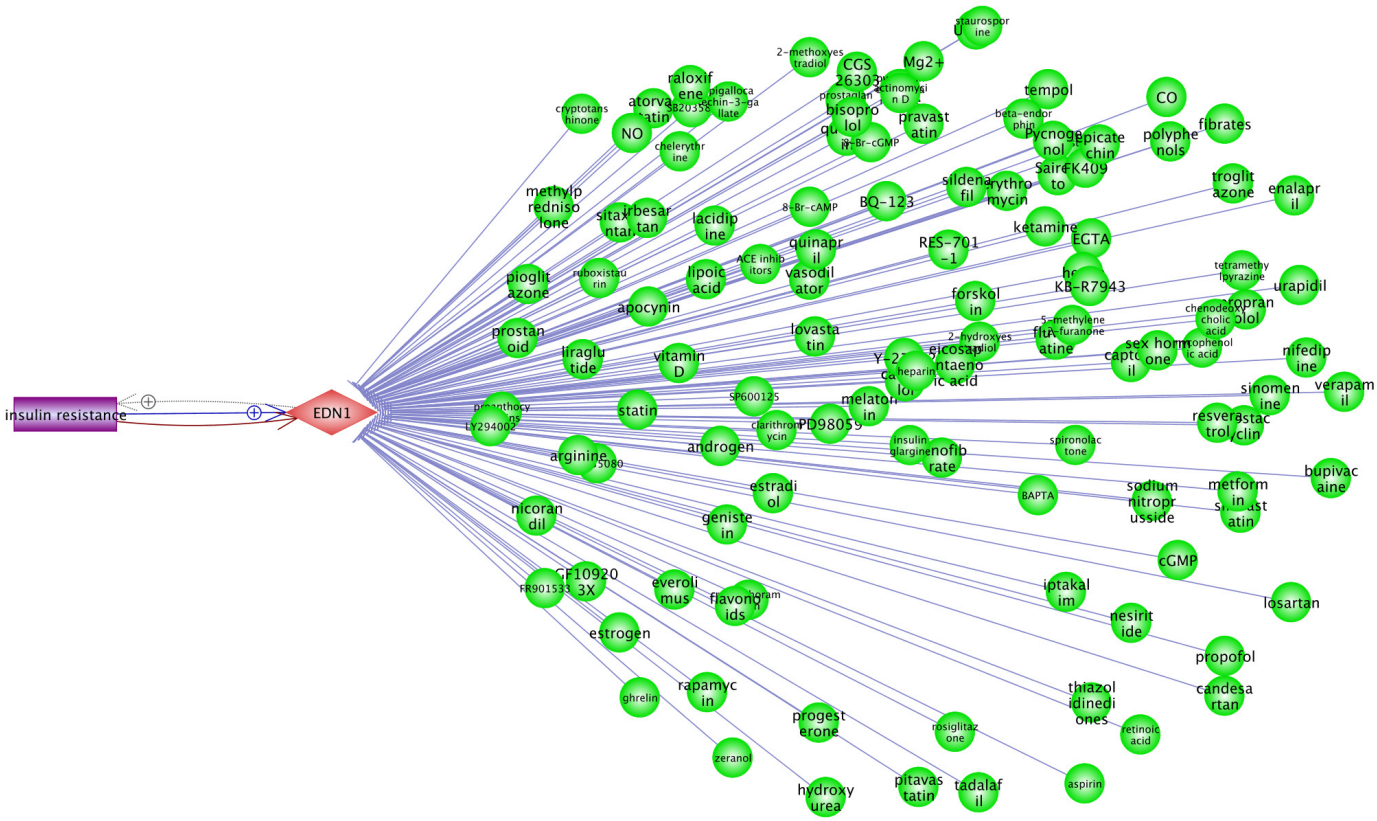
Question: Insulin resistance results in an increase in EDN1 levels, but is there any evidence that decreasing EDN1 levels is therapeutic?



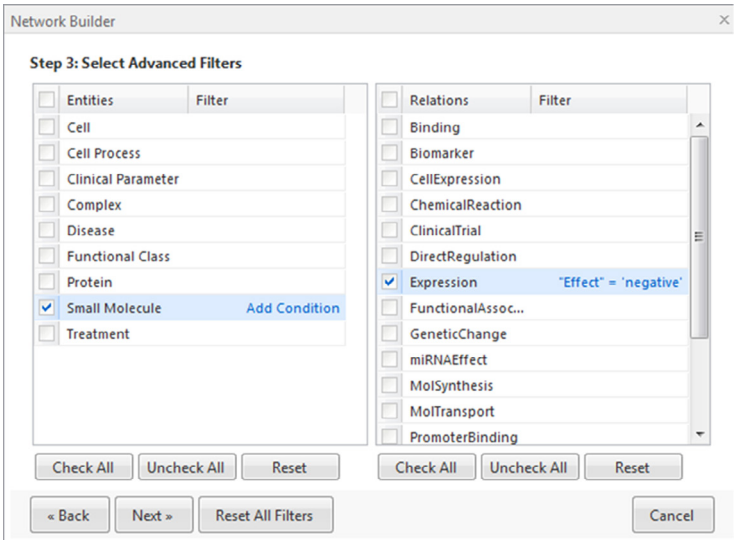
Answer: Evidence from animal models is that EDN1 is causal for insulin resistance and, therefore, might be a good candidate for therapeutic intervention.



Map small molecule inhibitors of EDN1:

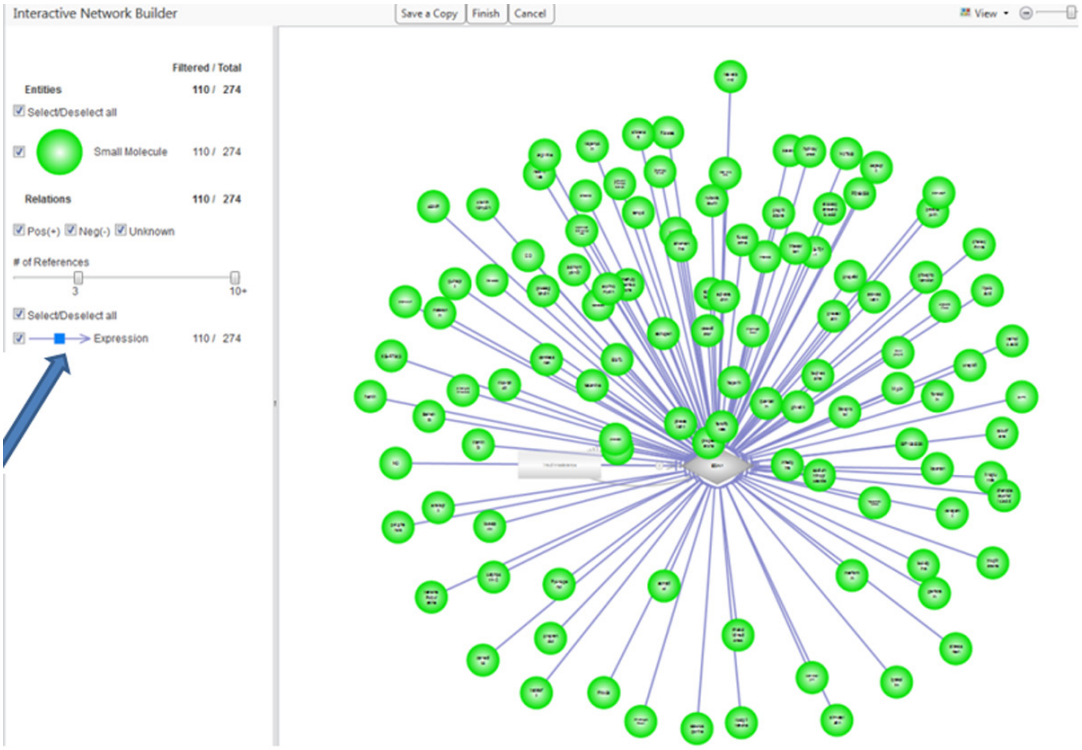


- Select EDN1.
- Add, Network Builder.
- Use Advanced Expand Pathway Tool.
- Select “Upstream.”*
- Use Advanced Filters.
- Select Small Molecule as Entity.
- Select Expression as Relation.
- Add Condition to Expression.
- Select Effect = negative.
- Click Next, then Entire Database.
- Click Next, then Launch Interactive Network Builder



* Select “Upstream” in the Expand Pathway Tool because you are looking for small molecules that act upon EDN1. In other words, they inhibit EDN1 gene expression or protein abundance. If you had selected “Downstream,” you would have been asking for small molecules that EDN1 acts upon instead.

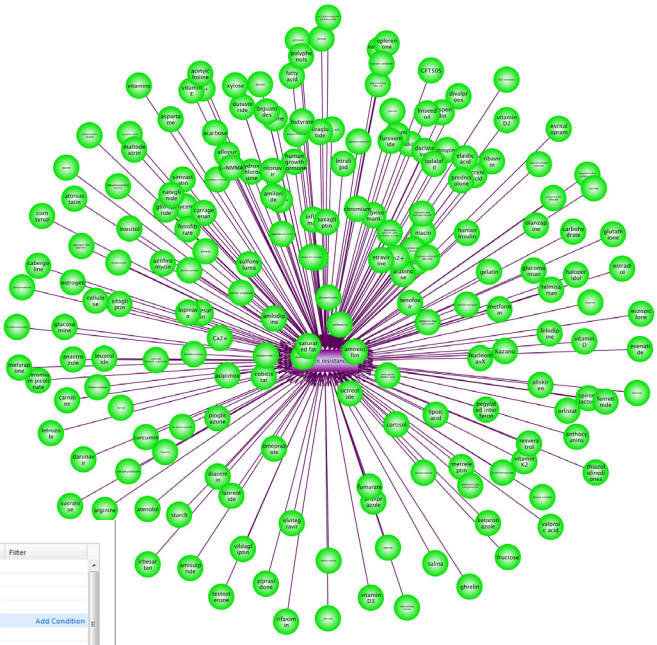
Use the Interactive Network Filter. Select References to a minimum of 4.

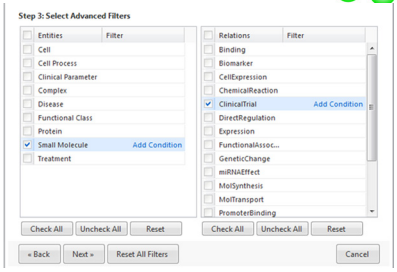


7.8 Mapping Candidate Drugs Both In and Outside of Current Clinical Trials

Map small molecules in clinical trials for the treatment of insulin resistance.

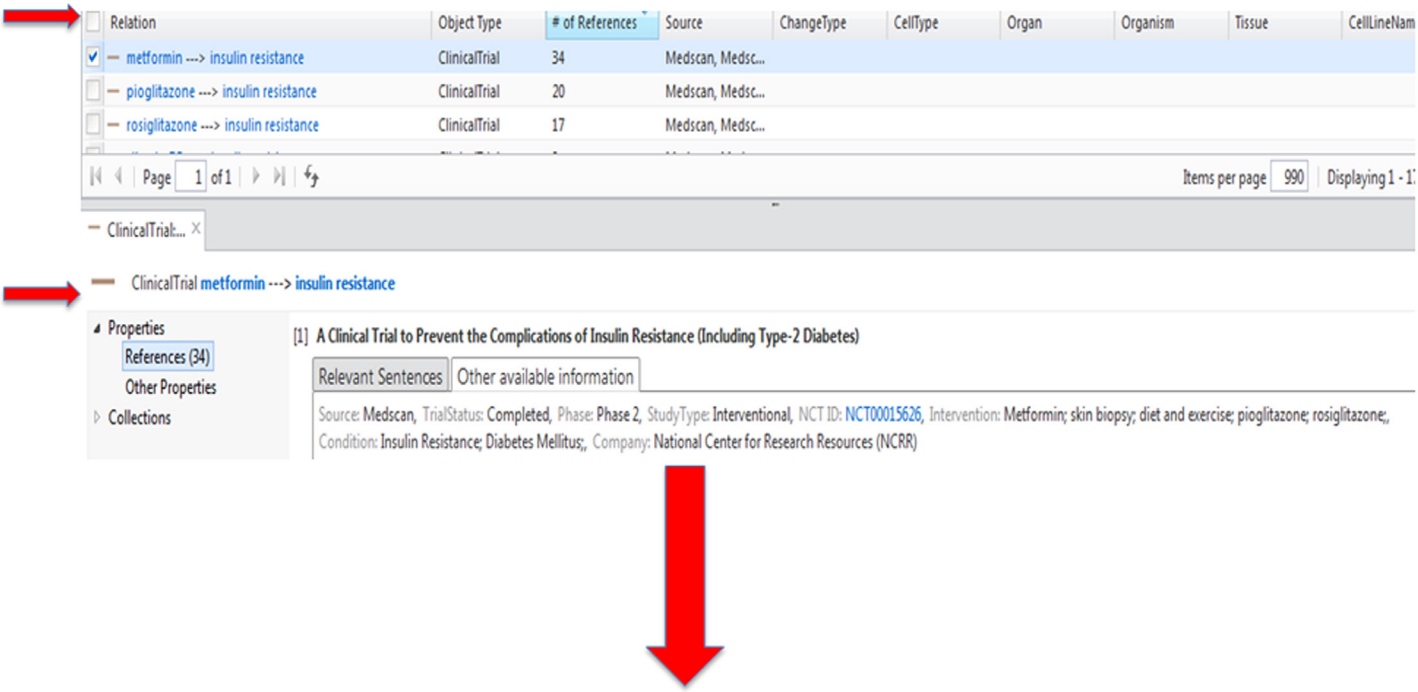
- Copy Insulin Resistance into a New Pathway.
- Select Insulin Resistance.
- Add Network Builder.
- Use Advanced Expand Pathway Tool.
- Select Direction "All."*
- Use Advanced Filters.
- Select Small Molecule as Entity.
- Select ClinicalTrial as Relation.
- Click Next, then Entire Database.
- Click Next, then Finish.

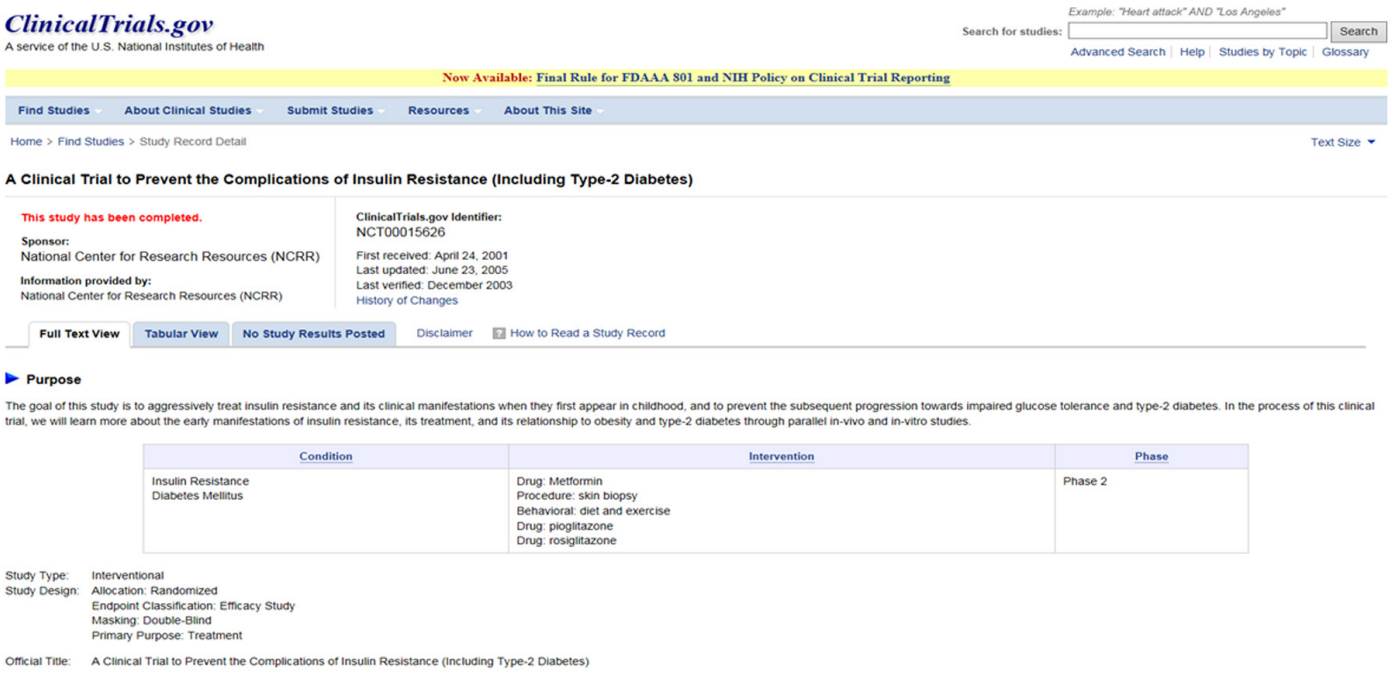


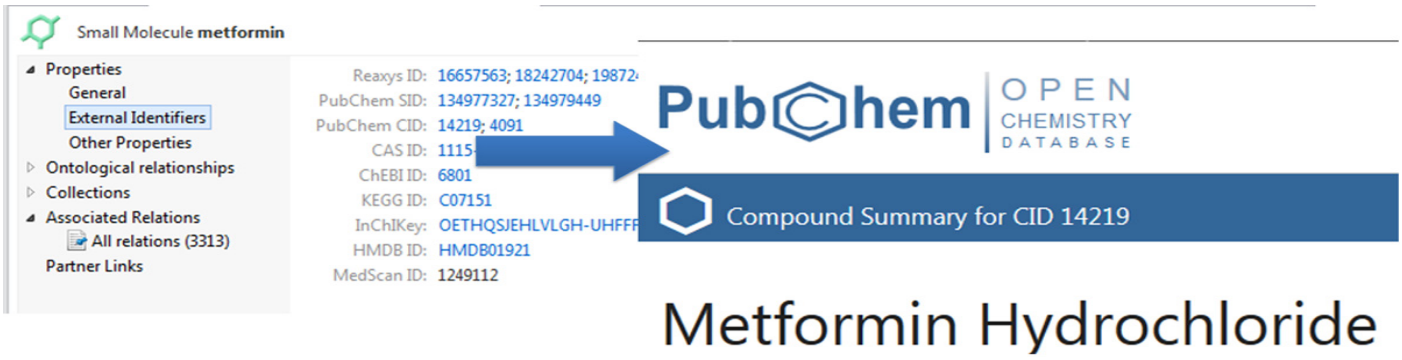


Inspect "Relation Table View" for additional information on clinical trials: example metformin.

Metformin, one of the leading drug treatments for insulin resistance, has been successfully repurposed as an anti-cancer drug.



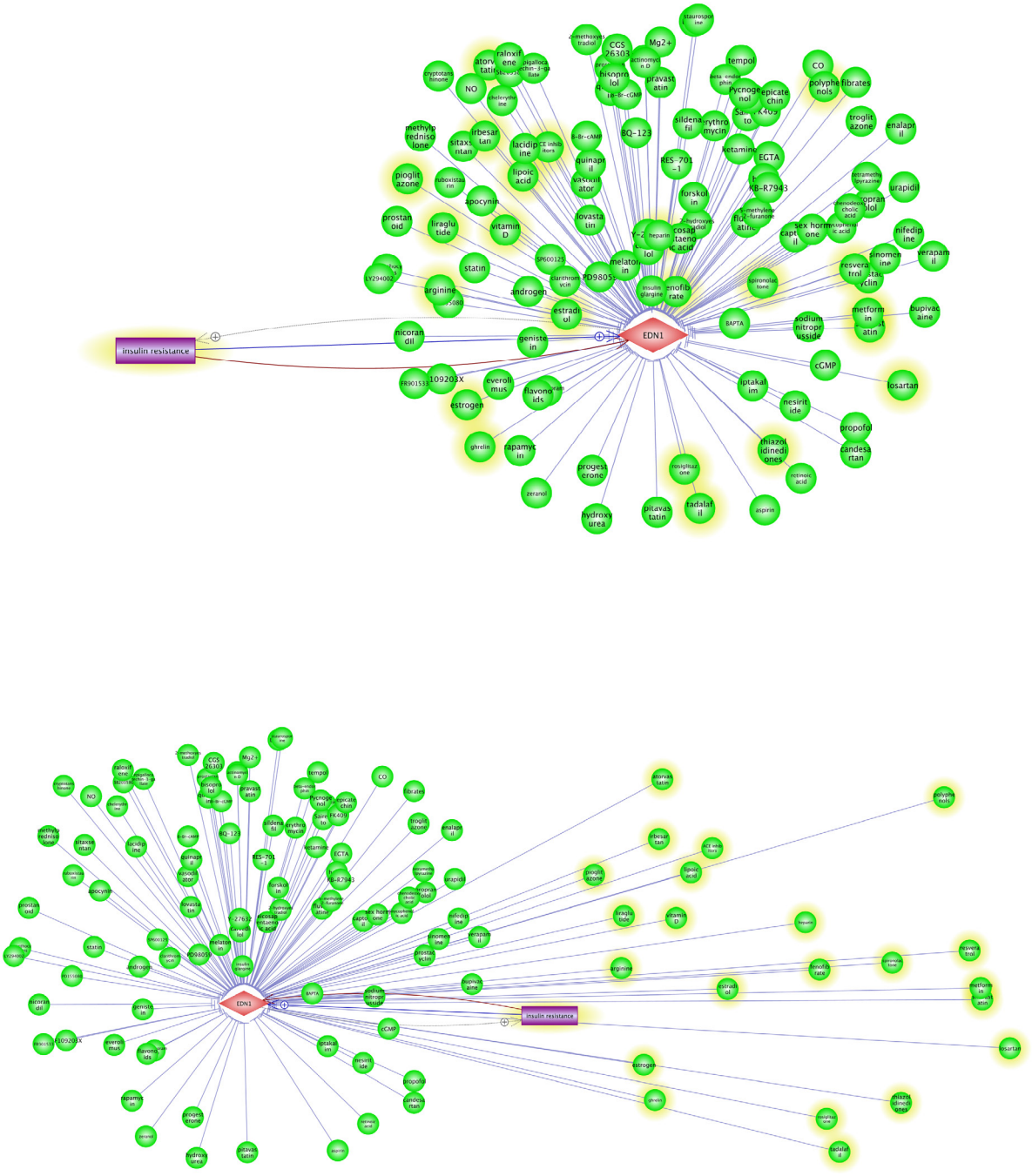




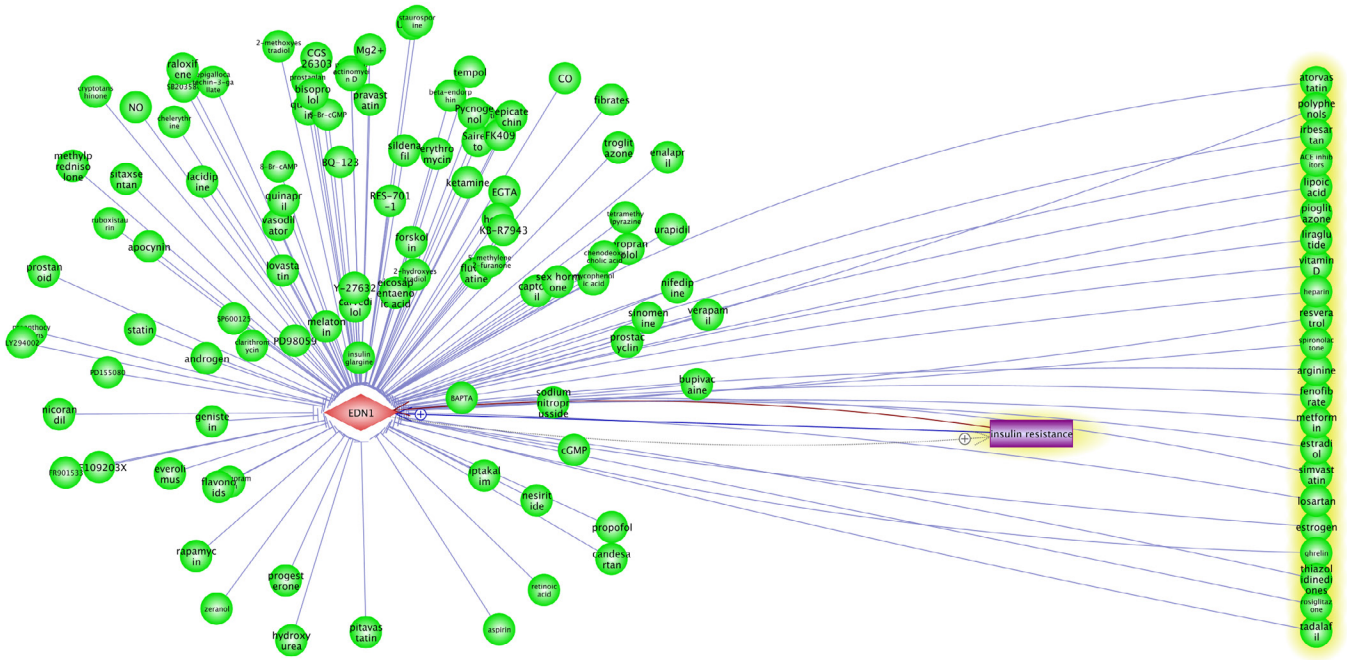
Metformin Hydrochloride is the hydrochloride salt of the biguanide metformin with antihyperglycemic and potential antineoplastic activities. Metformin inhibits complex I (NADPH:ubiquinone oxidoreductase) of the mitochondrial respiratory chain, thereby increasing the cellular AMP to ATP ratio and leading to activation of AMP-activated protein kinase (AMPK) and regulating AMPK-mediated transcription of target genes. This eventually prevents hepatic gluconeogenesis, enhances insulin sensitivity and fatty acid oxidation and ultimately leads to a decrease in glucose levels. **Metformin may exert antineoplastic effects through AMPK-mediated or AMPK-independent inhibition of mammalian target of rapamycin (mTOR), which is up-regulated in many cancer tissues. Furthermore, this agent also inhibits tumor cell migration and invasion by inhibiting matrix metalloproteinase-9 (MMP-9) expression, which is mediated through the suppression of transcription activator protein-1 (AP-1) activation.**

Now, map Small Molecule Inhibitors of EDN1, which are also in clinical trials for insulin resistance.

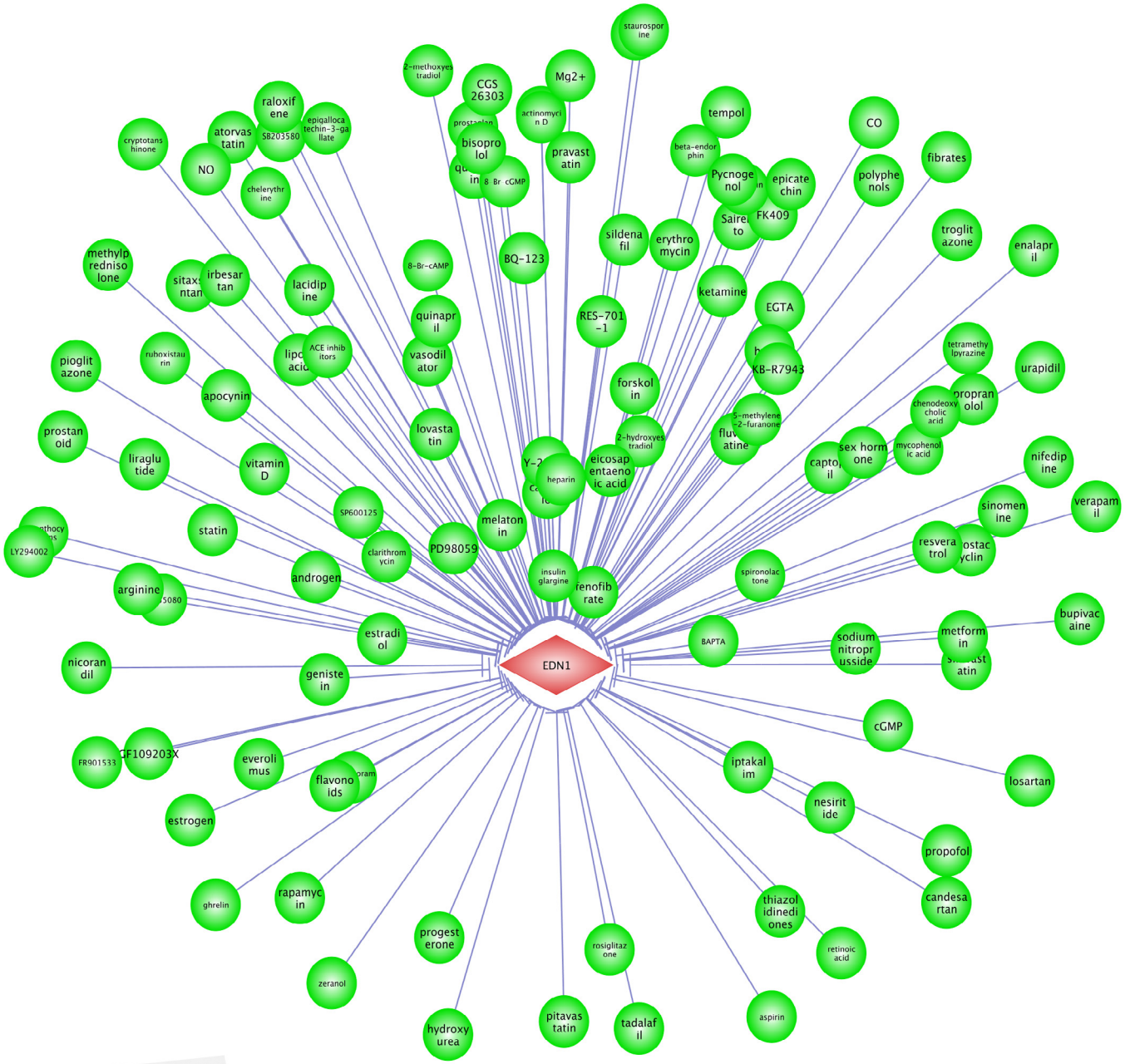
(Hint: Copy all small molecules from insulin resistance clinical trials, go to small molecule inhibitors of EDN1 pathway, select Clipboard Content, highlight selection in yellow and then drag these to one side.)



Now, you can easily see which small molecule inhibitors of EDN1 are in clinical trials or not.



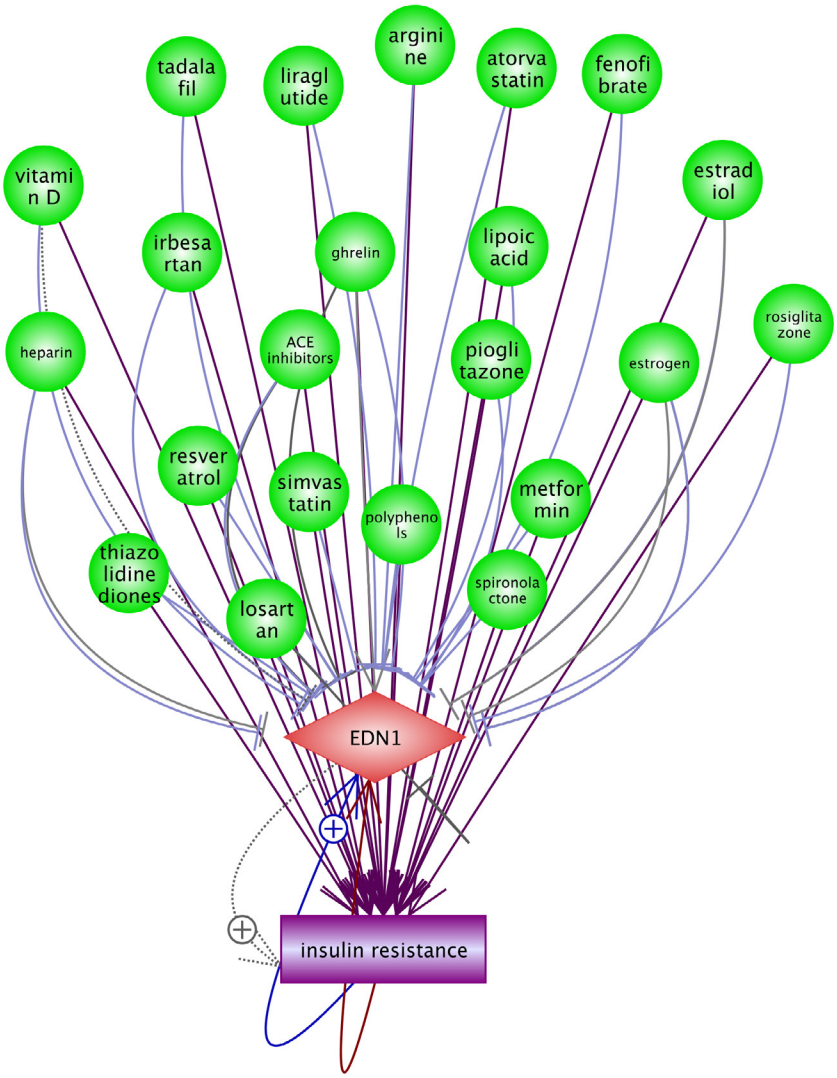
Isolate small molecule inhibitors of EDN1 not already in clinical trials for insulin resistance. (Hint: Simply delete the right hand side of the small molecule inhibitors of EDN1 pathway above.)



For extra credit! Identify three drugs that are strong inhibitors of EDN1 (but not in clinical trials for insulin resistance). Identify their current indication (i.e. what disease they are currently used to treat). Find three different disease-specific drugs that could be repurposed for the treatment of insulin resistance.

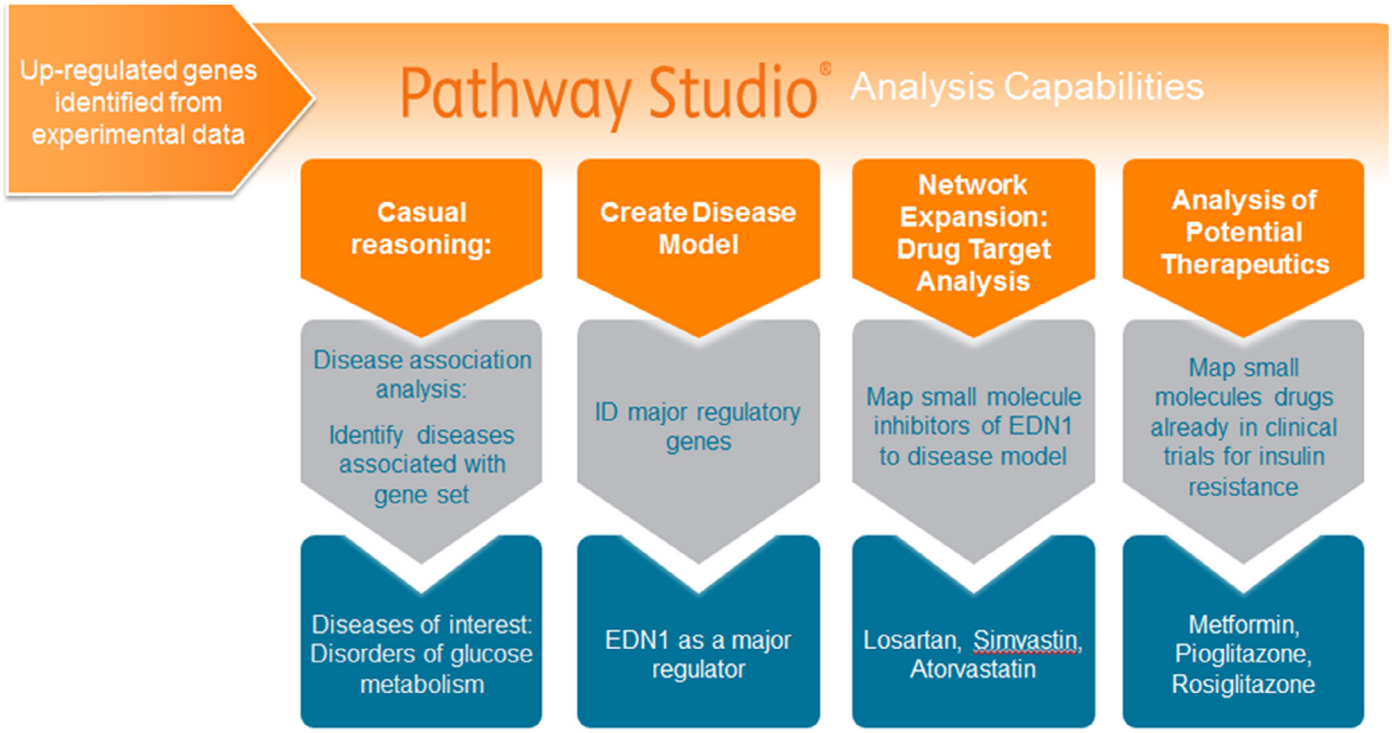
And finally, map...

Small Molecules in Clinical Trials for the treatment of Insulin Resistance mapped to EDN1 small molecule inhibitors only.



(Hint: Now simply delete the left hand side of the pathway above. Then, re-associate both clinical trial and expression relation data. It will be important for you to get creative!)

Time for a quick review!



Study Questions 7

1. Which file types should an adventurer use in his quest to master the Pathway Studio Import feature?
2. What options do you have for the combination of two pathways?
3. Identify top protein (by # of references) that mediates the translocation of EDN1. (Hint: MolTransport)
4. Identify the top small molecule (by # of references) that mediates the translocation of EDN1. (Hint: MolTransport)
5. How many small molecule inhibitors of Crohn's disease are also in clinical trials (for Crohn's disease)? Identify the top small molecule (by # of references).
6. List the top five small molecules (by number of clinical trials) for Crohn's disease.

For Protein/Small Molecule Transport, How Do I ...

Exercise 7.1: What protein mediates the translocation of a protein or small molecule?

Identifies proteins involved in the translocation of a protein or small molecule target.

Step 1: Create New Pathway or within Pathway, select a protein(s)

Step 2: Select Add- > Network Builder

Step 3: Select Expand Pathway, Advanced Expand Pathway Tool

Step 4: Select Direction: Upstream

Step 5: Entities: Select "Protein and Small Molecules" Relations: Select "MolTransport"

Continue to Network Builder Step 5 and click Launch Interactive Network Builder or Finish

Exercise 7.2: What small molecule mediates the translocation of a protein?

Identifies small molecules involved in the translocation of a protein target.

Step 1: Create New Pathway or within Pathway, select a protein(s)

Step 2: Select Add- > Network Builder

Step 3: Select Expand Pathway, Advanced Expand Pathway Tool

Step 4: Select Direction: Upstream

Step 5: Entities: Small Molecules Relations: MolTransport

Continue to Network Builder Step 5 and click Launch Interactive Network Builder or Finish

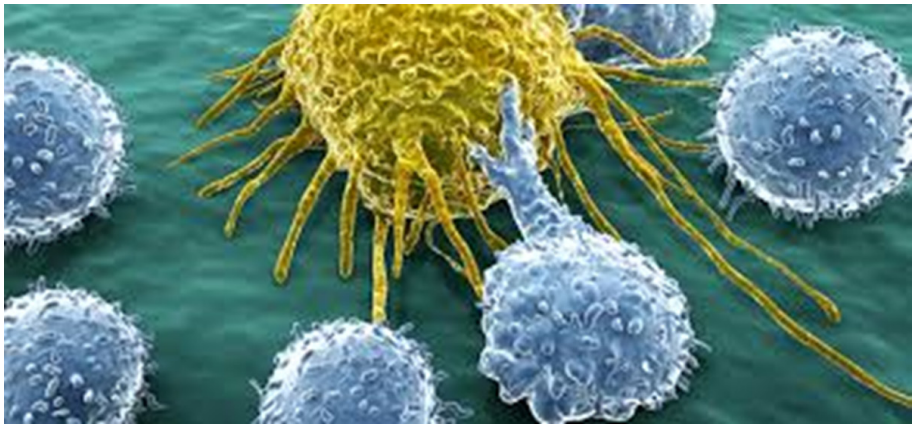
Module 8

Cancer Immunotherapy

Contents

8.1 Introduction

Cancer Immunotherapy



Immunotherapy has proven to be one of the most exciting new areas for oncology.

So what is it?

Well, cancer immunotherapy has arisen from the recognition that the human immune system does have effective defensive measures against malignant cells.

Knowledge of the basic mechanisms of the immune system as it relates to cancer has been increasing rapidly.

It is thought that the immune system is able to perceive and eliminate some tumors early on in their development.

The ability of T cells to constantly survey host tissues for newly transformed cells and to control and/or eliminate human cancers is known as **the cancer immunosurveillance hypothesis.**

So what's the problem?

Well unfortunately, some tumor cells can escape immunosurveillance and adapt counter measures. This entire process of surveillance and escape is known as the **cancer immunoediting hypothesis** (see below).

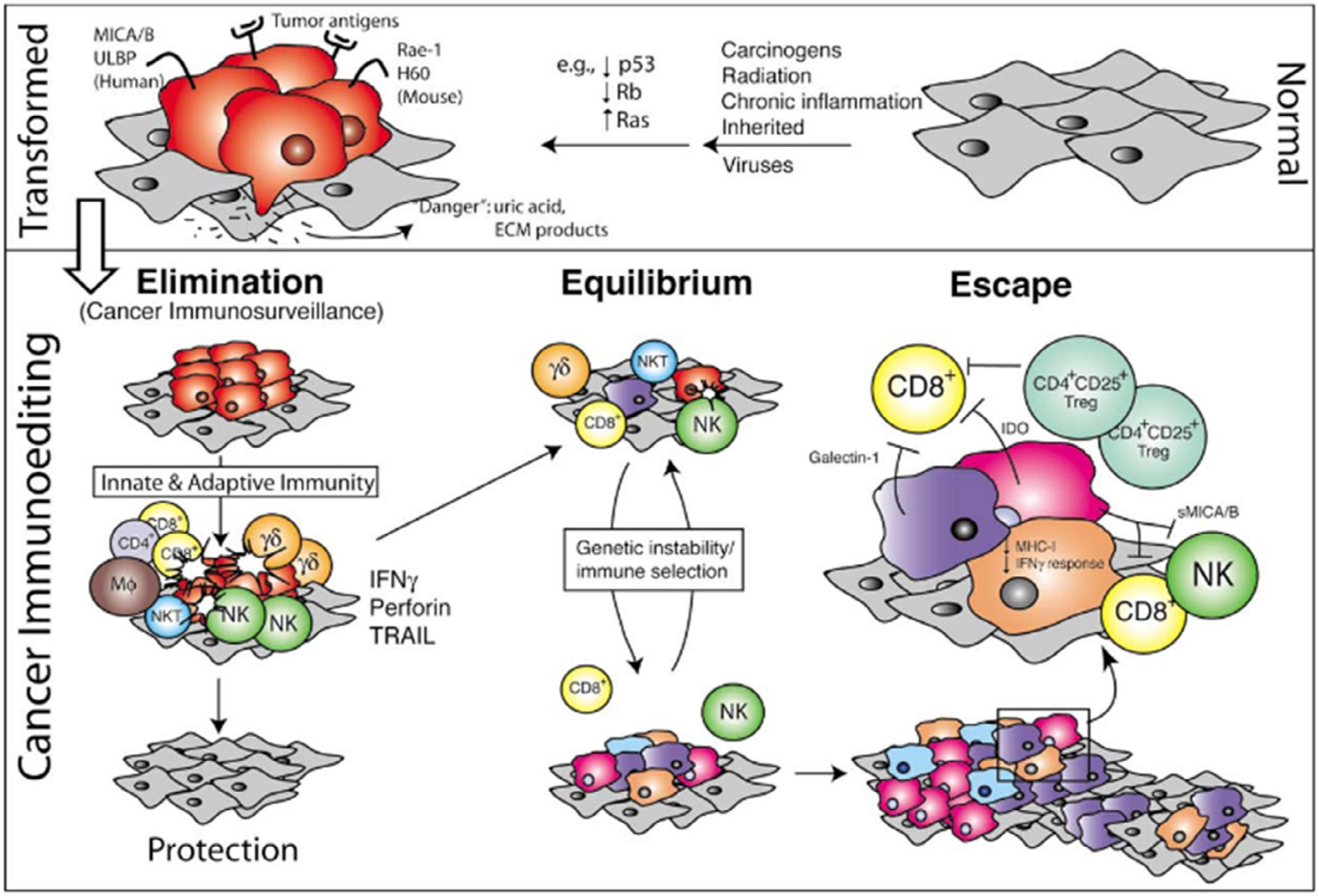
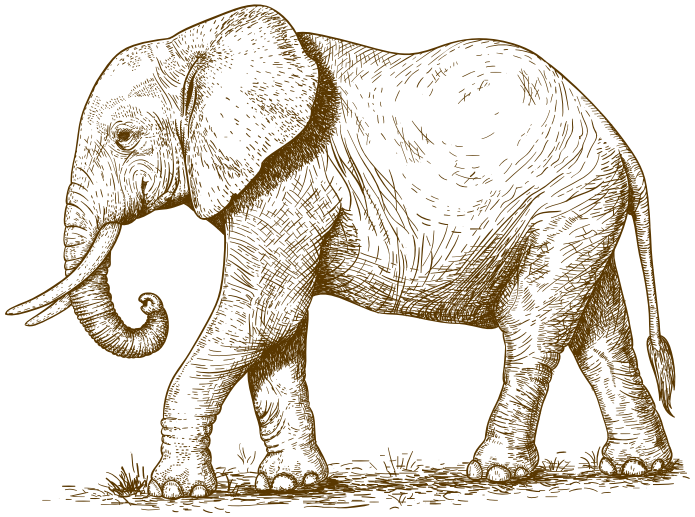


Figure 1. The Three Phases of the Cancer Immunoediting Process

Normal cells (gray) subject to common oncogenic stimuli ultimately undergo transformation and become tumor cells (red) (top). Even at early stages of tumorigenesis, these cells may express distinct tumor-specific markers and generate proinflammatory "danger" signals that initiate the cancer immunoediting process (bottom). In the first phase of elimination, cells and molecules of innate and adaptive immunity, which comprise the cancer immunosurveillance network, may eradicate the developing tumor and protect the host from tumor formation. However, if this process is not successful, the tumor cells may enter the equilibrium phase where they may be either maintained chronically or immunologically sculpted by immune "editors" to produce new populations of tumor variants. These variants may eventually evade the immune system by a variety of mechanisms and become clinically detectable in the escape phase.

The immunobiology of cancer immunosurveillance and immunoediting. Dunn GP, Old LJ, Schreiber RD. Immunity. 2004 Aug;21(2):137-48. Review.

How tumor cells escape immunosurveillance and the many different counter measures now under development is a BIG subject.

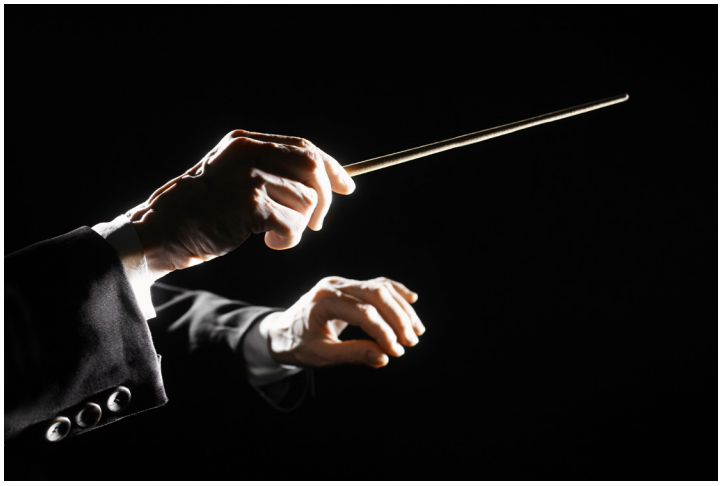


In general, contemporary immunotherapies can be classified into two types: active or passive. Active therapies attempt to induce an immune response in otherwise non-responsive patients. These include cytokines (e.g. IL-2), immunomodulatory mABs (e.g. checkpoint inhibitors), and cancer vaccines.

Passive therapies stimulate a patient's intrinsic immune response and include cell-based therapies (e.g. CAR-T), bispecific and multispecific antibodies (e.g. Blinatumomab CD3-CD19), oncolytic viruses (e.g. T-Vec), and tumor-targeting mABs (e.g. Rituximab, anti-CD20).

8.2 Secreted Protein of Melanoma Cells

The workflow presented here is based on the observation that at least one of the many ways that cancer cells escape immunosurveillance is through the secretion of protein products, which down-regulate critical cellular components of the immune system.

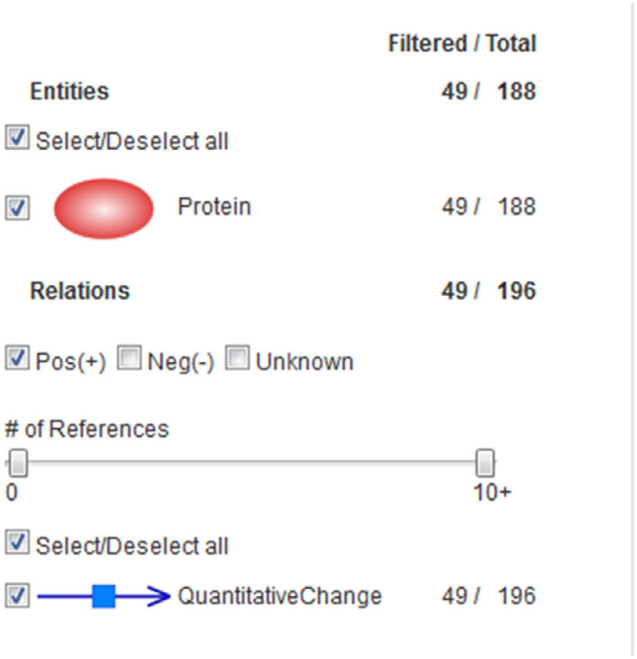


Shall we begin?!!!

Step 1: Search for the disease "melanoma" and place it on the pathway work space.

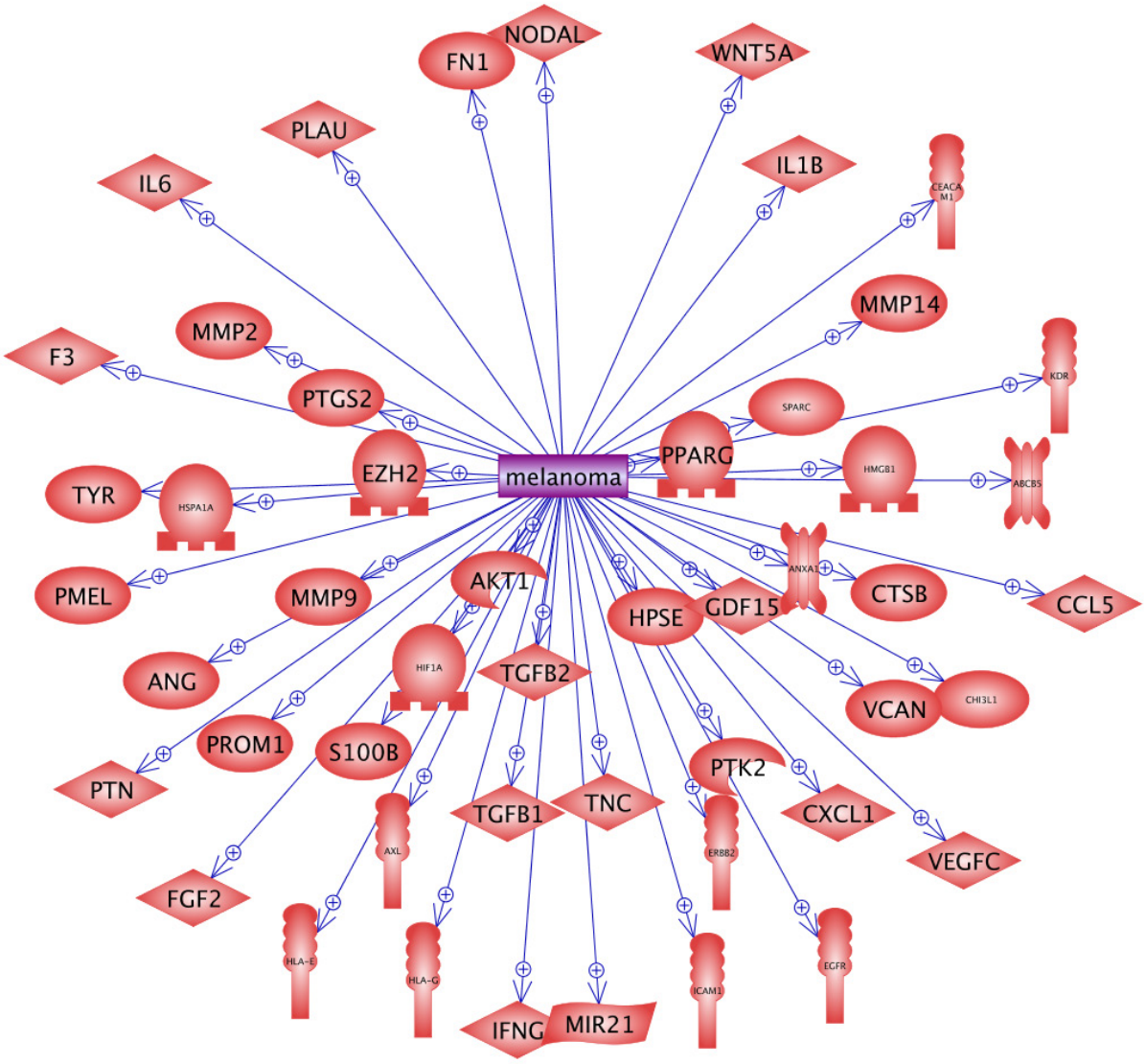
Step 2. Now let's "Add"

Tool category: Advanced Expand Pathway
Tool Direction: Downstream Protein: QuantitativeChange (QuantitativeType = secretion)



Use the Interactive Network Filter to select only positive relations.

This should result in about 50 proteins.



It may be obvious, but it's always a good idea after creating a network to take the time to examine a sample of the relations to ensure that the question you thought you were asking has indeed been answered properly. In this case, you were looking for proteins known to be secreted by melanoma cells. The pop-up box for CCL5 confirms the search strategy (Hint: Another way to look at this quickly and even more comprehensively is to go to the Relation Table View.)

8.3 Cells Inhibited by Melanoma Proteins

Now, go find all the cell types that are inhibited by these melanoma-secreted proteins.

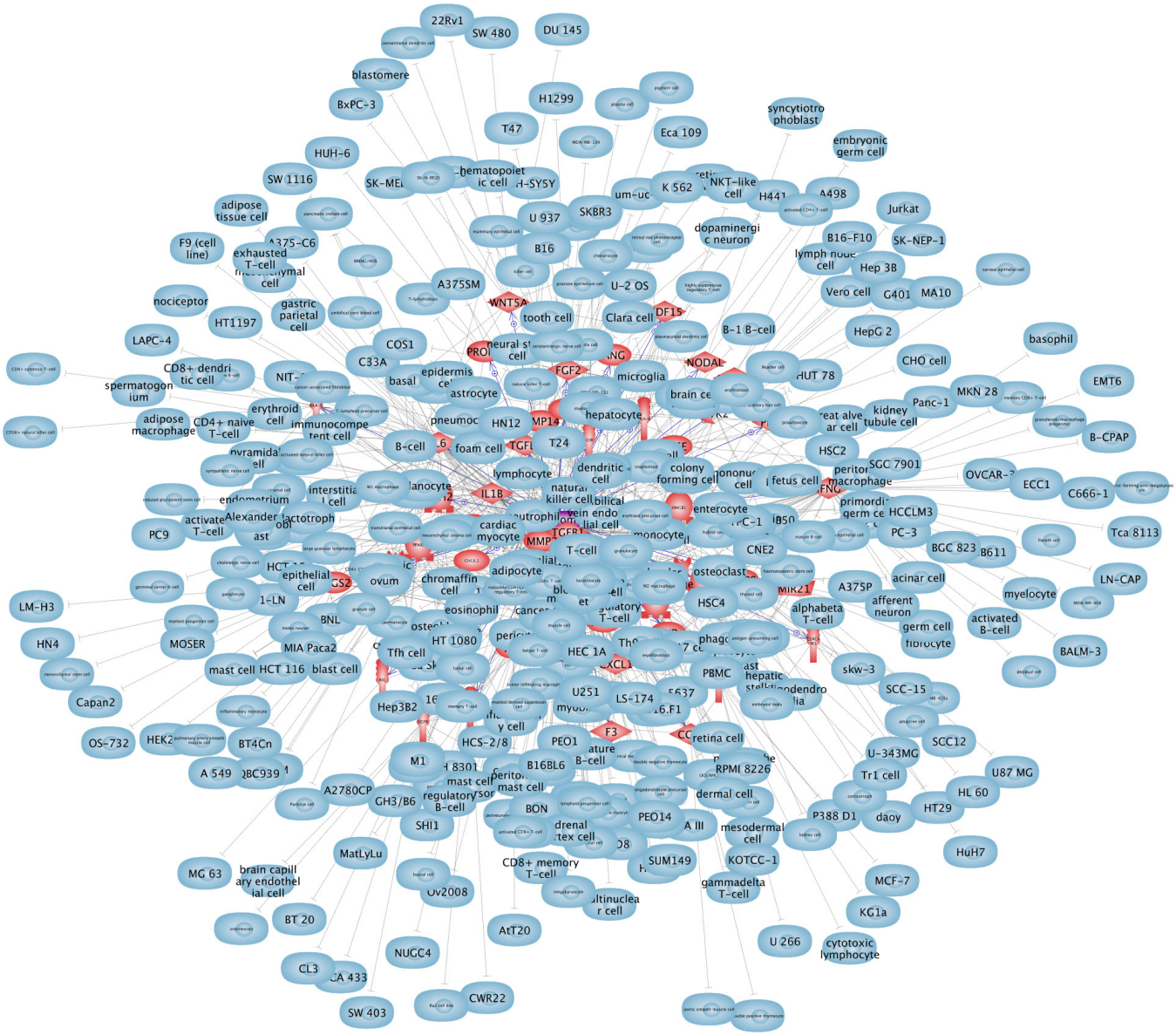
First, select "Entities by Type," and then "Protein."

Then, go to add **"Network Builder,"** and choose **"Advanced Expand Pathway Tool."**

Next, select "Downstream" (Hint: In effect, this tells the software that you are interested in finding cells that are inhibited by the melanoma proteins.).

Finally, select "Cell" (for Entities), "Regulation" (for Relations), and "Effect"= "negative."

You found nearly 400 different cell types! These are all cells that are inhibited by the melanoma-secreted proteins.



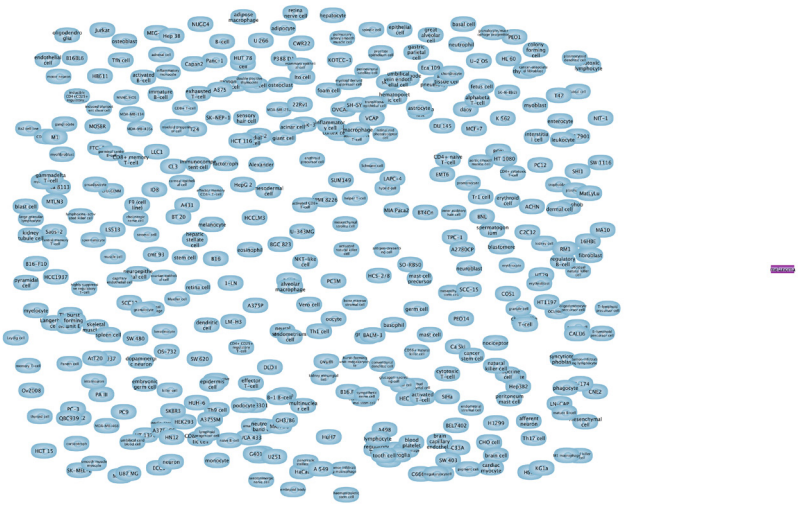
OK, this is where it gets a little tricky!

We need to find out which of these cells has activity against melanoma disease (in other words, identify a specific subset of these cells).

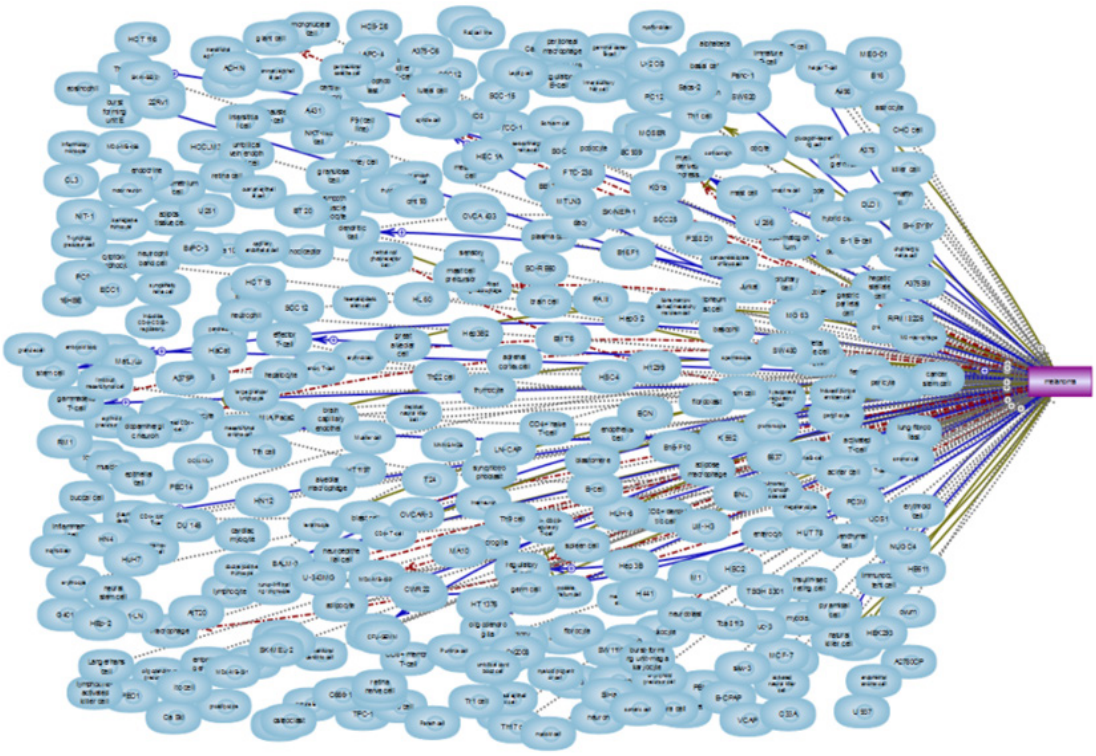
There are several different ways to approach this problem, and the one below might be the simplest (we hope!).

First step, make sure you've saved your work to this point, and then copy and paste just the cells and melanoma disease into a new workspace **("Create New Pathway")**.

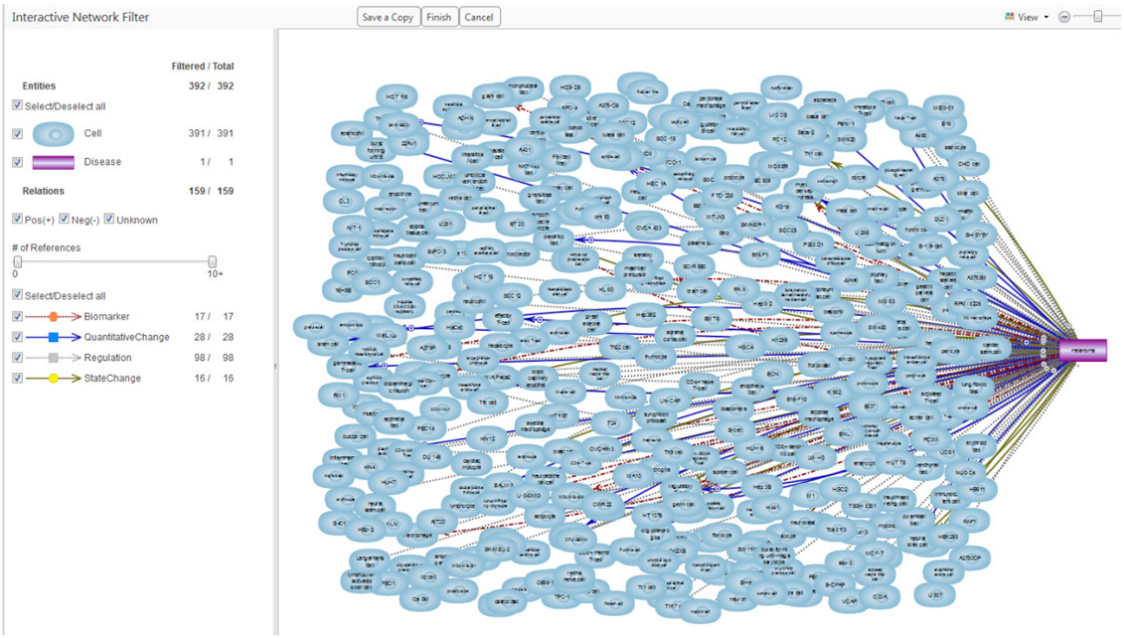
It should look something like this:

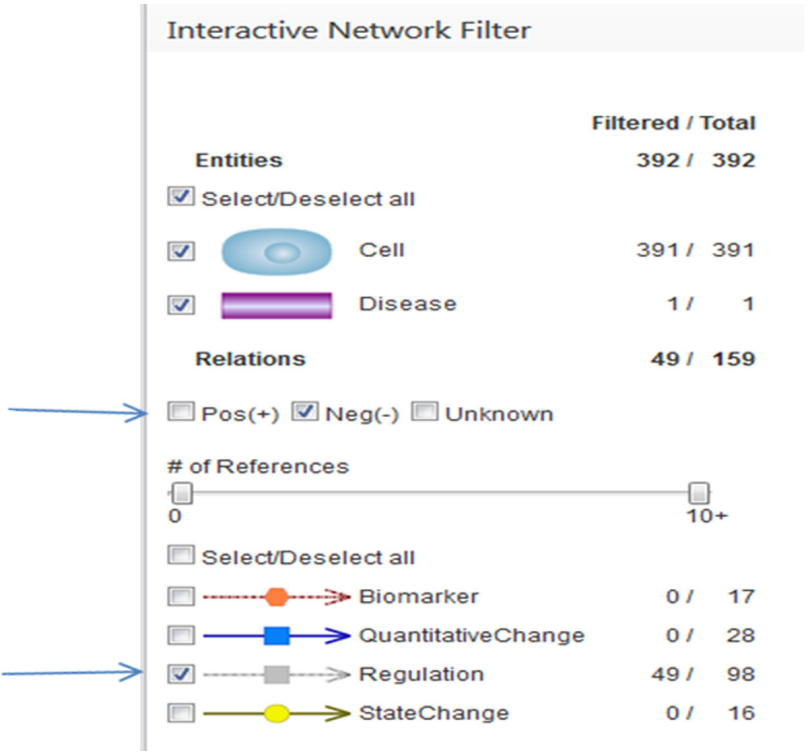


Select melanoma disease and add “Relations between Selected and Unselected.”



And now it’s time , once again, for....



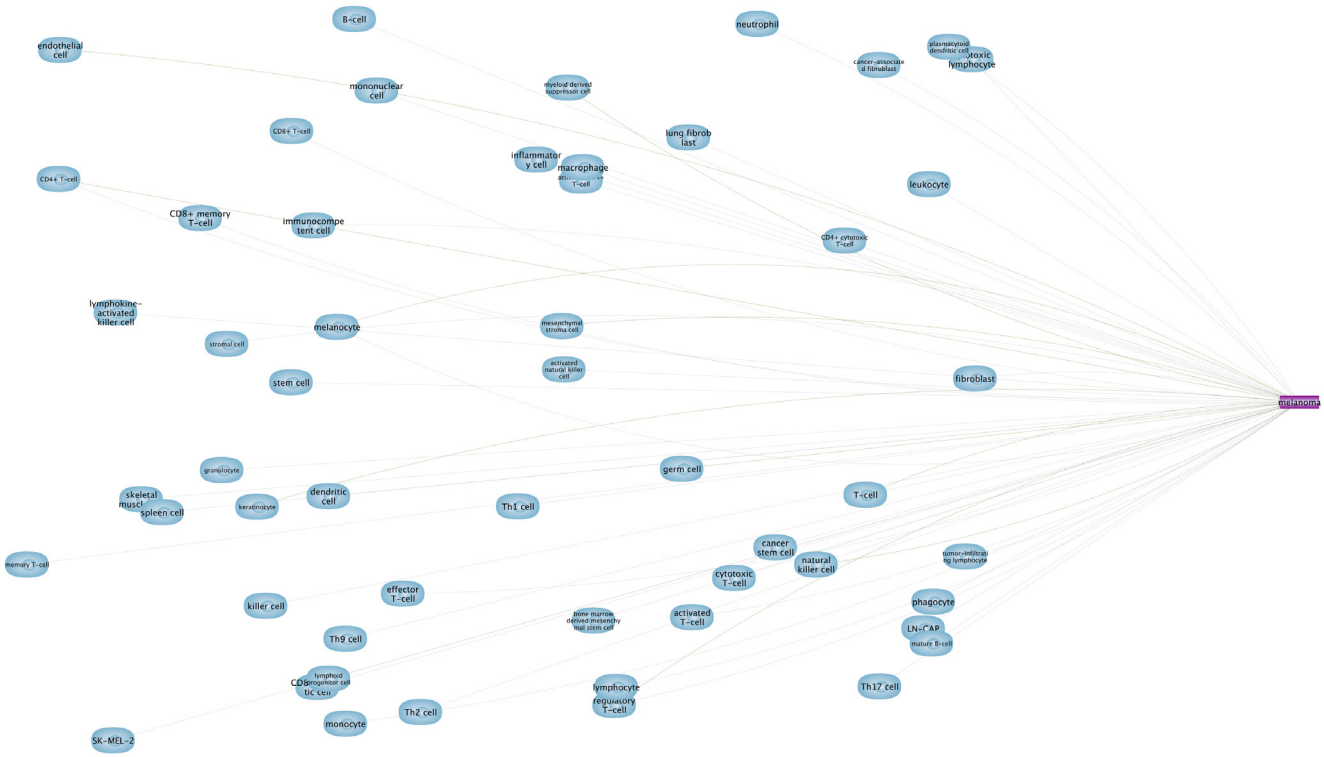


1. Select only Neg(-) as the effect, and Regulation as the relation.

2. Click “Finish”.

3. Select “Unconnected Entities”.

4. Edit, Remove.



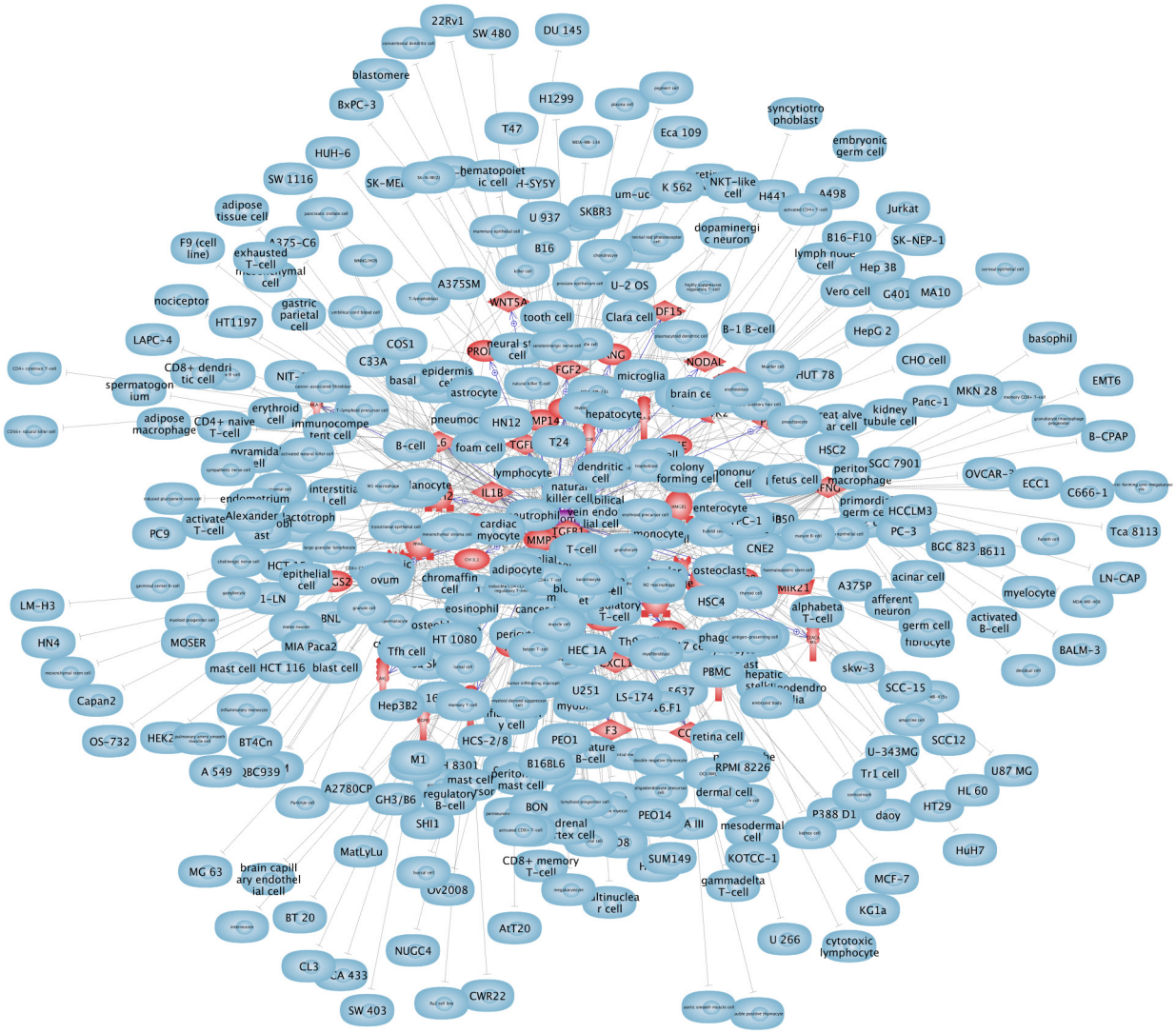
8.4 Cells Inhibiting Melanoma Cells

Now, the goal is to use this identified subset of cells to modify our earlier melanoma

-> secreted proteins-> negatively regulated cells.

At the end of the day, you only want cells that have known inhibitory action against melanoma.

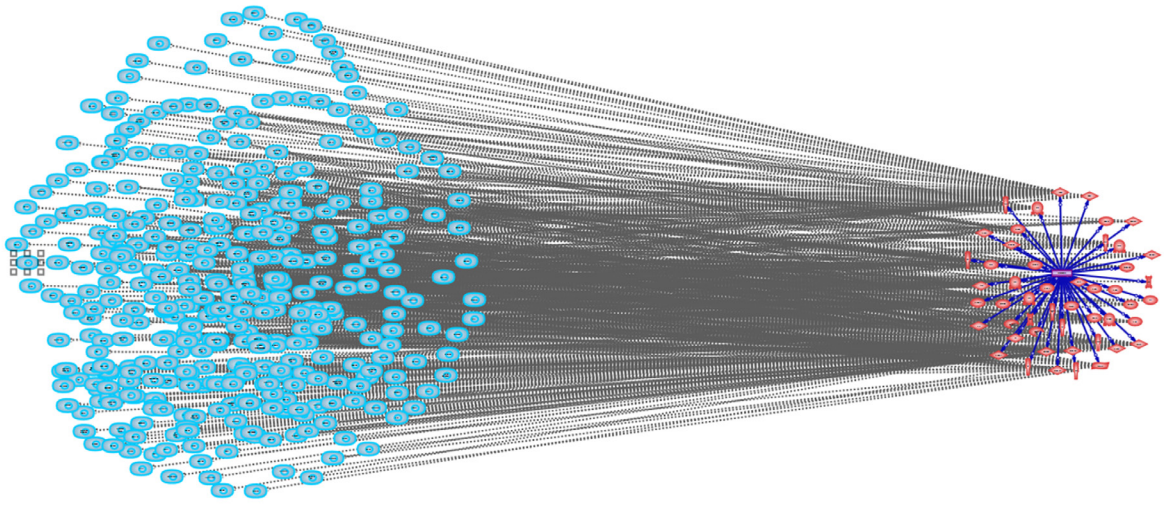
So now, copy these cells and go back to this pathway:



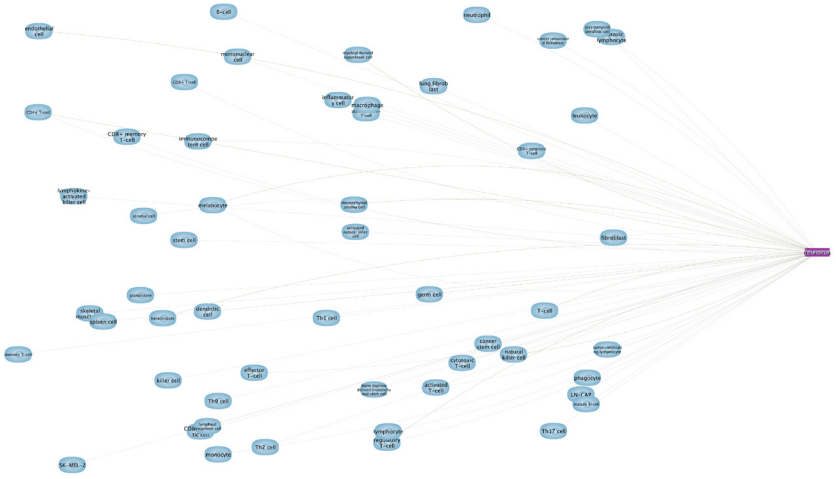
8.5 Integrating Three Data sets

But first, you need to make some big moves!

Select cells and drag them far to the left, like so:



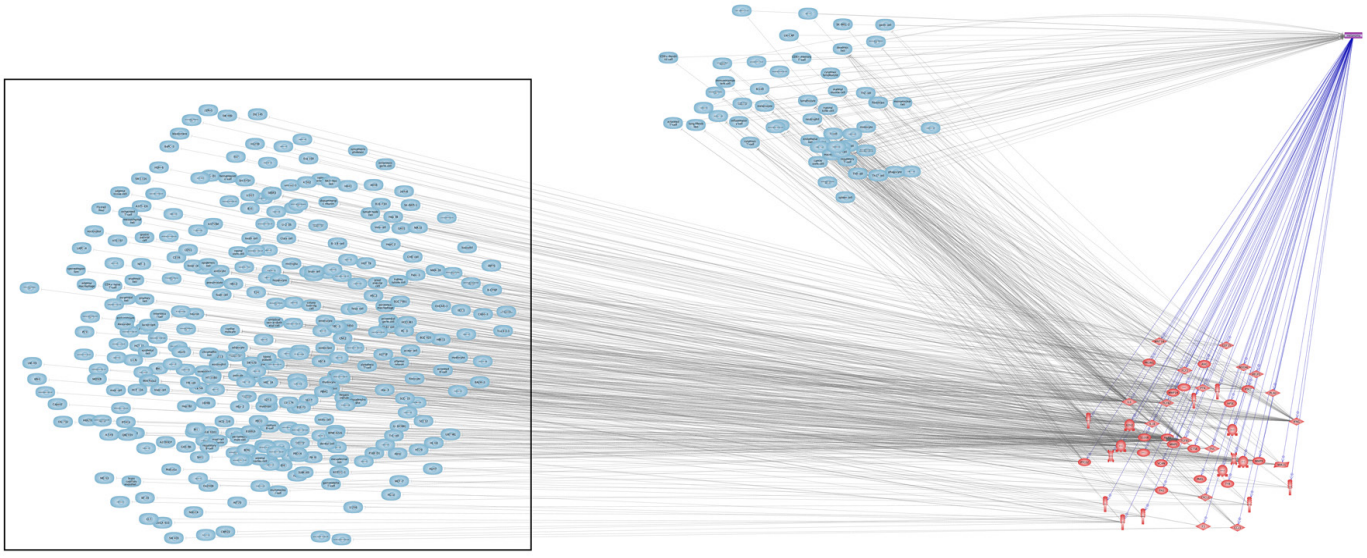
Next, select all and copy from here:



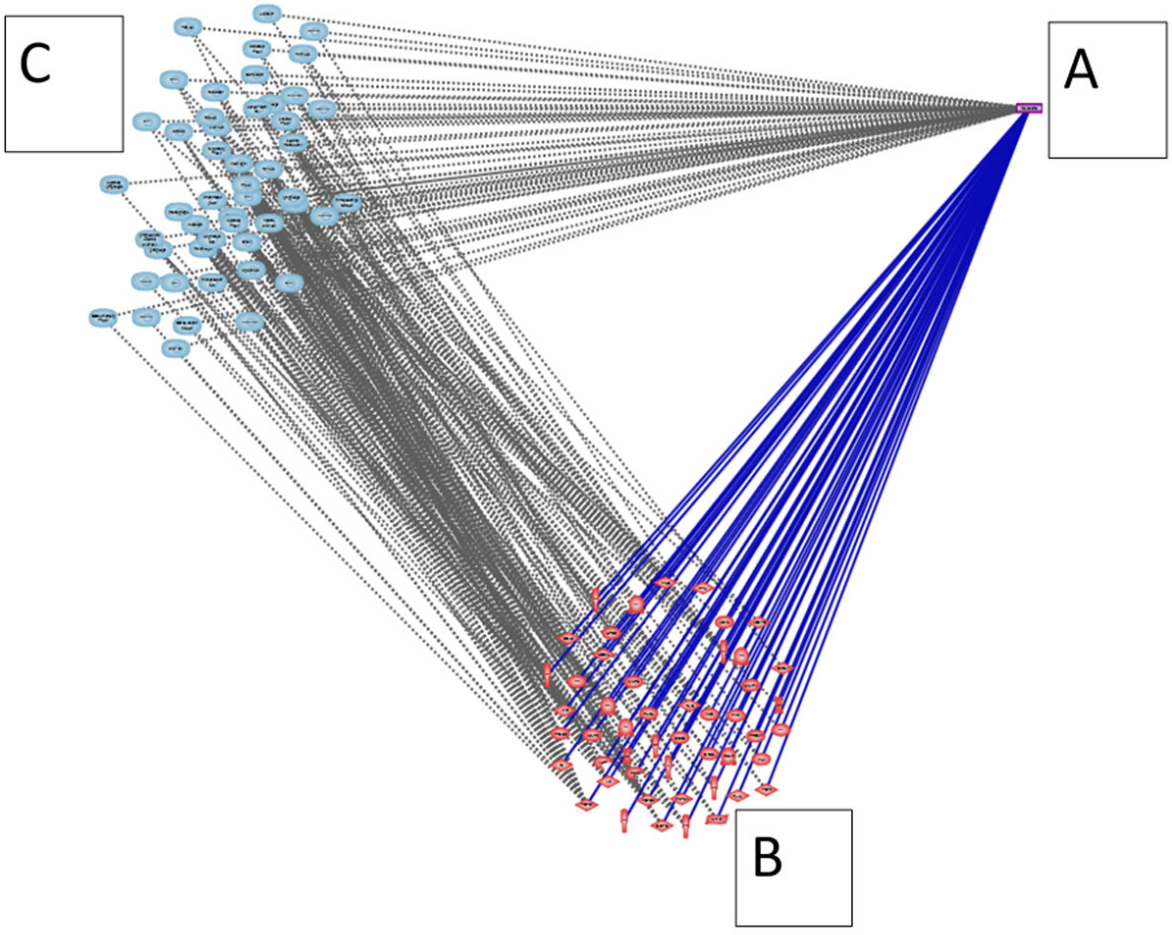
Return to the main pathway, paste and then drag the selected cells up and out of the way, like this:

8.6 Closing the Loop!

Now, you can select and remove the unwanted cells (on the left).



Voila!

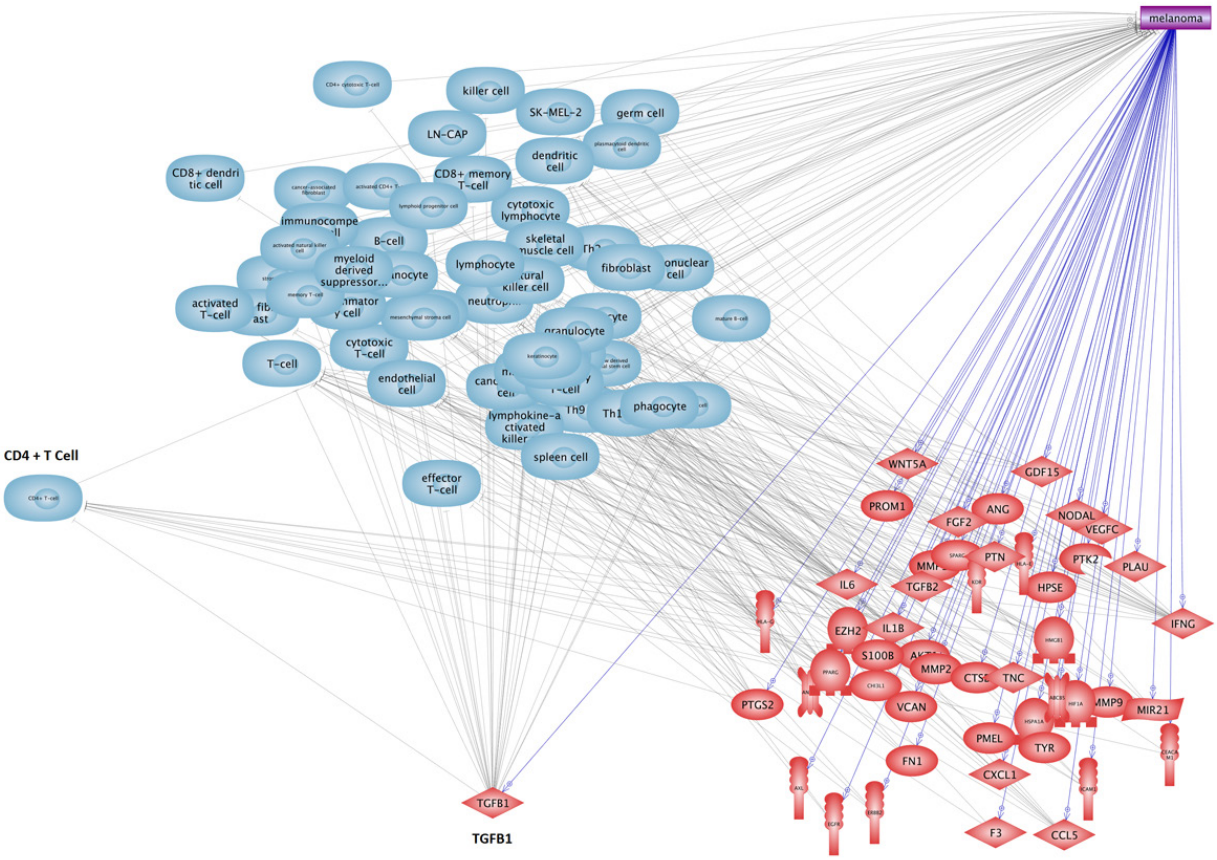


See how you just closed the loop? Melanoma (A) secreted proteins (B) inhibit cells (C) that inhibit melanoma (A)!

Now we can…analyze your data.

Drill Down!

We can see that some entities are highly connected (e.g. TGFB1 and CD4+ T cells).



Let's begin to explore this systematically by going to the "View Entities" Table,

Sort descending on column named "Local Connectivity" (if not there go to Customize table to select and display that column).

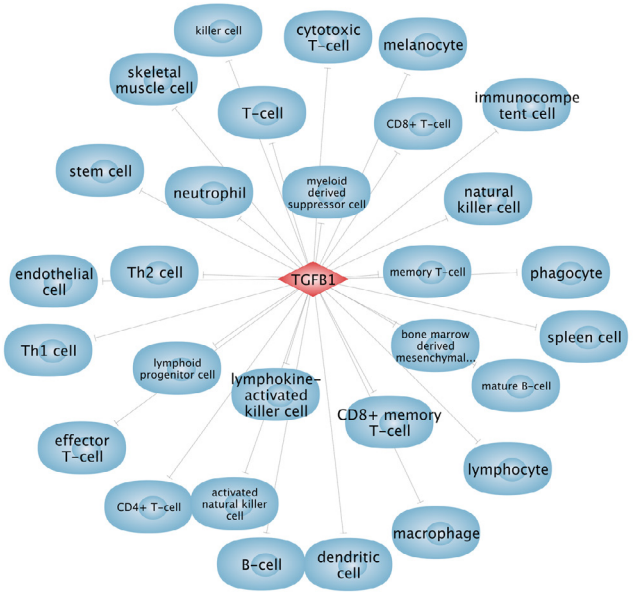
Name	Object Type	Total Connectivity	Local Connectivity
melanoma	Disease	7718	98
TGFB1	Protein	15758	28
IFNG	Protein	13019	21
IL6	Protein	18336	19
T-cell	Cell	12715	18
HLA-G	Protein	1181	12
PPARG	Protein	8178	12
neutrophil	Cell	8458	12
Th1 cell	Cell	3470	12
CD4+ T-cell	Cell	4763	12
dendritic cell	Cell	6608	11
IL1B	Protein	13355	11
PTGS2	Protein	9926	10
TGFB2	Protein	2703	10

From this we can see that TGFB1 is indeed the most highly connected protein (CD4+ T-cells are actually the fourth most connected cell type and connect with TGFB1 as well).

Let's take a closer look at this! (can you figure out how? Hint: simply delete all proteins, other than TGFB1, and melanoma disease, and then remove unconnected entities).

8.8 TGFBI as a Therapeutic Target

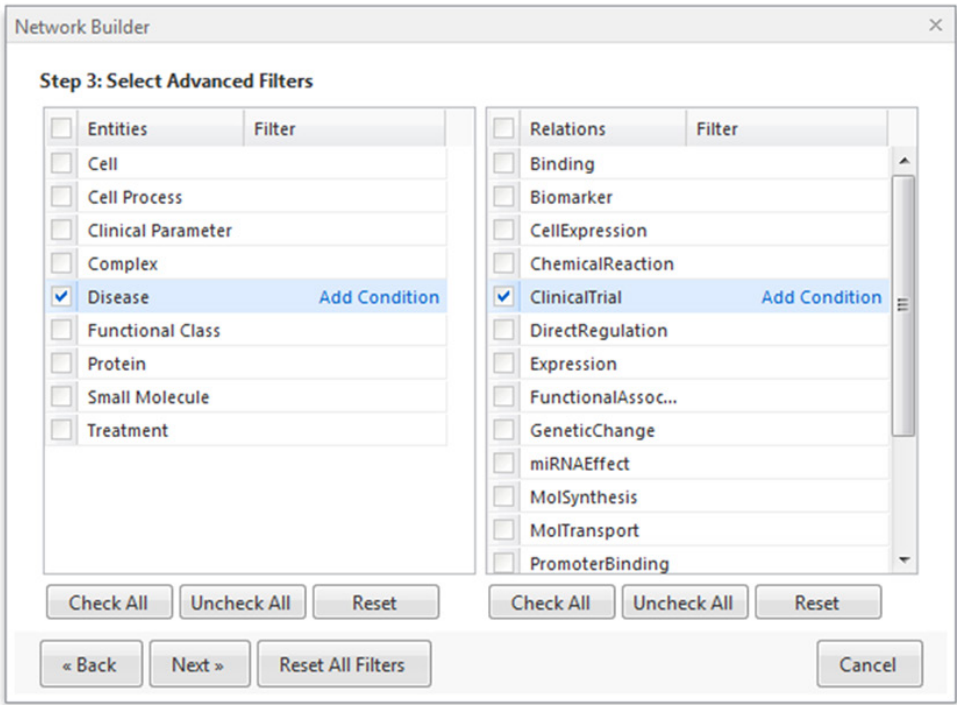
From this we can see that TGFB1 is indeed the most highly connected protein (CD4+ T-cells are actually the fourth most connected cell type and connect with TGFB1 as well).

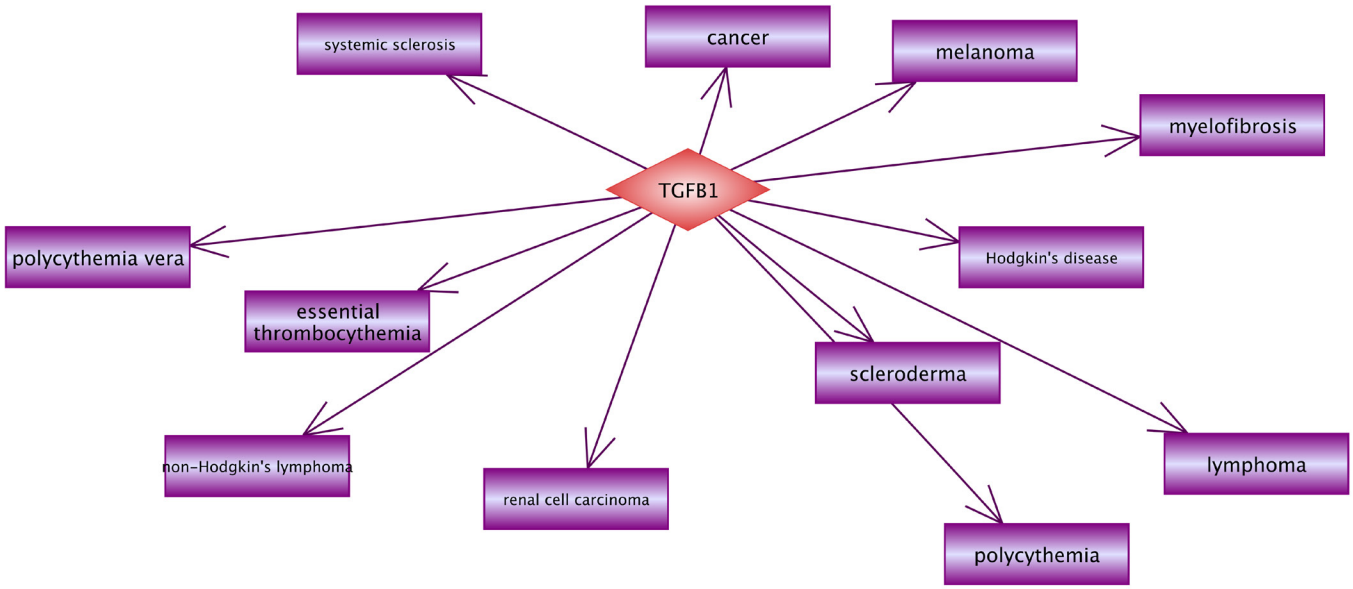


8.9 TGFBI in Clinical Trials

Has TGFB1 ever been investigated as a target point for therapeutic intervention in a clinical trial?

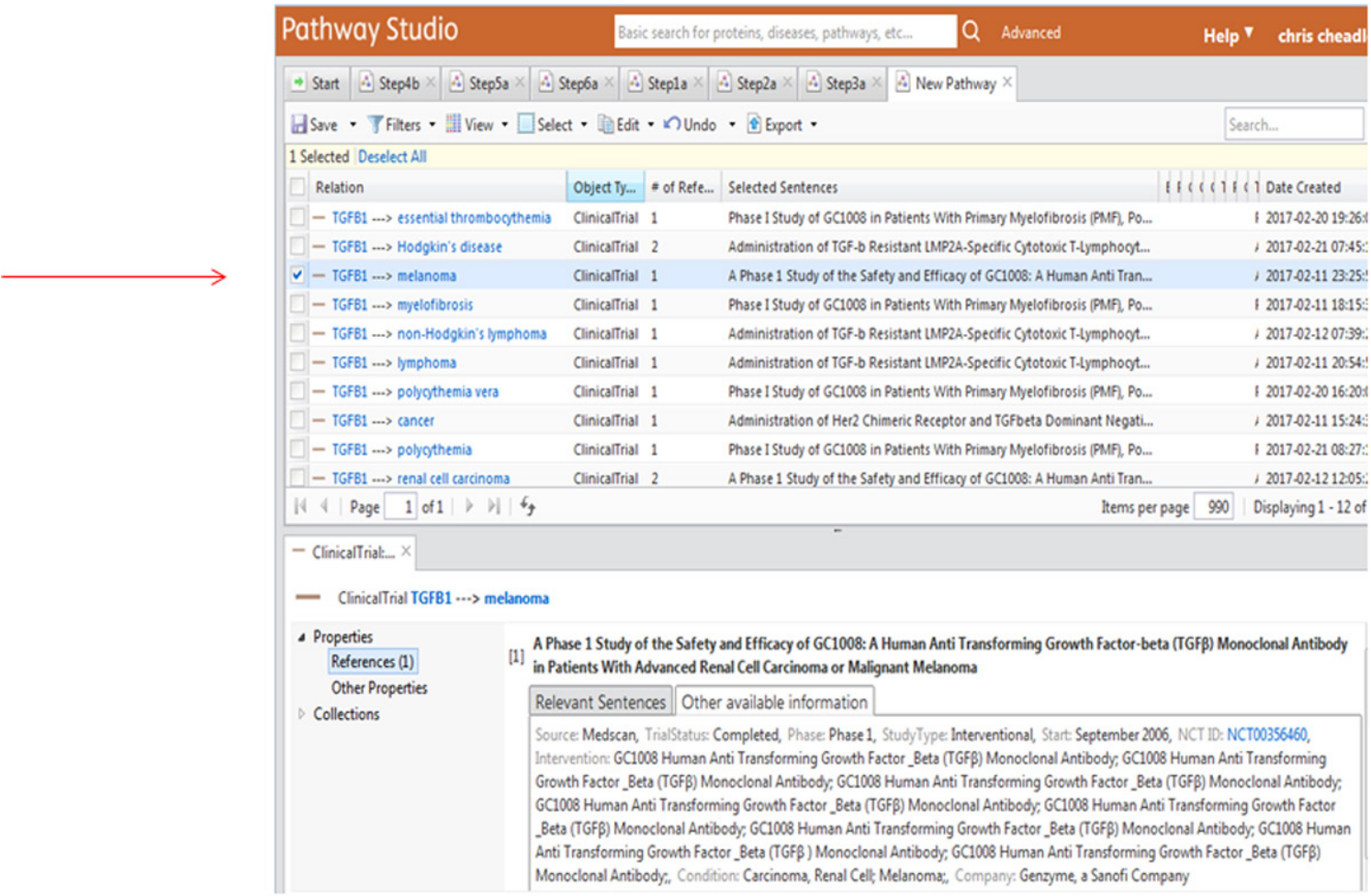
Start a new pathway and find out!





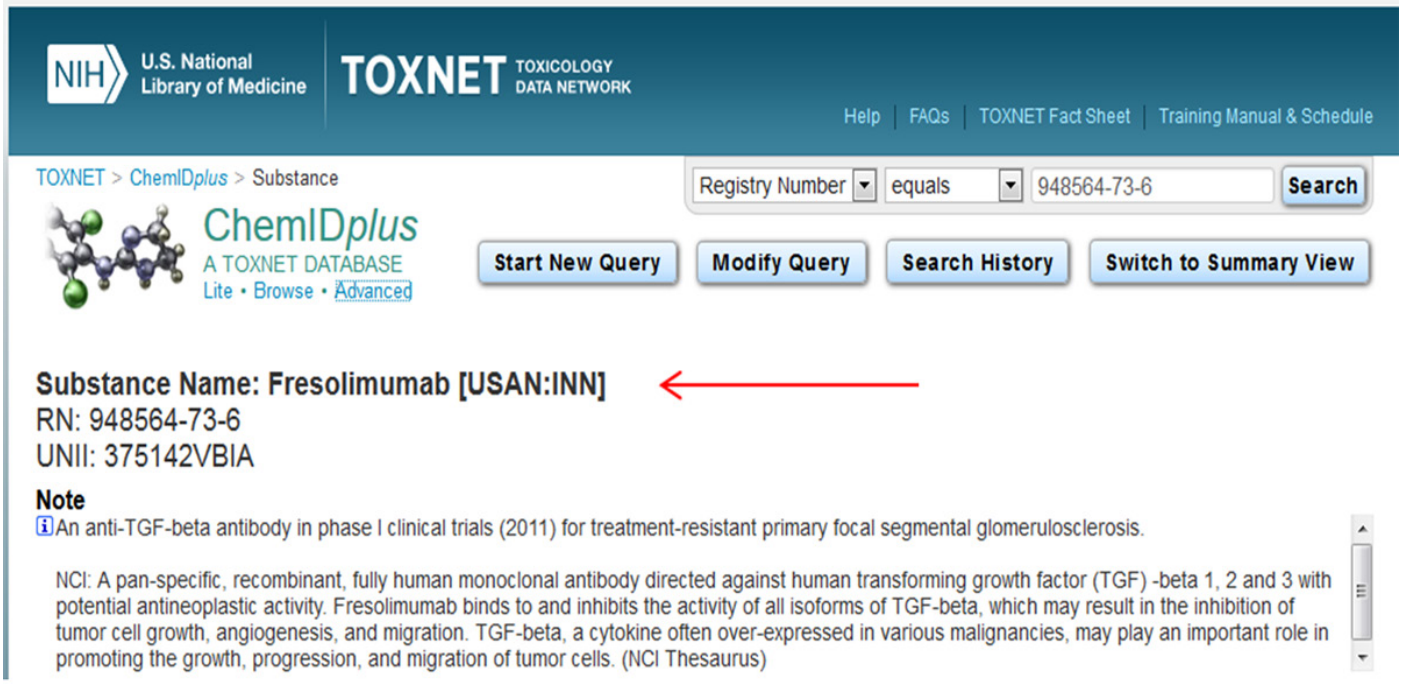
It looks like TGFB1 has been targeted across a broad range of disease indications.

We even have a clinical trial involving melanoma.



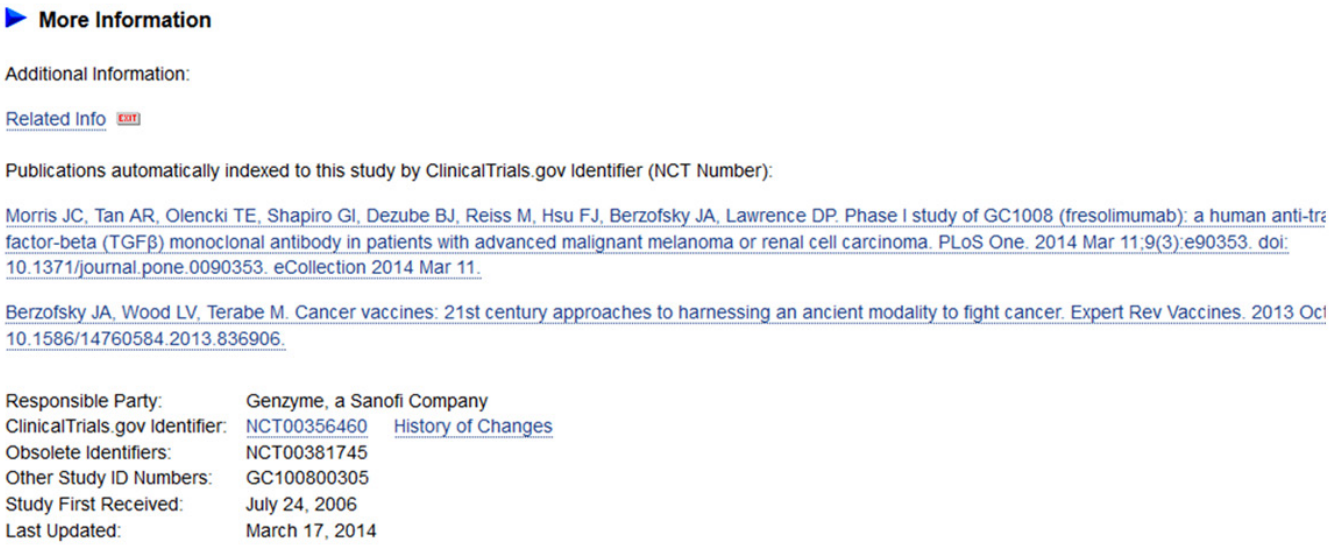
8.10 Fresolimumab

Can you find more information about the TGFB1 inhibitor (GC1008) currently being investigated for the treatment of melanoma in clinical trials? (Hint: Search for GC1008 in the Query box under Properties, click External Identifiers, CAS ID, and Open in ChemID plus.)



Wonder how this clinical trial turned out? It's completed.

Just go to the bottom of the clinical trial record page.



It looks like it helps!

PLoS One. 2014 Mar 11;9(3):e90353. doi: 10.1371/journal.pone.0090353. eCollection 2014.

Phase I study of GC1008 (fresolimumab): a human anti-transforming growth factor-beta (TGFβ) monoclonal antibody in patients with advanced malignant melanoma or renal cell carcinoma.

Morris JC[1], Tan AR[2], Olencki TE[3], Shapiro GI[4], Dezube BJ[5], Reiss M[2], Hsu FJ[6], Berzofsky JA[1], Lawrence DP[7].

⊕ Author information

Abstract

BACKGROUND: In advanced cancers, transforming growth factor-beta (TGFβ) promotes tumor growth and metastases and suppresses host antitumor immunity. GC1008 is a human anti-TGFβ monoclonal antibody that neutralizes all isoforms of TGFβ. Here, the safety and activity of GC1008 was evaluated in patients with advanced malignant melanoma and renal cell carcinoma.

METHODS: In this multi-center phase I trial, cohorts of patients with previously treated malignant melanoma or renal cell carcinoma received intravenous GC1008 at 0.1, 0.3, 1, 3, 10, or 15 mg/kg on days 0, 28, 42, and 56. Patients achieving at least stable disease were eligible to receive Extended Treatment consisting of 4 doses of GC1008 every 2 weeks for up to 2 additional courses. Pharmacokinetic and exploratory biomarker assessments were performed.

RESULTS: Twenty-nine patients, 28 with malignant melanoma and 1 with renal cell carcinoma, were enrolled and treated, 22 in the dose-escalation part and 7 in a safety cohort expansion. No dose-limiting toxicity was observed, and the maximum dose, 15 mg/kg, was determined to be safe. The development of reversible cutaneous keratoacanthomas/squamous-cell carcinomas (4 patients) and hyperkeratosis was the major adverse event observed. One malignant melanoma patient achieved a partial response, and six had stable disease with a median progression-free survival of 24 weeks for these 7 patients (range, 16.4-44.4 weeks).

CONCLUSIONS: GC1008 had no dose-limiting toxicity up to 15 mg/kg. In patients with advanced malignant melanoma and renal cell carcinoma, multiple doses of GC1008 demonstrated acceptable safety and preliminary evidence of antitumor activity, warranting further studies of single agent and combination treatments.

TRIAL REGISTRATION: Clinicaltrials.gov NCT00356460.

And that brings you to the end of your journey for today!

Before we go, what did we learn today?

- How to use QuantitativeChange to find proteins secreted because of melanoma disease.
- How to find cells inhibited by those secreted proteins.
- How to identify a subset of these cells that inhibit melanoma.
- Manipulate large networks to “close the loop”: Melanoma (A) secreted proteins (B) inhibit cells (C) which inhibit melanoma (A).
- Identification of clinical trials for specific diseases which involve TGFB1.
- Evaluation of clinical trial results for melanoma and a monoclonal antibody directed against TGFB1.

Study Questions 8

1. Identify the top five proteins (by # of references) that are secreted by melanoma. (Hint: Add: QuantitativeChange; QuantitativeType = secretion.)

2. Identify all cells that were inhibited by the five proteins from Question 1. How many of them? (Hint: Add: Regulation; Effect = negative.)

3. Identify all cells that inhibit melanoma. How many of them? (Hint: Negative regulation)

4. Find overlap of the cells from Questions 2 and 3. How many of them?

5. Identify the top five cells (by # of references) from Question 4 that are connected to the immune response.

For Proteins/Small Molecules Involved in Chemical Interactions, How Do I ...

Exercise 8.1: What enzymes are involved in a chemical reaction with a small molecule?

Identifies functional classes and proteins that catalyze chemical reactions of small molecules. Most metabolism enzymes in the metabolism pathways are represented by functional diseases.

Step 1: Create New Pathway or within Pathway, select a protein(s)

Step 2: Select Add- > Network Builder

Step 3: Select Expand Pathway, Advanced Expand Pathway Tool

Step 4: Select Direction: All

Step 5: Entities: Select “Protein, functional classes” Relations: Select “ChemicalReaction”

Continue to Network Builder Step 5 and click Launch Interactive Network Builder or Finish

Module 9

Finding Support for Your Hypothesis

Contents

9.1 Introduction

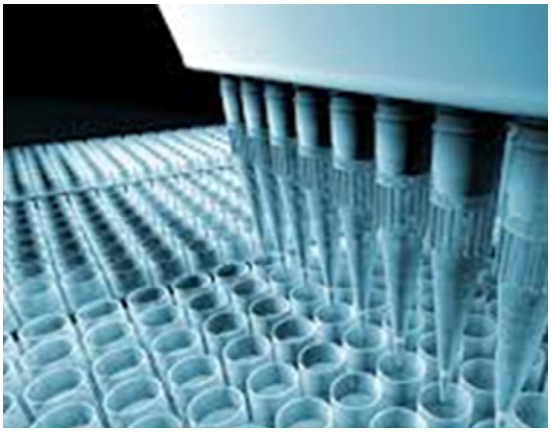
Let's say you have performed an assay or are analyzing genomic high-throughput data and now you have the answer!

But you'd like to know whether this finding is novel (i.e. not previously noted in the scientific literature)

But if it is novel, that's wonderful, but you now have a new problem!

How to get literature support for a novel finding? (since by definition, there will be no direct mention of this finding in the current literature).

Well here's a way!



Let's say you've performed a massive Genome Wide Association Study (GWAS)

In genetics, a genome-wide association study (GWA study, or GWAS…is an examination of a genome-wide set of genetic variants in different individuals to see if any variant is associated with a trait.

Wikipedia contributors. "Genome-wide association study." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 9 Mar. 2017. Web. 24 Mar. 2017.

9.2 A Real World Example

You've done a lot of work, getting the samples, performing the assay, processing and analyzing the data, and now you have an answer.

And it is....

Glutamate Metabotropic Receptor 3

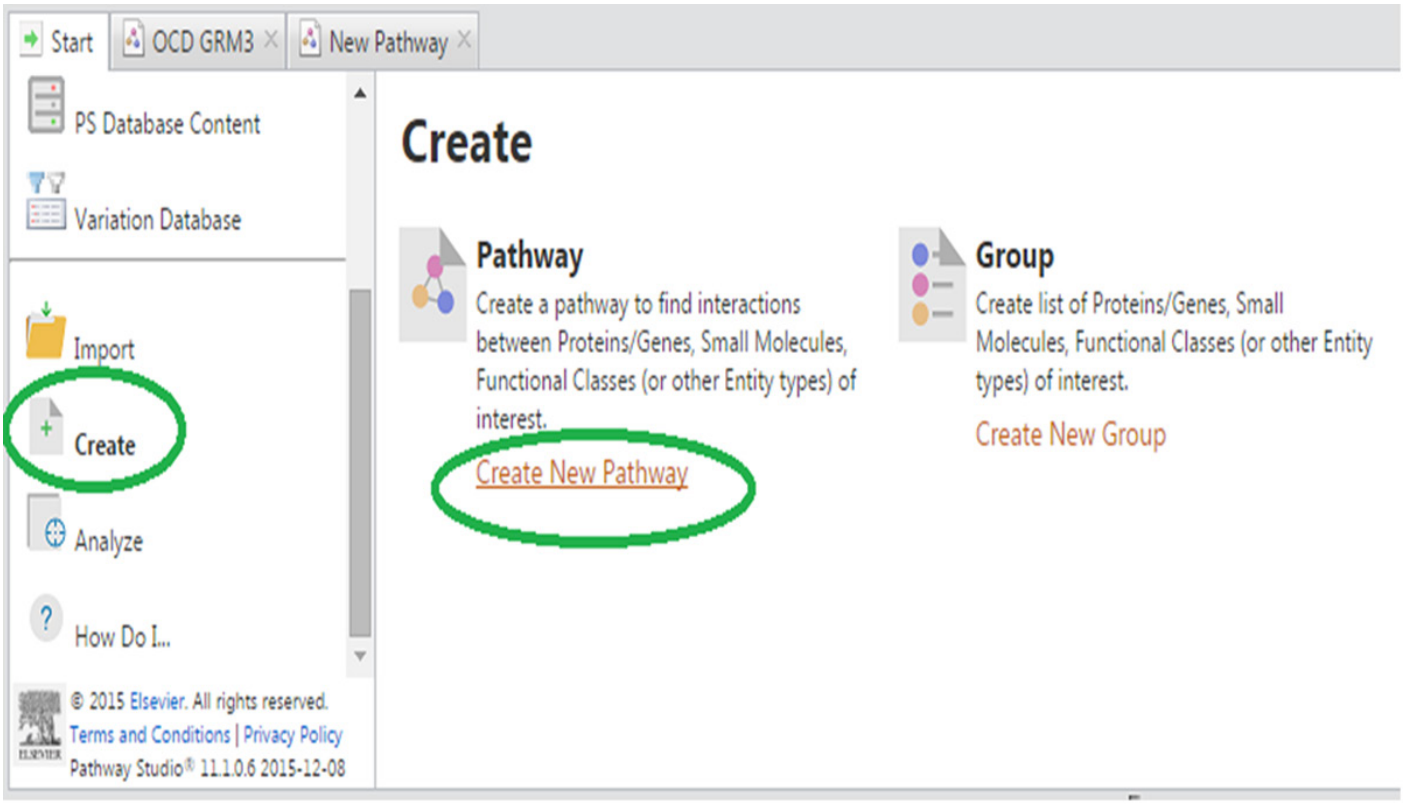
Glutamatergic neurotransmission is involved in most aspects of normal brain function.*

***Based on a real world example**

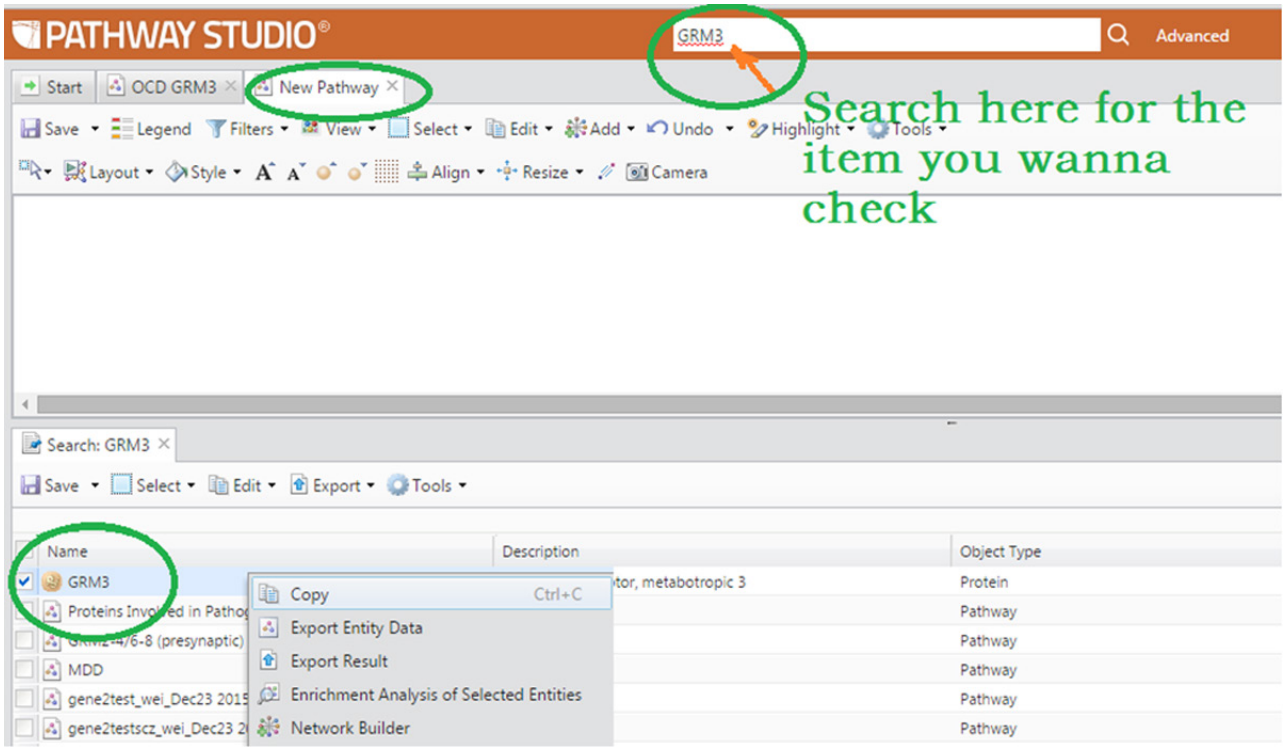
So, let's go to Pathway Studio and look for any information which might connect the GRM3 gene/protein with Obsessive-Compulsive Disorder (OCD):

9.3 Direct Interactions

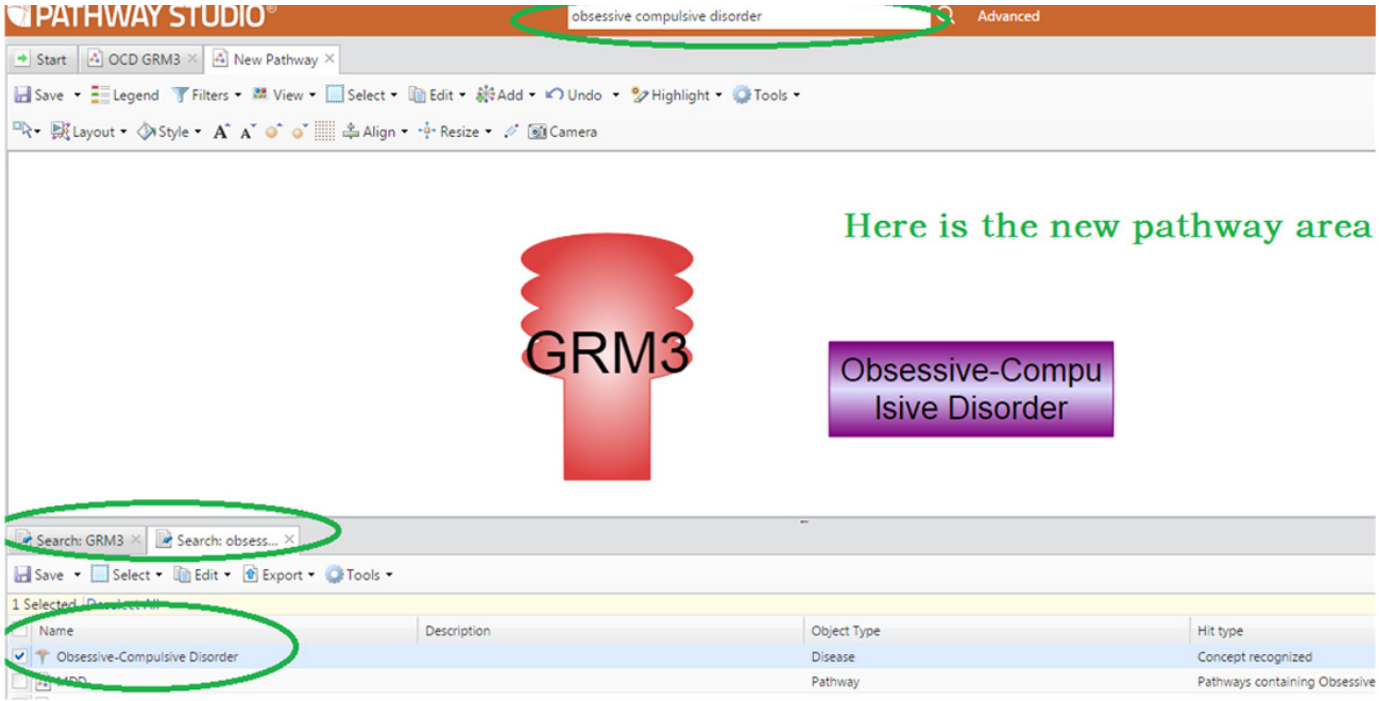
Step 1: Open a new Pathway...



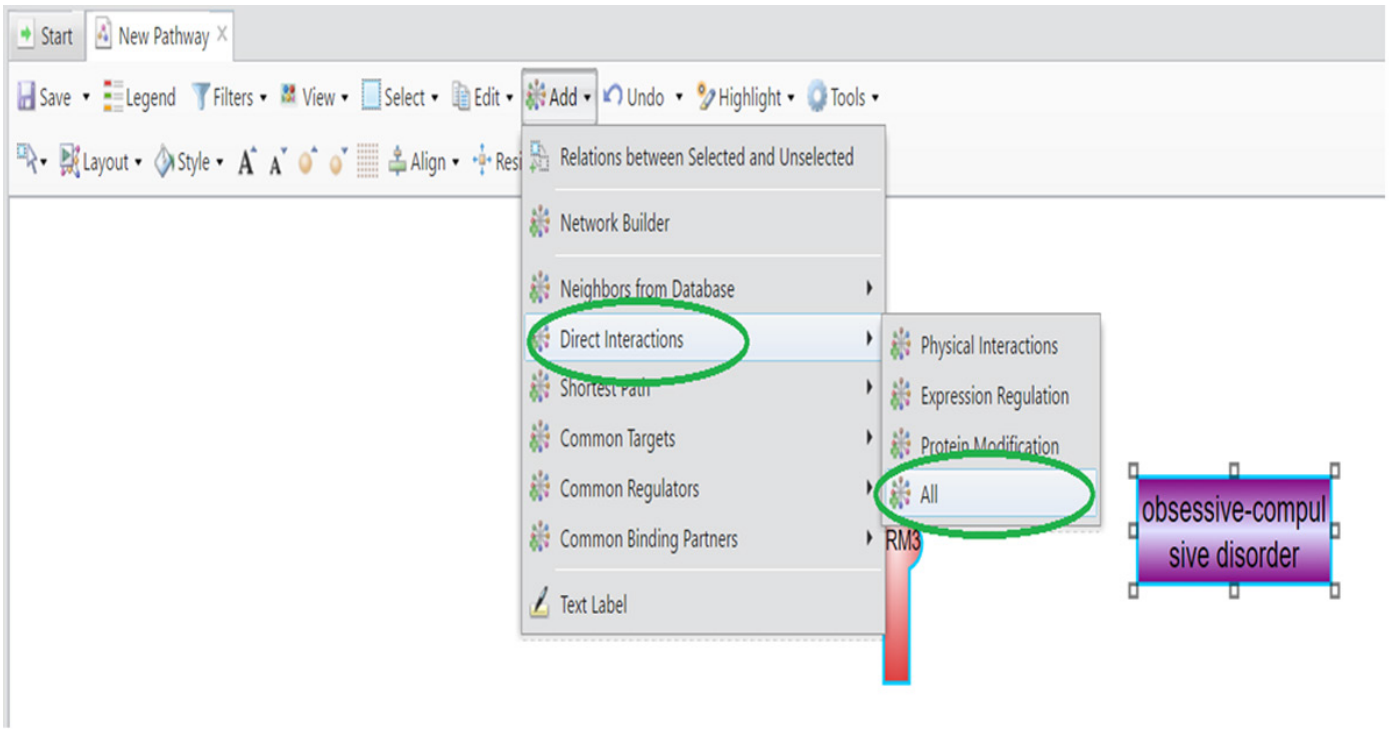
Step 2. Search and copy the item...



Step 3. Paste GRM3 to 'new Pathway area' and repeat the process for OCD.

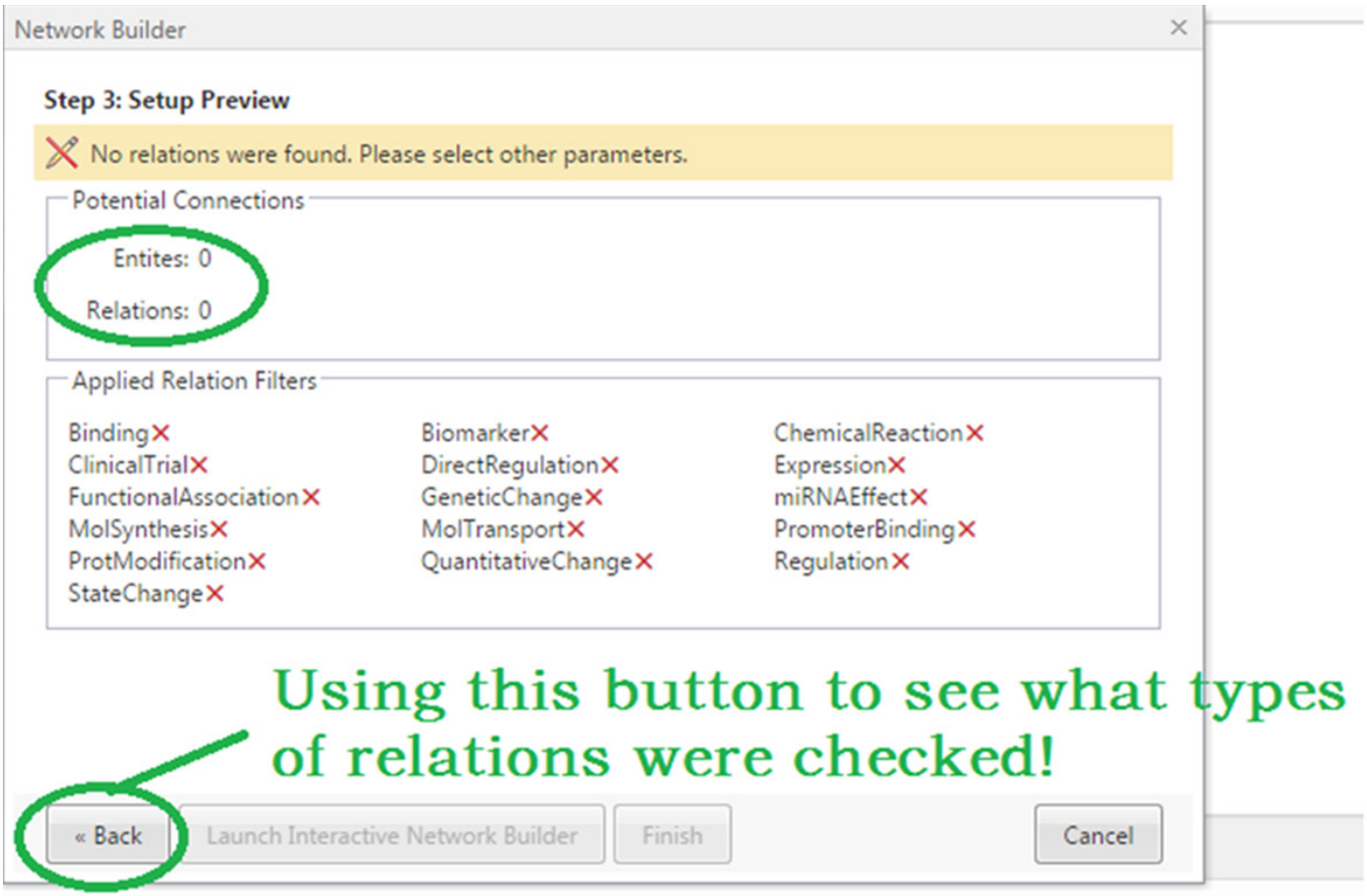


Step 4. Select both items and the press button 'Add'...



Step 5. The pop-up window shows the results:

Entities: Zero and **Relations:** Zero indicates that no direct connection between GRM3 and OCD has been detected in the literature!



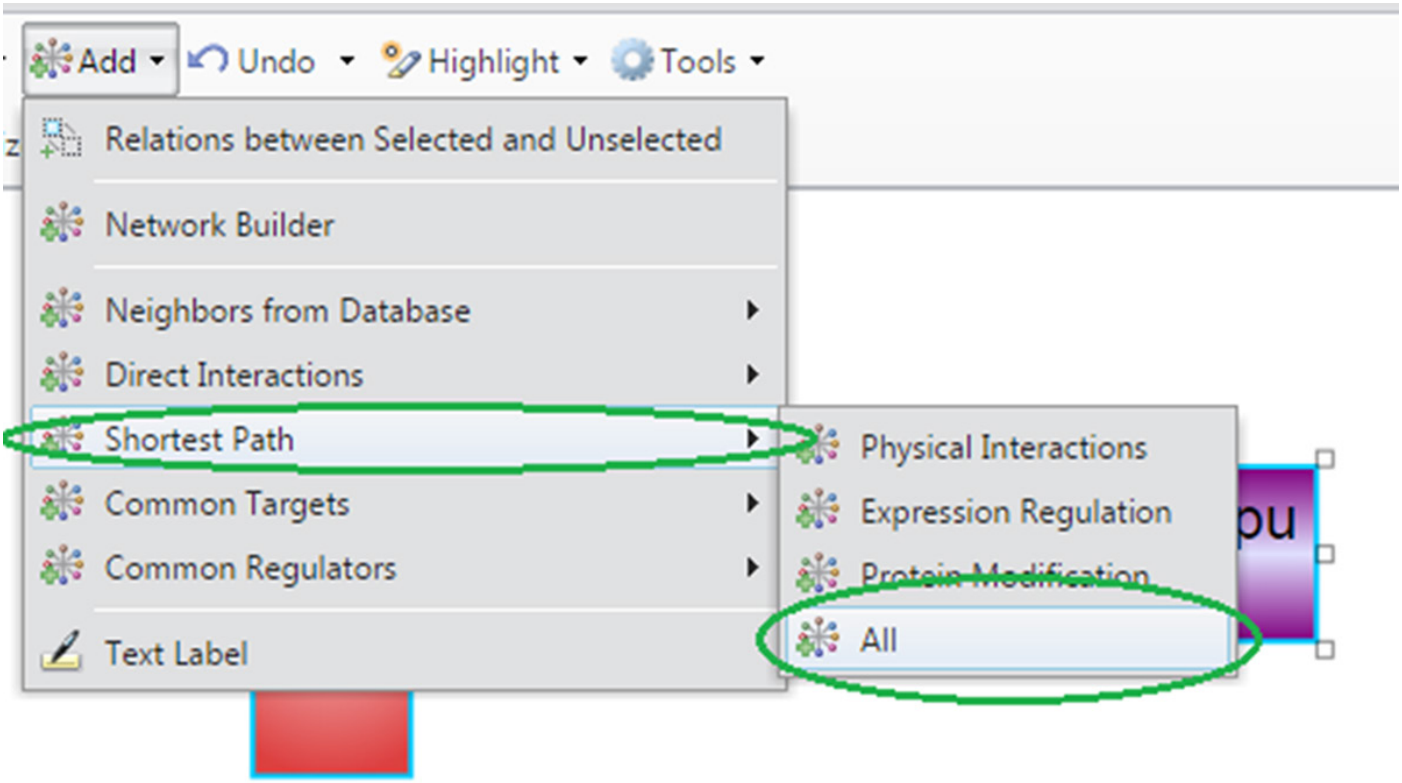
What to do?

9.4 Shortest Path Between Two Entities

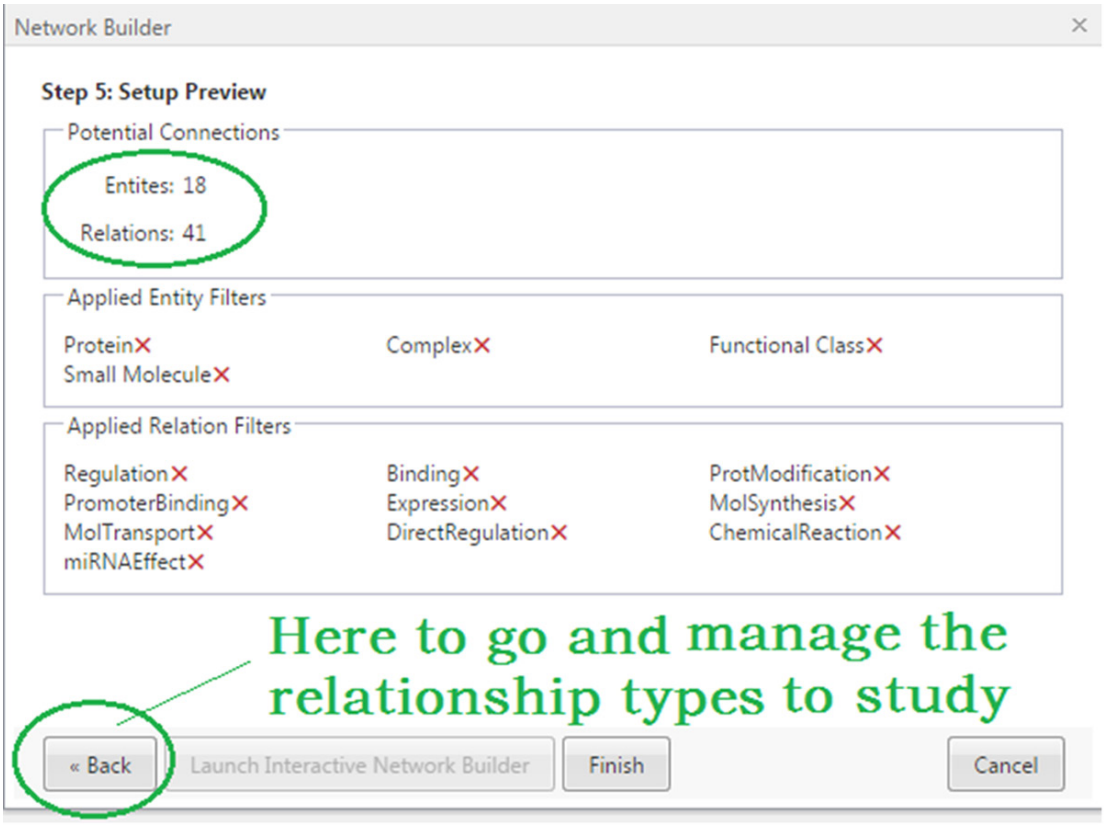
Let's look for the "shortest path" between two entities.

(Note: If there is a direct connection between two entities then that is the shortest path; if not, then the software will look for connections requiring one intervening step, increasing steps as necessary until a connection is made.)

Step 6. Select both entities (GRM3 and OCD) and click on **"Add, Shortest Path, All".**

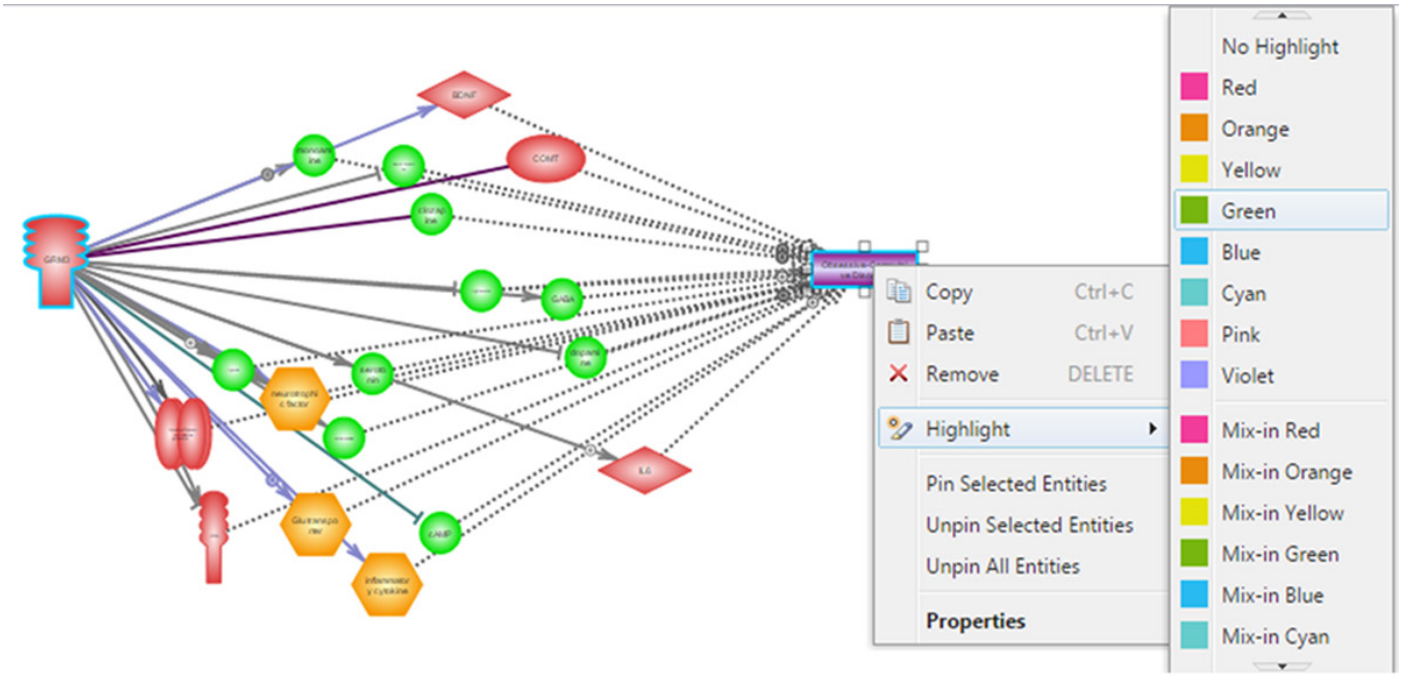


Step 7. Search results for 'Shortest Path'.



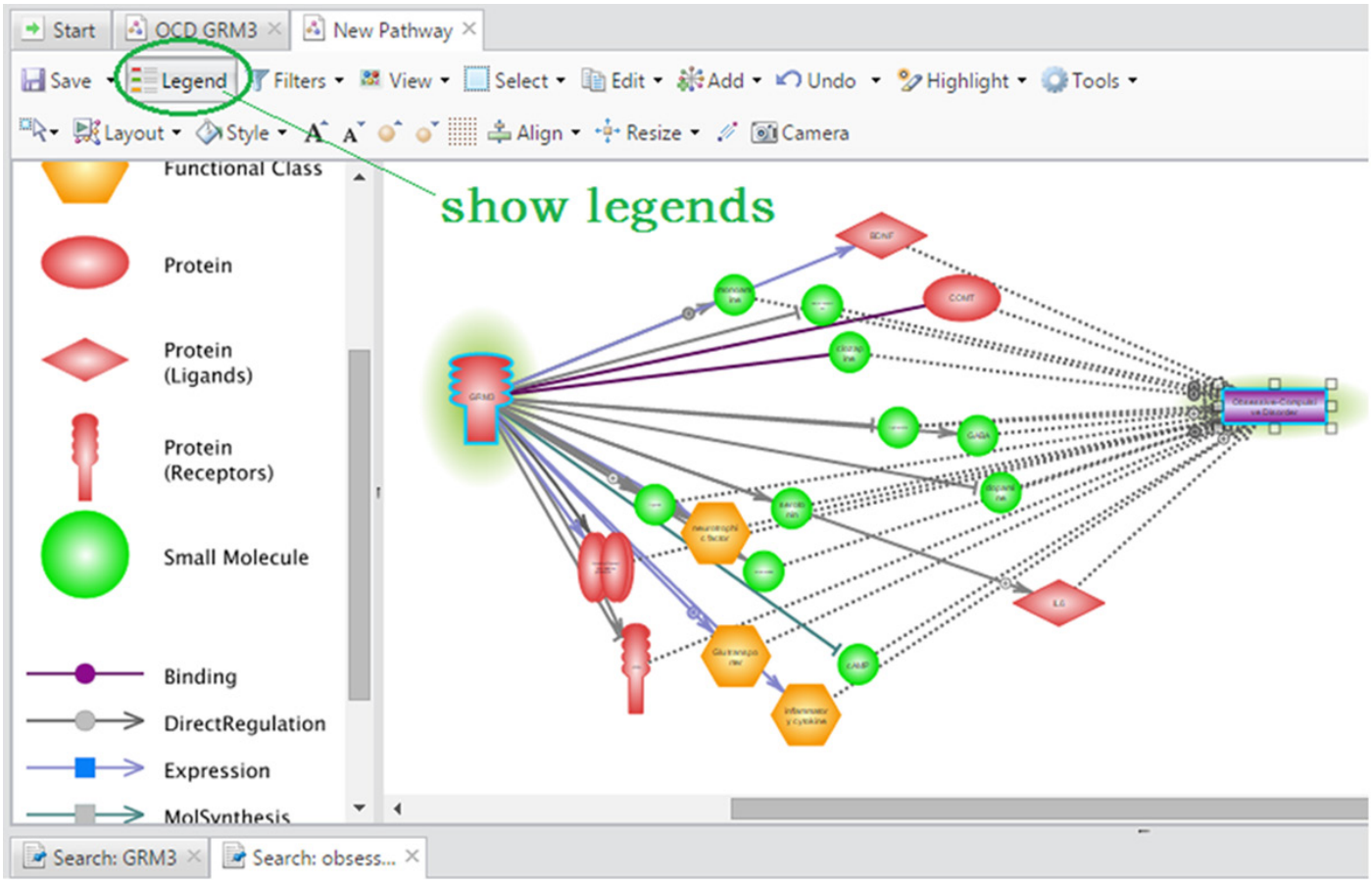
Entities: 18 and **Relations: 41** indicates that the indirect connections between GRM3 and OCD involves a total of 18 Entities and 41 Relations.

Step 8. Click 'Finish' and highlight GRM3 and OCD.

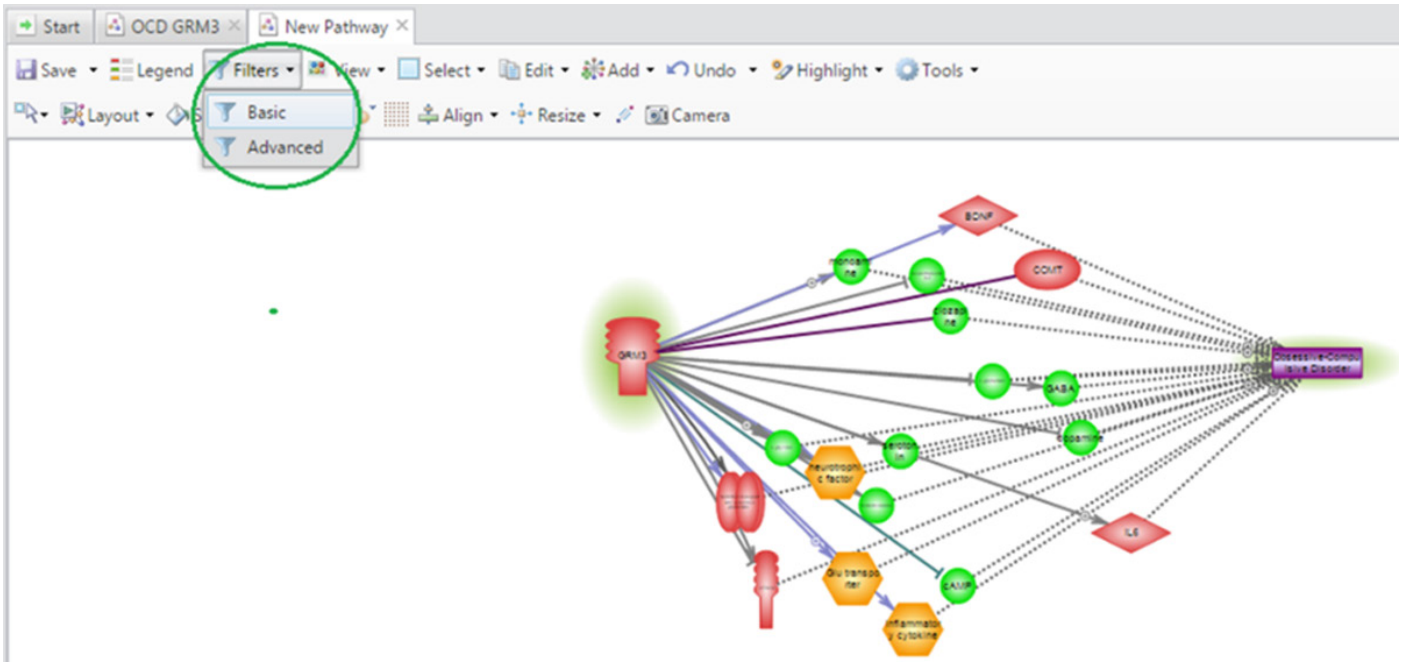


9.5 Show the Legend

Step 9. Show the legend for the entities and relations.

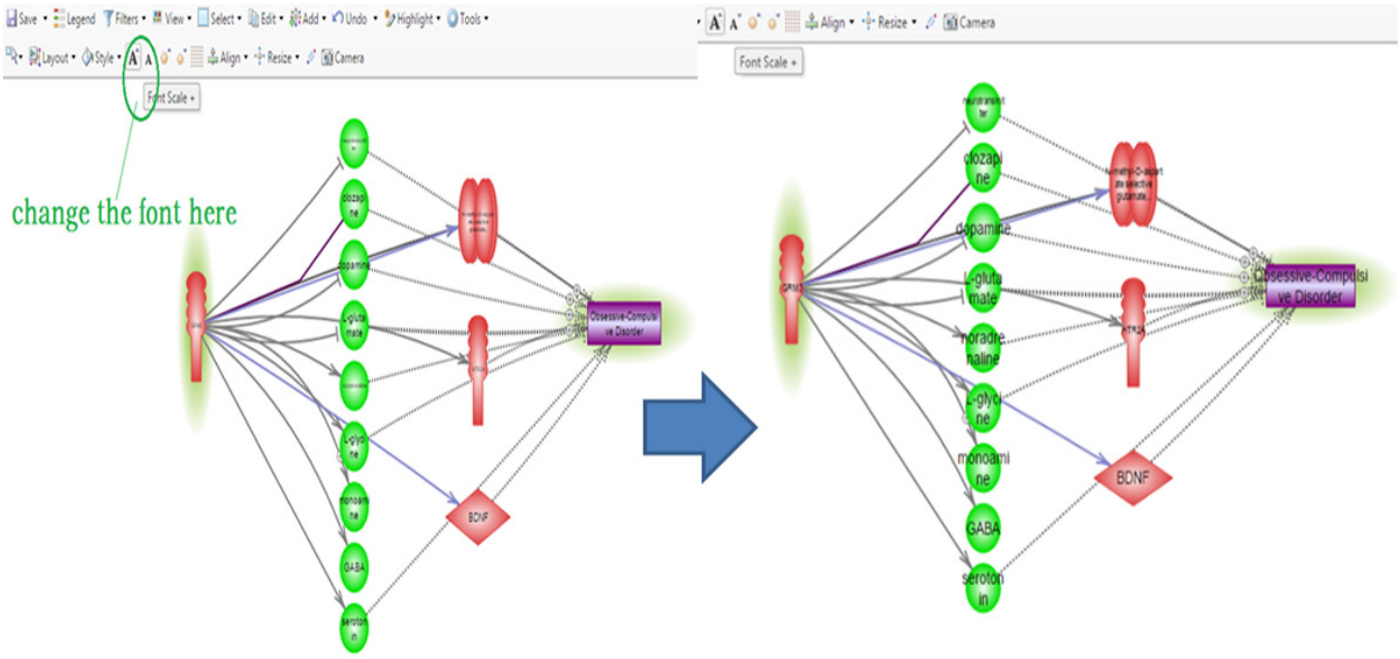


Step 10. Using 'Filter' to manage the relations and entities (for this example, filter by Ref # >=3, proteins and small molecules only, then remove entities which no longer connect OCD and GRM3).

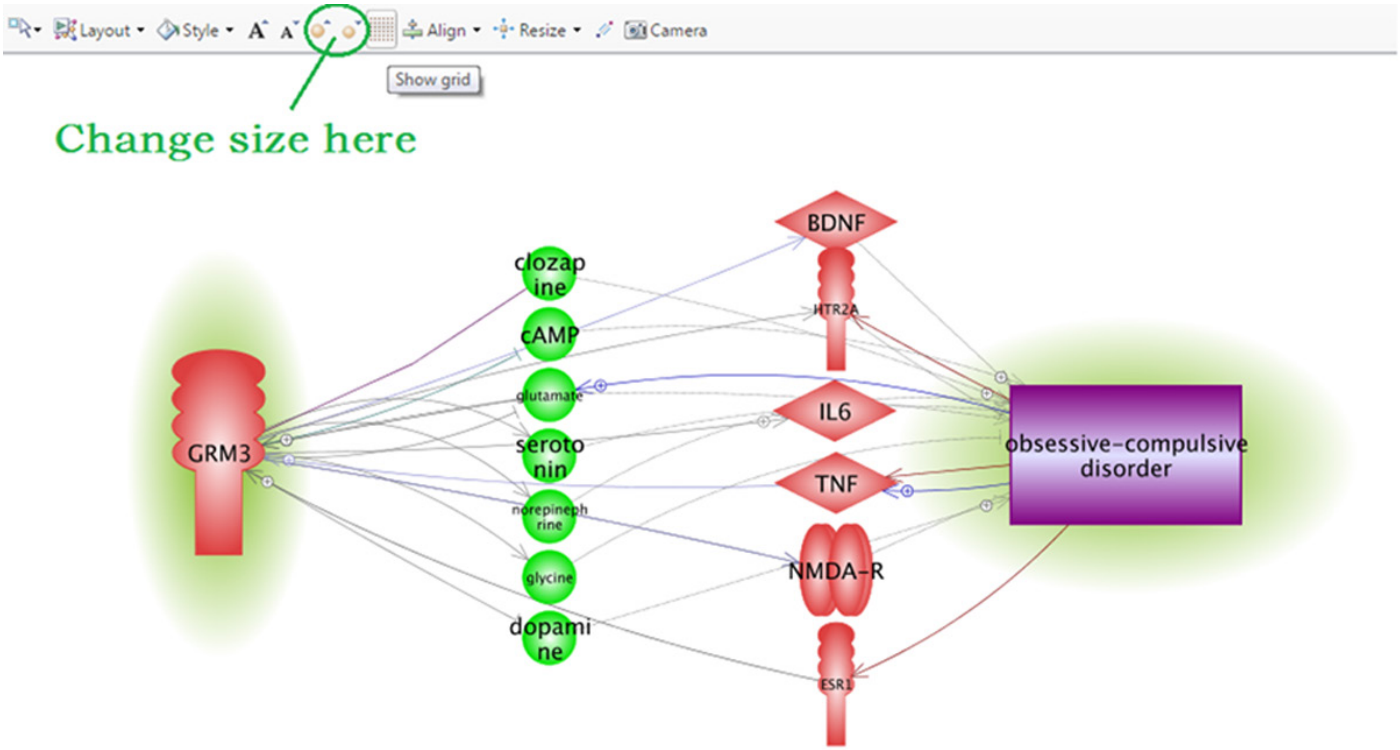


9.6 Adjust Font and Object Size

Step 11. Adjust the font size.

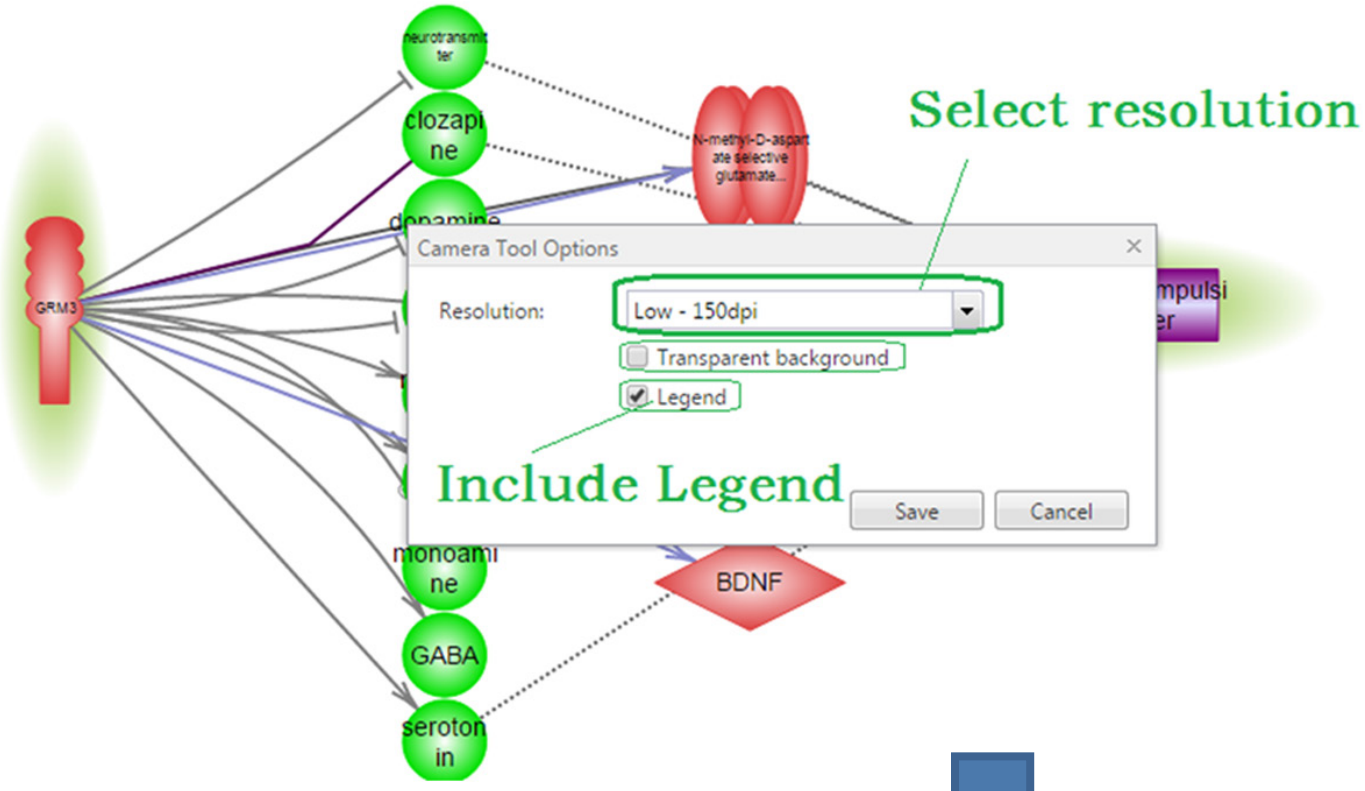


Step 12. Adjust the object size.

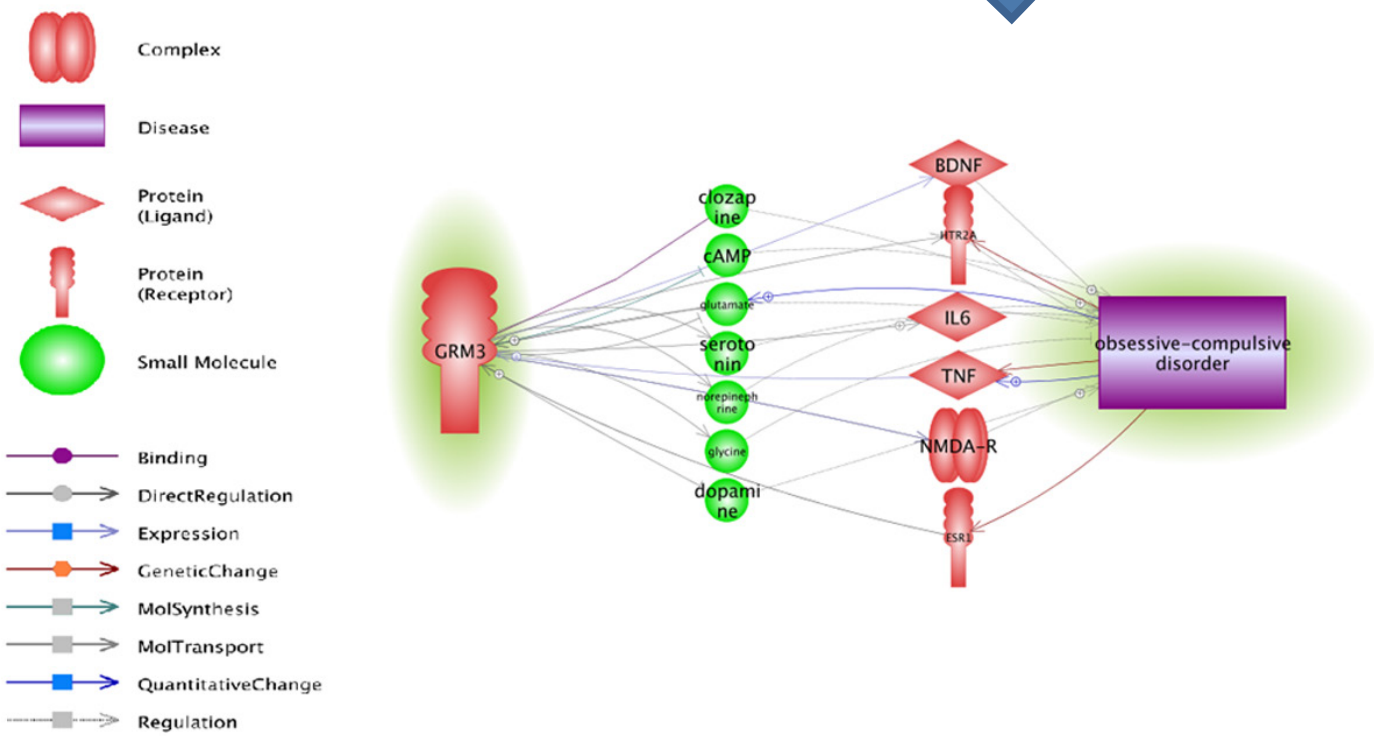


9.7 Save the Image

Step 13. Save the image.

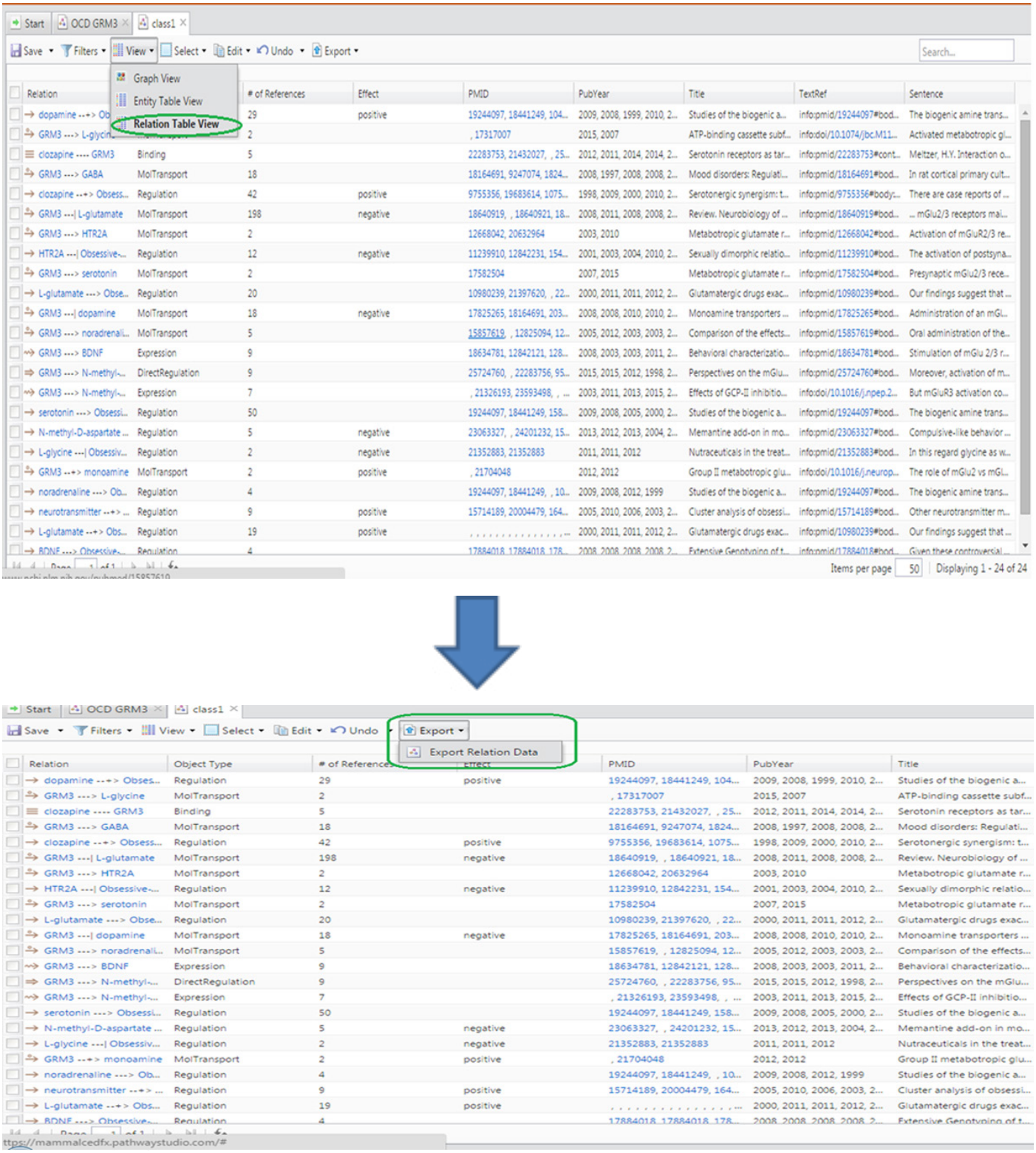


The saved image is going to look like this:



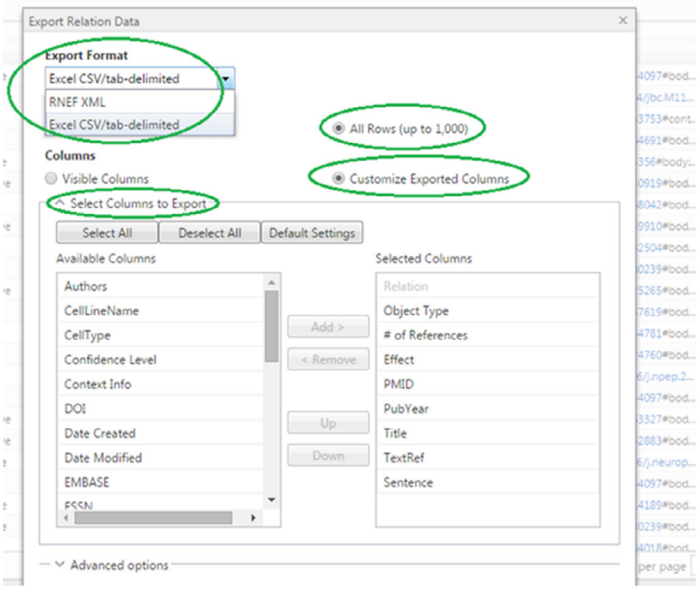
9.8 Export the Relation Table

Step 14. Review and export the “Relation Table”.



Step 15. Customize the output.

- **Select output format: ‘CVS’ or ‘XML’**
- **Select the columns of the output**



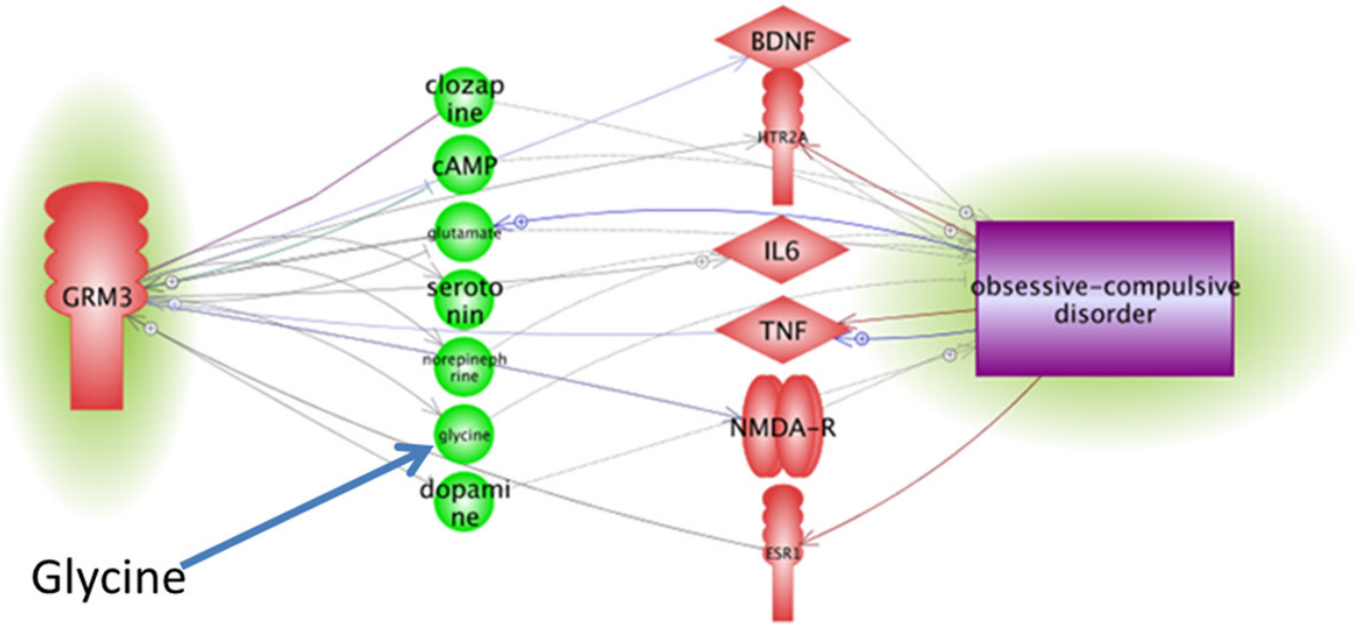
Step 16. View the saved Table.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	RelationSymbolicName	Effect	Relation	PubYear	TextRef	Authors	Journal	nsrc	Title													
2	MolTransport: GRM3 ---> glycine		2	2007;201	info:pmi	Malone, S.	Pain: Jou	ID{134=A	The antinociceptive effect of 2-chloro-2'- C-methyl-N 6-cyclopentyladenosine (2'-Me-CCPA), a highly selective adenosine A 1													
3	negative Regulation: glycine ---| obsessive-compulsive disorder	negative	2	2011;201	info:pmi	Canfield,	Progress In this		Nutraceuticals in the treatment of Obsessive Compulsive Disorder (OCD): A review of mechanistic and clinical evidence; Nutrac													
4	MolTransport: GRM3 ---> serotonin		2	2007;201	info:pmi	Palucha,	Pharmaco	Presynap	Metabotropic glutamate receptor ligands as possible anxiolytic and antidepressant drugs; When is a proof-of-concept (POC) not													
5	positive MolTransport: GRM3 --+) IL6	positive	2	2005;200	info:pmi	Aronica,	Neurosci	The ID{2	Activation of metabotropic glutamate receptor 3 enhances interleukin (IL)-1β-stimulated release of IL-6 in cultured human a													
6	GeneticChange: obsessive-compulsive disorder ---> ESR1		2	2011;201	info:pmi	Labad, J.	Journal	Although	Reproductive hormone sensitivity and obsessive-compulsive disorder: Are there differences in the genetic predisposition betw													
7																						
8	positive Regulation: cAMP --+) obsessive-compulsive disorder	positive	2	2002;200	info:pmi	Marazzit	Psychoneu	On the ba	Decreased inhibitory activity of PKC in OCD patients after six months of treatment; Altered cAMP-dependent protein kinase A i													
9	positive Expression: TNF --+) GRM3	positive	3	2012;201	info:doi	Berger, J.	Neurosci	Focusing	Opposite regulation of metabotropic glutamate receptor 3 and metabotropic glutamate receptor 5 by inflammatory stimuli in cu													
10	Regulation: norepinephrine ---> obsessive-compulsive disorder		3	2012;200	info:doi	Ari, M. ;	Journal	Both irr	Serum adiponectin and resistin levels in patients with obsessive compulsive disorder; Studies of the biogenic amine transport													
11	positive QuantitativeChange: obsessive-compulsive disorder --+) TNF	positive	4	2012;200	info:doi	Unsal, C.	Journal	Increase	Low plasma adiponectin levels in panic disorder; A cytokine study in children and adolescents with Tourette's disorder; Cytoki													
12	GeneticChange: obsessive-compulsive disorder ---> TNF		4	2008;200	info:pmi	Hounie, A.	Neurosci	The sing	TNF-alpha polymorphisms are associated with obsessive-compulsive disorder; TNF-alpha polymorphisms are associated with obsess													
13	Expression: GRM3 ---> NMDA-R		5	2003;201	info:doi	Carpente	Neuropep	But ID{2	Effects of GCP-II inhibition on responses of dorsal horn neurones after inflammation and neuropathy: an electrophysiological													
14	MolTransport: GRM3 ---> norepinephrine		5	2005;201	info:pmi	Lorrain, D	Neuropha	Oral adm	Comparison of the effects of diazepam, the CRF1 antagonist CP-154,526 and the group II mGlu receptor agonist LY379268 on str													
15	Regulation: BDNF ---> obsessive-compulsive disorder		5	2008;200	info:pmi	Alonso, P.	Biologic	Given th	Extensive Genotyping of the BDNF and NTRK2 Genes Define Protective Haplotypes Against Obsessive-Compulsive Disorder: Extensiv													
16	negative Regulation: NMDA-R ---| obsessive-compulsive disorder	negative	6	2012;201	info:pmi	Ghaleiha,	Journal	Compulsi	Memantine add-on in moderate to severe obsessive-compulsive disorder: Randomized double-blind placebo-controlled study; Gluta													
17	Binding: GRM3 ---- clozapine		7	2012;201	info:pmi	Meltzer, H	Current	Meltzer,	Serotonin receptors as targets for drugs useful to treat psychosis and cognitive impairment in schizophrenia; The novel objec													
18	DirectRegulation: GRM3 ---> NMDA-R		8	2015;201	info:enh	Purkayas	Current	Group 2	A review on GABA/glutamate pathway for therapeutic intervention of ASD and ADHD; Glutamatergic system controls synchronizatio													
19	positive DirectRegulation: ESR1 --+) GRM3	positive	9	2010;201	info:pmi	Grove-St	Journal	Further,	Membrane estrogen receptors activate the metabotropic glutamate receptors mGluR5 and mGluR3 to bidirectionally regulate CREB													
20	Expression: GRM3 ---> BDNF		9	2008;200	info:pmi	Bespalov,	European	Stimulat	Behavioral characterization of the mGlu group II/III receptor antagonist, LY-341495, in animal models of anxiety and depress													
21	positive Regulation: glutamate --+) GRM3	positive	9	2005;200	info:pmi	Moran, M.	M. ; McFarl	In the first experiment, the capacity of ID{1197745=glutamate} derived from xc- to regulate excitatory transmission via ID{0,2912,291														
22	positive DirectRegulation: glutamate --+) GRM3	positive	13	2011;201	info:pmi	Ryan, P. J.	Neurosci	However,	Nucleus incertus-An emerging modulatory role in arousal, stress and memory; Using human brain imaging studies as a guide towa													
23	negative Regulation: HTR2A ---| obsessive-compulsive disorder	negative	13	2001;200	info:pmi	Enoch, M.	Biologic	The acti	Sexually dimorphic relationship of a 5-HT 2A promoter polymorphism with obsessive-compulsive disorder; How to treat OCD in pa													

9.9 Reality Check

Do does this really work?

Let's take a closer look at some of our results



Take glycine, for example; we can examine the two relations which connect glycine to both GRM3 and OCD.

GRM3 → Glycine

Activated ID{2913=metabotropic glutamate receptor 3} inhibits subsequent ID{1197745=glutamate} and ID{1178899=glycine} release (Robert et al. 2015).

Glycine → OCD

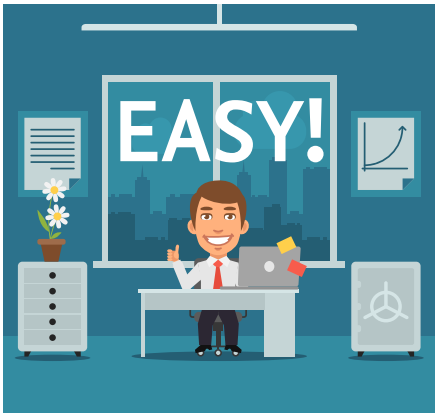
The ID{7000363=patient} was administered ID{1178899=glycine} over a 5-year period, which led to a large reduction in OCD symptoms, and resumption of education and social life (Camfield et al., 2011).

So, activated GRM3 inhibits glycine release, and the administration of glycine proved to be an effective treatment for OCD in at least one documented case.

Thus, variations in the GRM3 gene may affect glycine release that, in turn, impact the symptoms of OCD.

And now you see that GRM3 and OCD are indeed connected!

That was .....



Before you go, what did you learn today?

- You learned how to connect two entities not directly connected in the literature by using the "Add, Shortest Path" option.
- You learned how to make publication-quality images using a combination of highlighting, filtering, and by adjusting font and object size.
- You learned how to save a high resolution image.
- You learned how to export a relation table.
- You reviewed evaluating an indirect connection in order to see whether it made biological sense.

Study Questions 9

1. What is the most studied small molecule connection between GRM3 and OCD?

2. What specific "pro-inflammatory cytokines" play a role in OCD? (Hint: Use "Relation Table View" to find relations and references for this functional class and OCD.)

3. GRM3 influences which member of the nerve growth factor family of proteins?

4. GRM3 influences which member of the dopamine receptor family of proteins?

5. Find cell processes that are influenced by GRM3 and also linked with OCD. How many of them? (Hint: Add Cell Processes separately to GRM3 and OCD, then check overlap.)

6. Find cells that are influenced by GRM3 and also linked to OCD. (Hint: This is the same as for Question 5.) How many of them? Name four of these cell types.

For Protein/Small Molecule Association with Diseases and Cell Processes, How Do I ...

Exercise 9.1: What proteins are known to be associated with a disease or cellular process?

Identifies proteins known to be associated with a specific disease or cellular process. (More specific data relating proteins to diseases is available in DiseaseFx data including StateChange, GeneticChange and QuantitativeChange.)

Step 1: Create New Pathway or within Pathway, select disease or cell process

Step 2: Select Add- > Network Builder

Step 3: Select Expand Pathway, Advanced Expand Pathway Tool

Step 4: Select Direction: Upstream

Step 5: Entities: Select “Protein” Relations: Select "Regulation"

Continue to Network Builder Step 5 and click Launch Interactive Network Builder or Finish

Exercise 9.2: What small molecules are associated with a disease or cellular process?

Identifies small molecules that are associated with diseases or cellular processes. Small molecule association with diseases and cell processes through regulation relations are found in the ChemEffect Database. In addition, more information about small molecules associated with diseases can be found in the DiseaseFx database through QuantitativeChange and biomarker relations.

Step 1: Create New Pathway or within Pathway, select disease or cell process

Step 2: Select Add- > Network Builder

Step 3: Select Expand Pathway, Advanced Expand Pathway Tool

Step 4: Select Direction: Upstream

Step 5: Entities: Select "Small Molecules" Relations: "Regulation"

Continue to Network Builder Step 5 and click Launch Interactive Network Builder or Finish

Exercise 9.3: What proteins are known to change in expression, activity, or abundance in a disease?

Identifies proteins that are changed in activity abundance or expression in a disease. QuantitativeChange relations are found only in DiseaseFx data.

Step 1: Create New Pathway or within Pathway, select disease

Step 2: Select Add- > Network Builder

Step 3: Select Expand Pathway, Advanced Expand Pathway Tool

Step 4: Select Direction: Downstream

Step 5: Entities: Select "Protein" Relations: Select "QuantitativeChange"

Add condition: Quantitative Type is equal to: expression, abundance, activity

Continue to Network Builder Step 5 and click Launch Interactive Network Builder or Finish

Exercise 9.4: What small molecules are known to change in abundance in a disease?

Identifies small molecules that are changed in abundance in a disease. QuantitativeChange relations are found in DiseaseFx data.

Step 1: Create New Pathway or within Pathway, select disease

Step 2: Select Add- > Network Builder

Step 3: Select Expand Pathway, Advanced Expand Pathway Tool

Step 4: Select Direction: Downstream

Step 5: Entities: Select “Small Molecule” Relations: Select "QuantitativeChange"

Add condition: QuantitativeType is equal to : abundance

Continue to Network Builder Step 5 and click Launch Interactive Network Builder or Finish

Exercise 9.5: What proteins with genetic mutations are associated with a disease?

Identifies proteins with genetic changes (gene deletions, amplifications, mutations, epigenic changes, or methylation) associated with a disease. GeneticChange relations are found in DiseaseFx data.

Step 1: Create New Pathway or within Pathway, select disease

Step 2: Select Add- > Network Builder

Step 3: Select Expand Pathway, Advanced Expand Pathway Tool

Step 4: Select Direction: Downstream

Step 5: Entities: Select “Protein” Relations: Select "GeneticChange"

Add condition: ChangeType is equal to : gene deletions, amplifications, mutations, epigenic methylation

Continue to Network Builder Step 5 and click Launch Interactive Network Builder or Finish

Exercise 9.6: What proteins or small molecules are diagnostic for a disease?

Identifies proteins/small molecules known to be diagnostic for a disease. Biomarker relations are found in DiseaseFx data.

Step 1: Create New Pathway or within Pathway, select disease

Step 2: Select Add- > Network Builder

Step 3: Select Expand Pathway, Advanced Expand Pathway Tool

Step 4: Select Direction: Downstream

Step 5: Entities: Select "Protein and Small Molecules" Relations: Select "Biomarker"

Add condition: Biomarker Type is equal to : diagnostic

Continue to Network Builder Step 5 and click Launch Interactive Network Builder or Finish

Exercise 9.7: What proteins or small molecules are prognostic for a disease?

Identifies proteins/small molecules known to be prognostic for a disease. Biomarker relations are found in DiseaseFx data.

Step 1: Create New Pathway or within Pathway, select disease

Step 2: Select Add- > Network Builder

Step 3: Select Expand Pathway, Advanced Expand Pathway Tool

Step 4: Select Direction: Downstream

Step 5: Entities: Select "Protein and Small Molecules" Relations: Select "Biomarker"

Add condition: Biomarker Type is equal to : prognostic

Continue to Network Builder Step 5 and click Launch Interactive Network Builder or Finish

Exercise 9. 8 What protein phosphorylation/dephosphorylation events are associated with a disease?

Identifies post translational protein phosphorylation/dephosphorytation events associated with a disease. StateChange relations are found in DiseaseFx data.

Step 1: Create New Pathway or within Pathway, select disease

Step 2: Select Add- > Network Builder

Step 4: Select Direction: Downstream

Step 5: Entities: Select “Protein” Relations: Select "StateChange"

Add condition: ChangeType is equal to : phosphorylation

Continue to Network Builder Step 5 and click Launch Interactive Network Builder or Finish

Exercise 9. 9 What protein/gene splice variants are associated with a disease?

Identifies alternate gene splicing events/ splice variants associated with a disease. StateChange relations are found only in DiseaseFx data.

Step 1: Create New Pathway or within Pathway, select disease

Step 2: Select Add- > Network Builder

Step 3: Select Expand Pathway, Advanced Expand Pathway Tool

Step 4: Select Direction: Downstream

Step 5: Entities: Select “Protein” Relations: Select "StateChange"

Add condition: ChangeType is equal to : alternative splicing

Continue to Network Builder Step 5 and click Launch Interactive Network Builder or Finish

Module 10

Understanding a Rare Disease

Contents

10.1 Definition of a Rare Disease

So usually we would have an introduction outlining the description and the background of a disease which we are choosing to study more in depth at the molecular level (see, for example, Pre-eclampsia Disease, Module II).

A disease or disorder is defined as rare in the USA when it affects fewer than 200,000 Americans at any given time

- 80% of rare diseases have identified genetic origins whilst others are the result of infections.
- 50% of rare diseases affect children.

Excerpted from Rare Disease Day 2017

10.2 A Research Challenge

But what if you are a researcher and someone comes and asks you to come up to speed on a rare disease?

And they only give you a single clue as to what that rare disease might be!

SCA3*

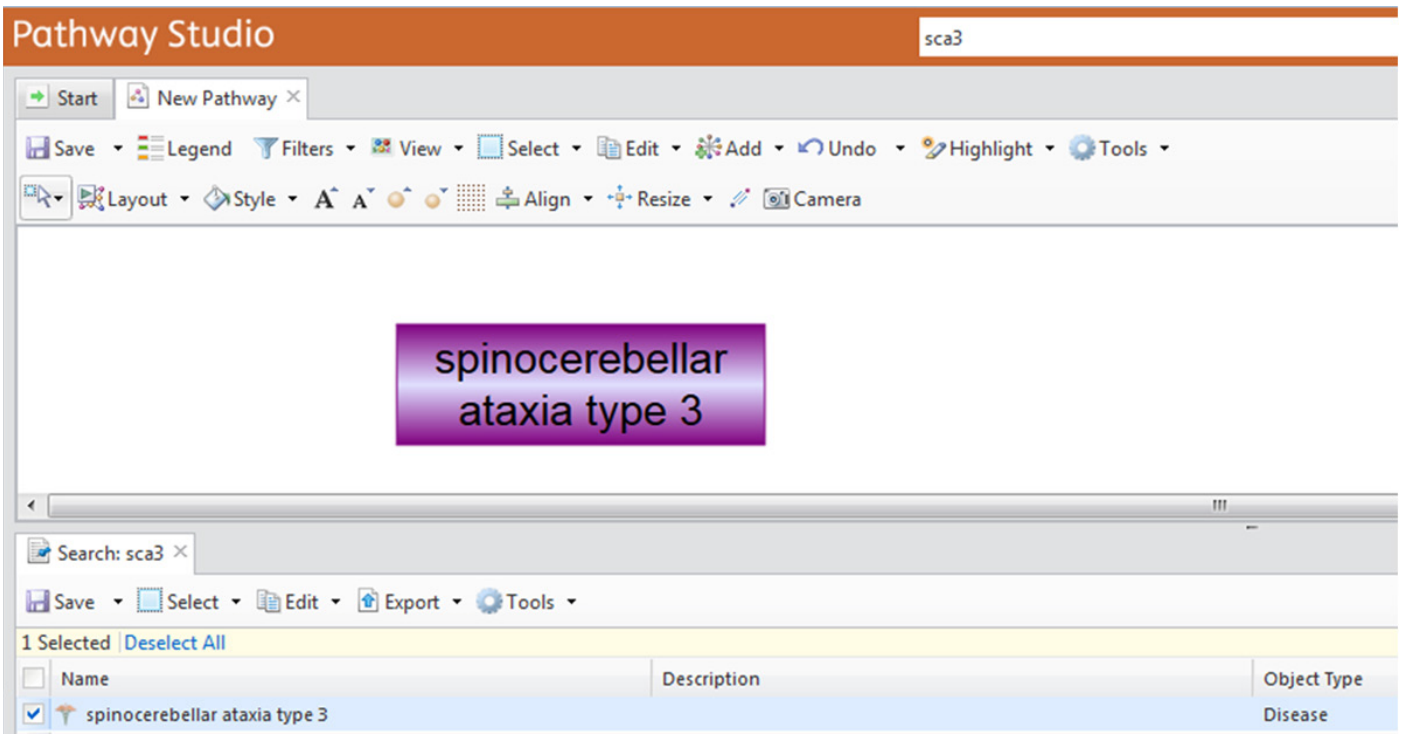
What do you say, should we take the challenge?

Yes, let's!

For this exercise let's agree to stay strictly within Pathway Studio and see how much we can learn about "SCA3" (in a very short period of time).

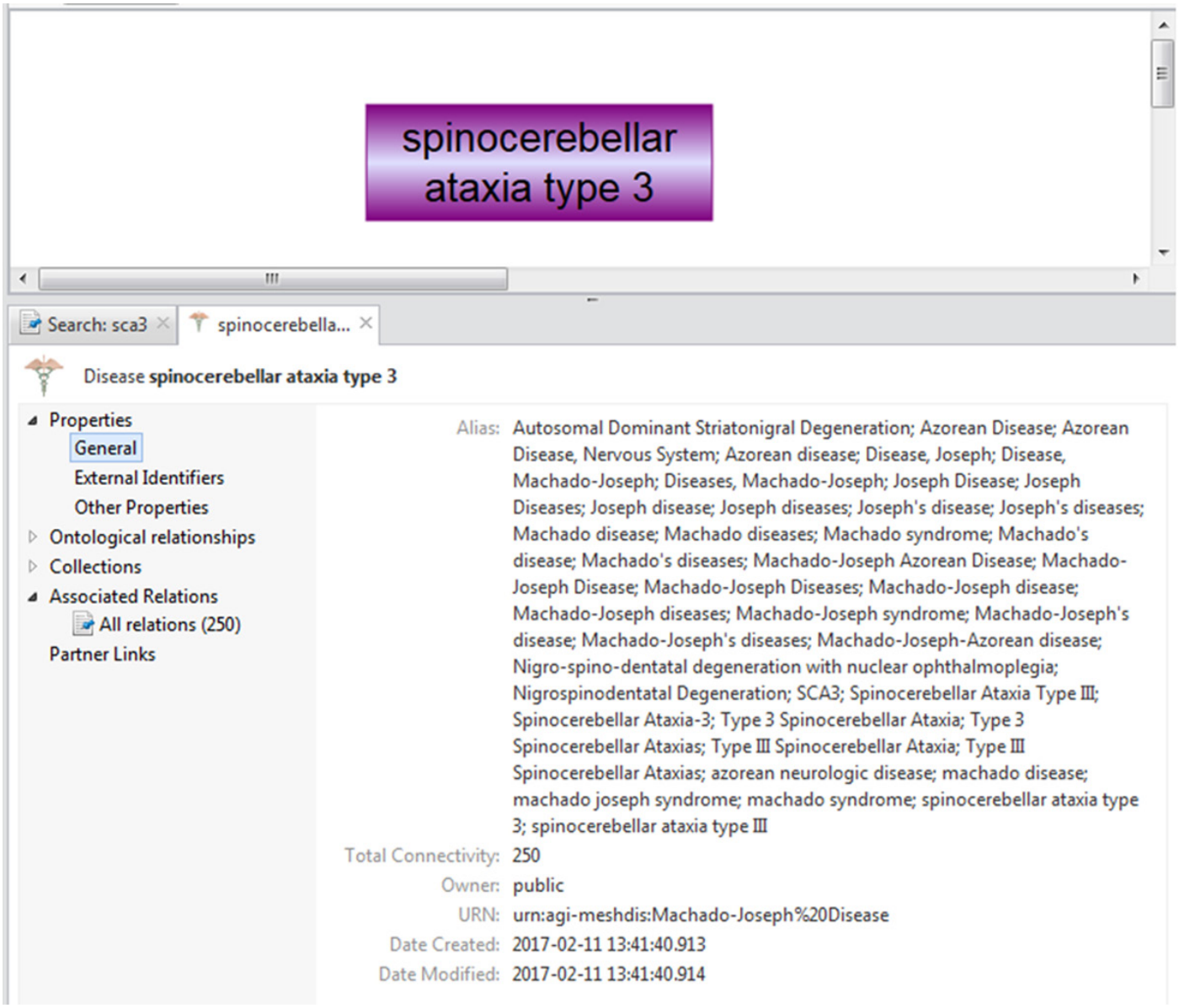
10.3 SCA3 = Spinocerebellar Ataxia Type 3

First Step, Search for **SCA3:**



And now you have a name: SCA3 = spinocerebellar ataxia type 3.

Now let's inspect the properties of this disease (double-click on disease icon).



Wow, this disease sure has a lot of different names, and Pathway Studio will recognize all of them in the scientific literature! This is exceptionally useful as it ensures that you are less likely to overlook important information.

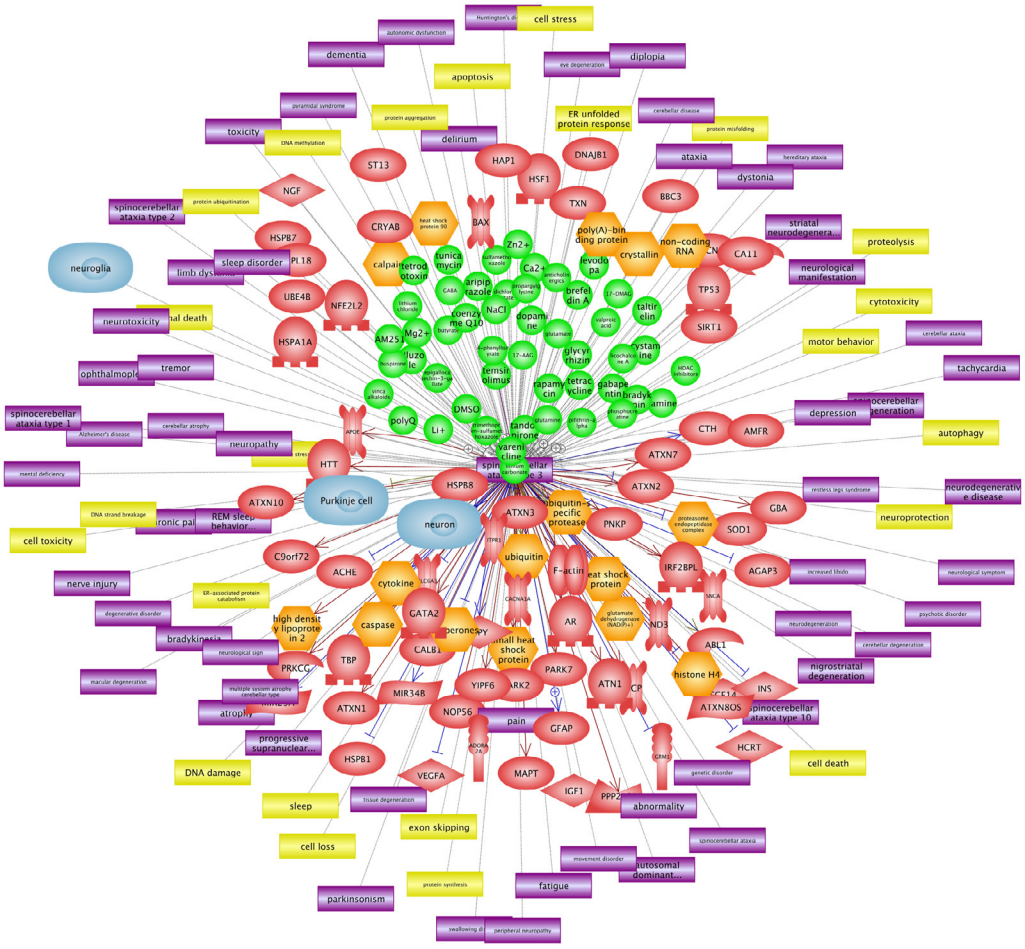
Don't Miss it!

10.4 SCA3, All Relations

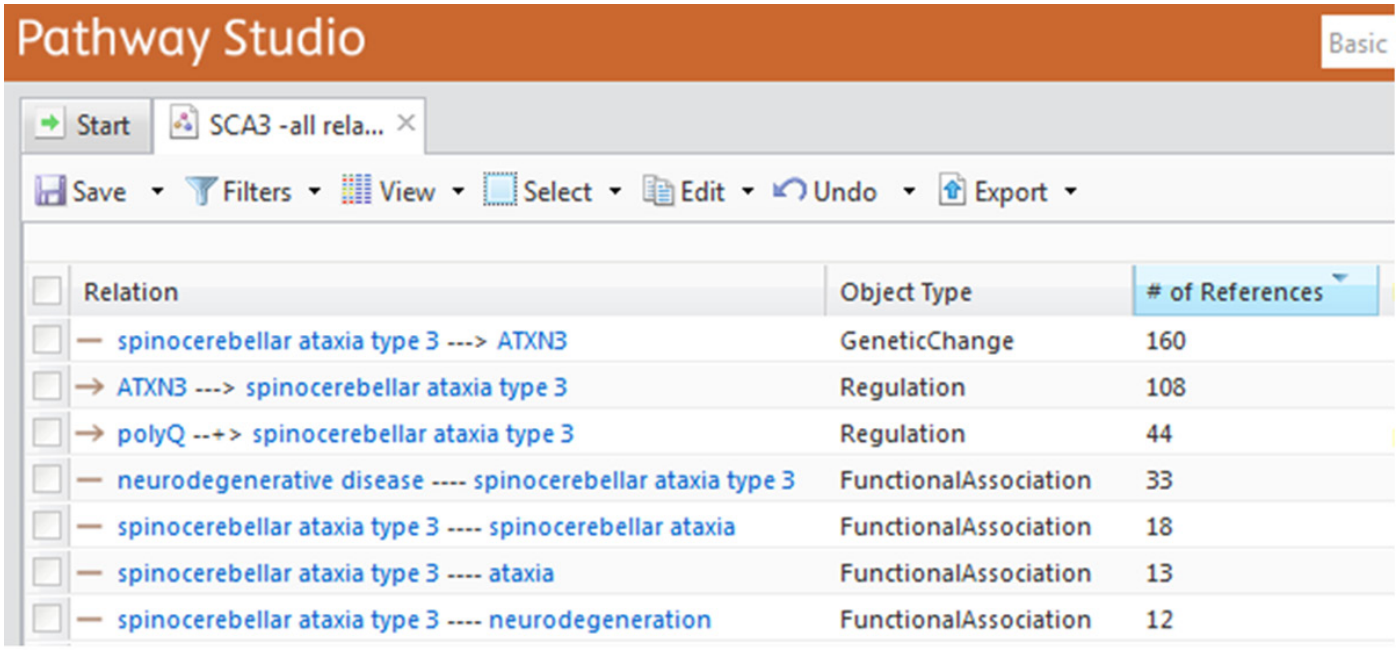
We note also that Pathway Studio has a total number of 250 unique facts or relations connected with SCA3 (double-click on all relations icon).

Let's examine them all!

Select All, Edit and Copy into workspace:



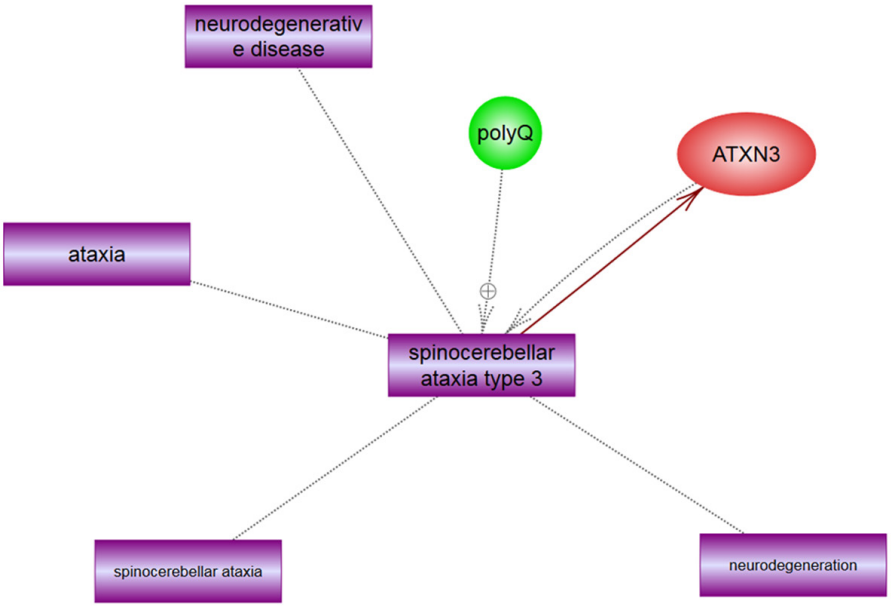
View Relation Table and sort by the numbers of references descending (relations with the most references will be at top).



Relation	Object Type	# of References
spinocerebellar ataxia type 3 ---> ATXN3	GeneticChange	160
ATXN3 ---> spinocerebellar ataxia type 3	Regulation	108
polyQ --+> spinocerebellar ataxia type 3	Regulation	44
neurodegenerative disease ---- spinocerebellar ataxia type 3	FunctionalAssociation	33
spinocerebellar ataxia type 3 ---- spinocerebellar ataxia	FunctionalAssociation	18
spinocerebellar ataxia type 3 ---- ataxia	FunctionalAssociation	13
spinocerebellar ataxia type 3 ---- neurodegeneration	FunctionalAssociation	12

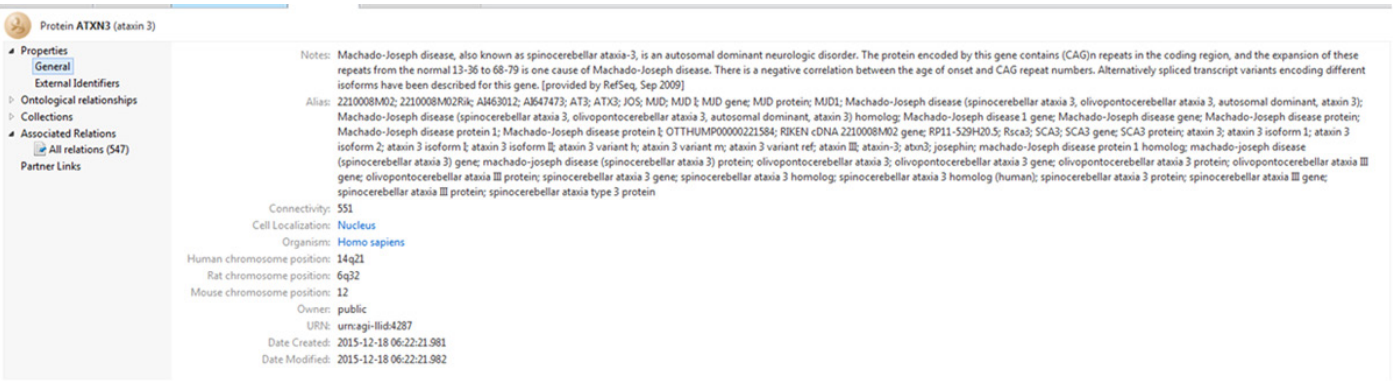
10.5 Disease Description

We see four diseases, one small molecule, and one protein all of which are strongly linked to SCA3. We can tell right away that SCA3 is:



- A subset of spinocerebellar ataxia diseases, in general.
- And is characterized by neurodegeneration and ataxia, defined as impairment of the ability to perform smoothly coordinated voluntary movements (drill down on properties of ataxia diseases, click on External Identifiers, MeSH Heading).

Double-click on the protein: ATXN3



Wow, this protein has a lot of different names (as is often the case!) one of which is Machado-Joseph disease (another common name for SCA3 disease).

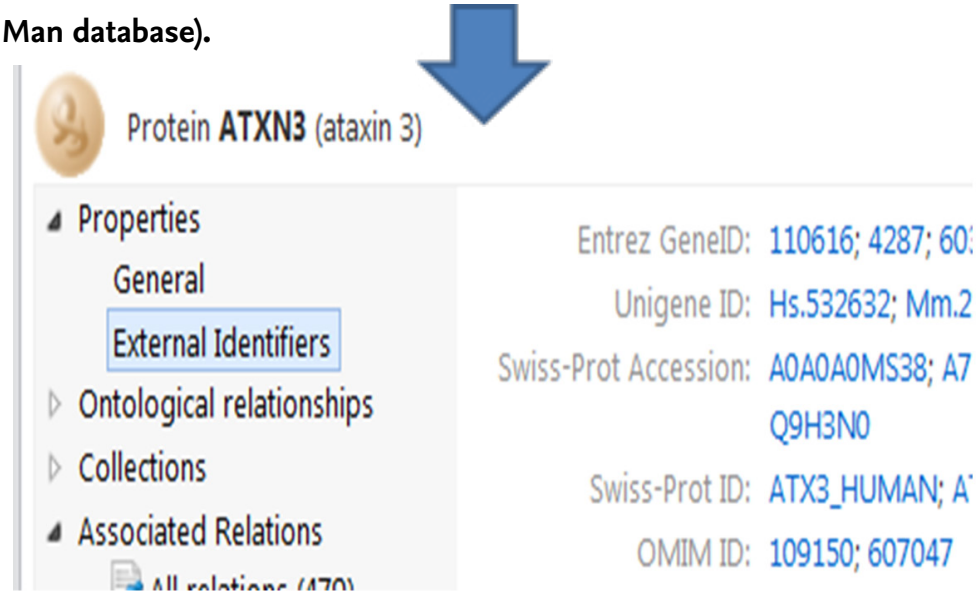
spinocerebellar ataxia type 3 ---> ATXN3	GeneticChange	160
ATXN3 ---> spinocerebellar ataxia type 3	Regulation	108

Machado-Joseph disease, also known as spinocerebellar ataxia -3.

10.6 ATXN3: a Mendelian Gene Disorder

We can kind of figure this protein is central to the disease based on the 160 references indicating that a GeneticChange in ATXN3 is associated with SCA3 and also that 108 references indicate that ATXN3 regulates the disease as well.

Let's click on External identifiers and we will see that there is a link to OMIM (the Online Mendelian Inheritance in Man database).



This database was initiated in the early 1960s by Dr. Victor A. McKusick as a catalog of Mendelian traits and disorders and is maintained at the Johns Hopkins University. Dr. McKusiek is widely known as the "father of medical genetics."

109150

MACHADO-JOSEPH DISEASE; MJD

Alternative titles; symbols

SPINOCEREBELLAR ATAXIA 3; SCA3
SPINOCEREBELLAR ATROPHY III
AZOREAN NEUROLOGIC DISEASE
SPINOPONTINE ATROPHY
NIGROSPINODENTATAL DEGENERATION

Phenotype-Gene Relationships

Location	Phenotype	Phenotype MIM number	Inheritance	Phenotype mapping key	Gene/Locus	Gene/Locus MIM number
14q32.12	Machado-Joseph disease	109150	AD	3	ATXN3	607047

Clinical Synopsis | Phenotypic Series

TEXT

A number sign (#) is used with this entry because Machado-Joseph disease (MJD), also known as spinocerebellar ataxia-3 (SCA3), is caused by a heterozygous (CAG)n trinucleotide repeat expansion encoding glutamine repeats in the ataxin-3 gene (ATXN3; 607047) on chromosome 14q32.

Normal individuals have up to 44 glutamine repeats, and MJD patients have between 52 and 86 glutamine repeats. Incomplete penetrance is associated with 45 to 51 repeats (Todd and Paulson, 2010).

Machado-Joseph disease (MJD), also known as spinocerebellar ataxia-3 (SCA3), is caused by a heterozygous (CAG)n trinucleotide repeat expansion encoding glutamine that repeats in the ataxin-3 gene.

There is a wealth of information on SCA3 disease in this OMIM entry. Scroll down through the Description and Clinical Features:

Machado-Joseph disease, named for affected families of Azorean extraction, is an autosomal dominant progressive neurologic disorder characterized principally by ataxia, spasticity, and ocular movement abnormalities. Although independently described as a seemingly separate disorder, **spinocerebellar ataxia-3** is now known to be the same as Machado-Joseph disease.

The **molecular genetics** Of MJD/SCA3 has shown that in normal individuals, the ATXN3 gene was found to contain between 13 and 36 CAG repeats, whereas most of the patients with clinically diagnosed MJD and all of the affected members of a family with the clinical and pathologic diagnosis of MJD showed expansion of the repeat number to the range of 68 to 79.

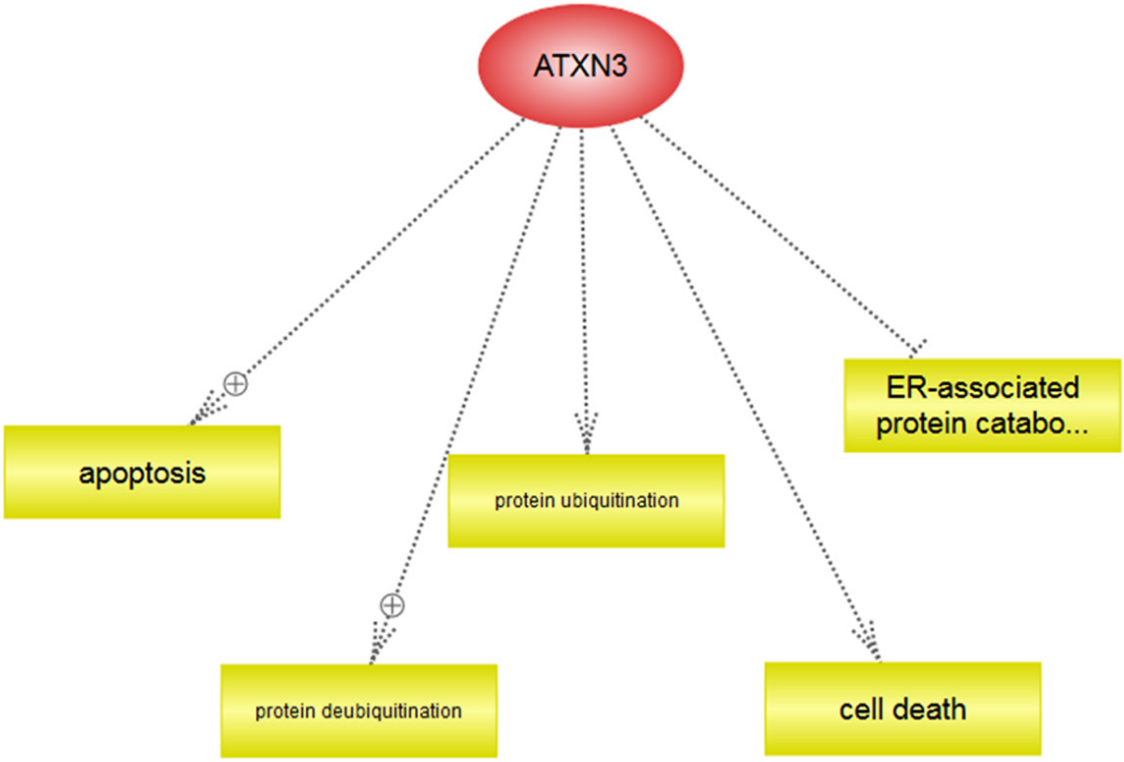
The **pathogenesis** of MJD/SCA3 is clearly related to the expanded CAG repeats since cell death in cultured cells expressing a portion of the ATXN3 gene that included the expanded CAG repeats occurred only when the CAG repeat was translated into polyglutamine residues[1].

The **neuron-specific phenotype** of Machado-Joseph disease can be explained by experiments showing that L-glutamate-induced excitation of patient-specific induced pluripotent stem cell (iPSC)-derived neurons initiates calcium-dependent proteolysis of ATXN3 followed by the formation of SDS-insoluble aggregates. Aggregate formation was further dependent on functional sodium and potassium channels as well as ionotropic and voltage-gated calcium channels, and was not observed in iPSCs, fibroblasts, or glia[2].

1: Ikeda H, Yamaguchi M, Sugai S, Aze Y, Narumiya S, Kakizuka A. Expanded polyglutamine in the Machado-Joseph disease protein induces cell death in vitro and in vivo. Nat Genet. 1996 Jun;13(2):196-202. PubMed PMID: 8640226.

2: Koch P, Breuer P, Peitz M, Jungverdorben J, Kesavan J, Poppe D, Doerr J, Ladewig J, Mertens J, Tüting T, Hoffmann P, Klockgether T, Evert BO, Wüllner U, Brüstle O. Excitation-induced ataxin-3 aggregation in neurons from patients with Machado-Joseph disease. Nature. 2011 Nov 23;480(7378):543-6.. PubMed PMID: 22113611.

10.7 ATXN3 and Cell Processes



Let's learn more about the ATXN3 gene by copying it into a new workspace and adding cell processes to it, filtering for 10 or more references.

We can see by inspecting the relation table and some of the references that normal ATN3 has positive protein deubiquitinating activity and is involved in the endoplasmic reticulum-associated degradation pathway:

"The action of ataxin-3 in this context requires its deubiquitinase activity as well as the intact proteasome function, suggesting that ataxin-3-mediated deubiquitination may promote proteasomal degradation of misfolded or dysfunctional proteins to alleviate polyQ-associated toxicity."

From Liu Y, Ye Y. Roles of p97-associated deubiquitinases in protein quality control at the endoplasmic reticulum. Curr Protein Pept Sci. 2012 Aug;13(5):436-46. Review. PubMed PMID: 22812527

We can drill down further into the article from which this knowledge was extracted and we come across a very useful summary of the biological conditions under which normal ATXN3 functions:

"Polyubiquitination of misfolded proteins… is thought to be associated with the formation of inclusion bodies… The diseases associated with protein misfolding and aggregation are recognized as "conformational diseases" … The common feature of these diseases is the tendency of misfolded protein to form aggregates. Misfolded proteins can be refolded by molecular chaperones or cleared by the ubiquitin-proteasome system (UPS)."

On the other hand, mutant ATXN3 clearly promotes the apoptosis of neurons:

"PolyQ-expanded ataxin-3 is neurotoxic and induces neuronal apoptosis through the mitochondrial pathway"

And,

"Ataxin-3 has also been proposed to regulate protein degradation via endoplasmic reticulum-associated protein degradation."

From Huang Q, Figueiredo-Pereira ME. Ubiquitin/proteasome pathway impairment in neurodegeneration: therapeutic implications. Apoptosis. 2010 Nov;15(11):1292-311. doi: 10.1007/s10495-010-0466-z. Review. PubMed PMID: 20131003

Also from the definition of the cell process: ER-associated protein catabolism,

(External Identifiers, GO ID 0030433 at AMIGO),

ER-associated protein degradation (ERAD) is the series of steps necessary to target endoplasmic reticulum (ER)-resident proteins for degradation by the cytoplasmic proteasome. Begins with recognition of the ER-resident protein, includes retrotranslocation (dislocation) of the protein from the ER to the cytosol, protein ubiquitination necessary for correct substrate transfer, transport of the protein to the proteasome, and ends with degradation of the protein by the cytoplasmic proteasome.

The picture is getting clearer!

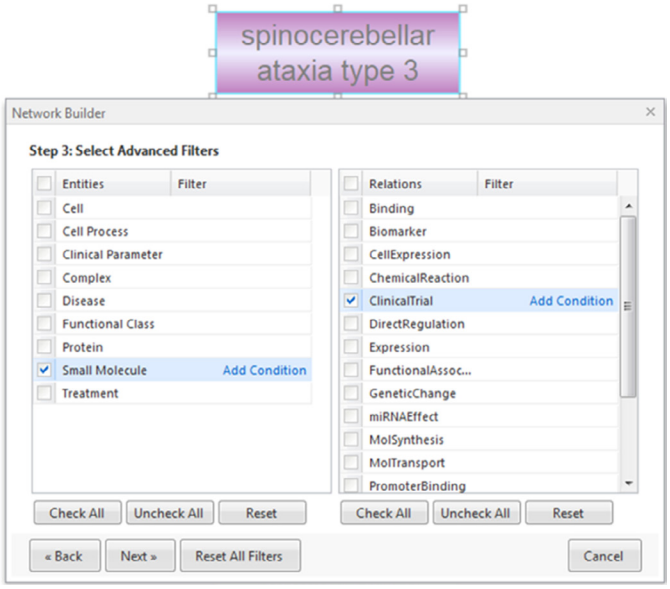
Mutant ATXN3 deregulates the normal processes of protein degradation with the resulting build-up of toxic accumulations of misfolded proteins into insoluble aggregates.

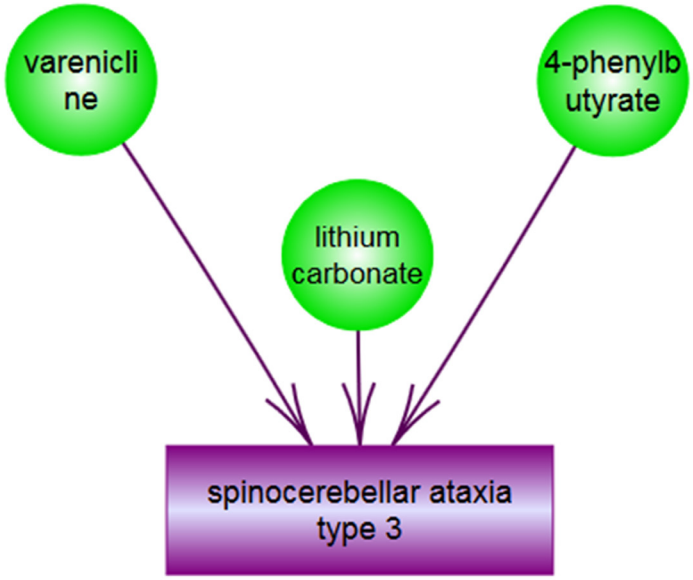
So, what about therapeutic treatments, do any exist for this unfortunate condition?

--let's go find out:

10.8 Therapeutic Treatments: Clinical Trials

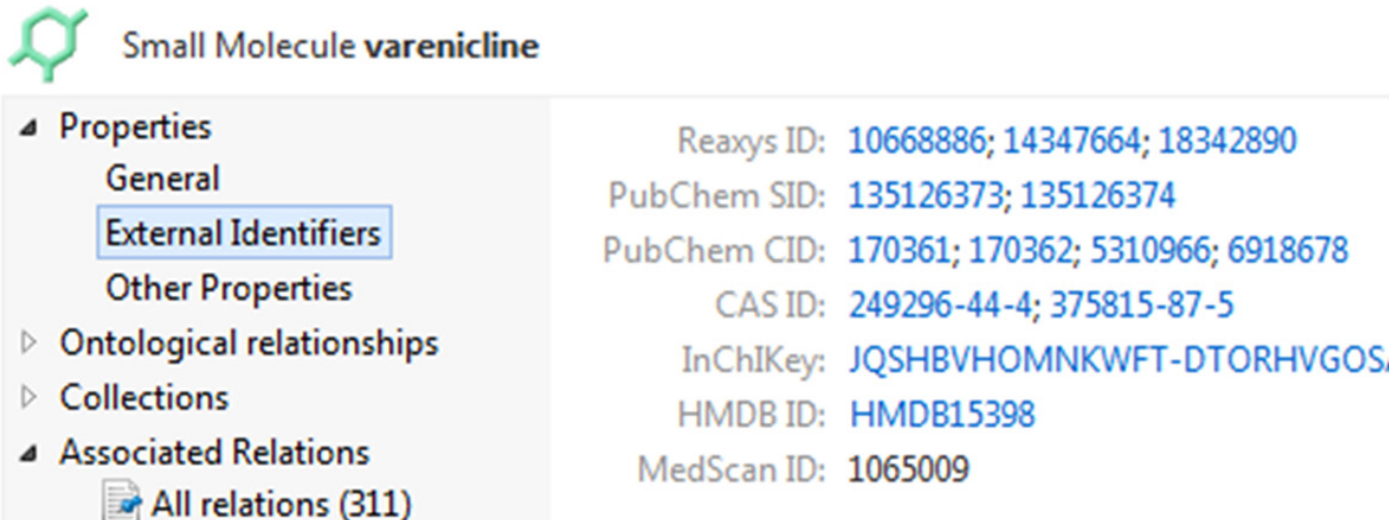
(Add, Network Builder, Advanced Expand Pathway, Small Molecule, Clinical Trial)



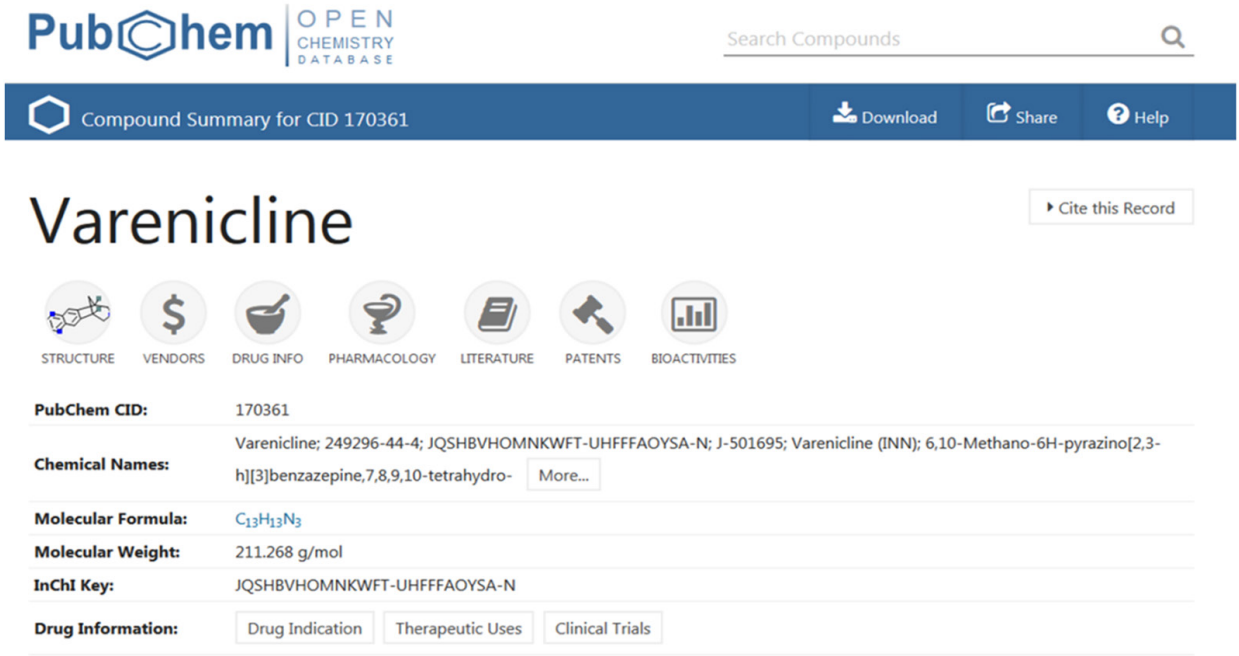


Wow, not too many drugs in clinical trials (maybe this is because SCA3 is a rare disease).

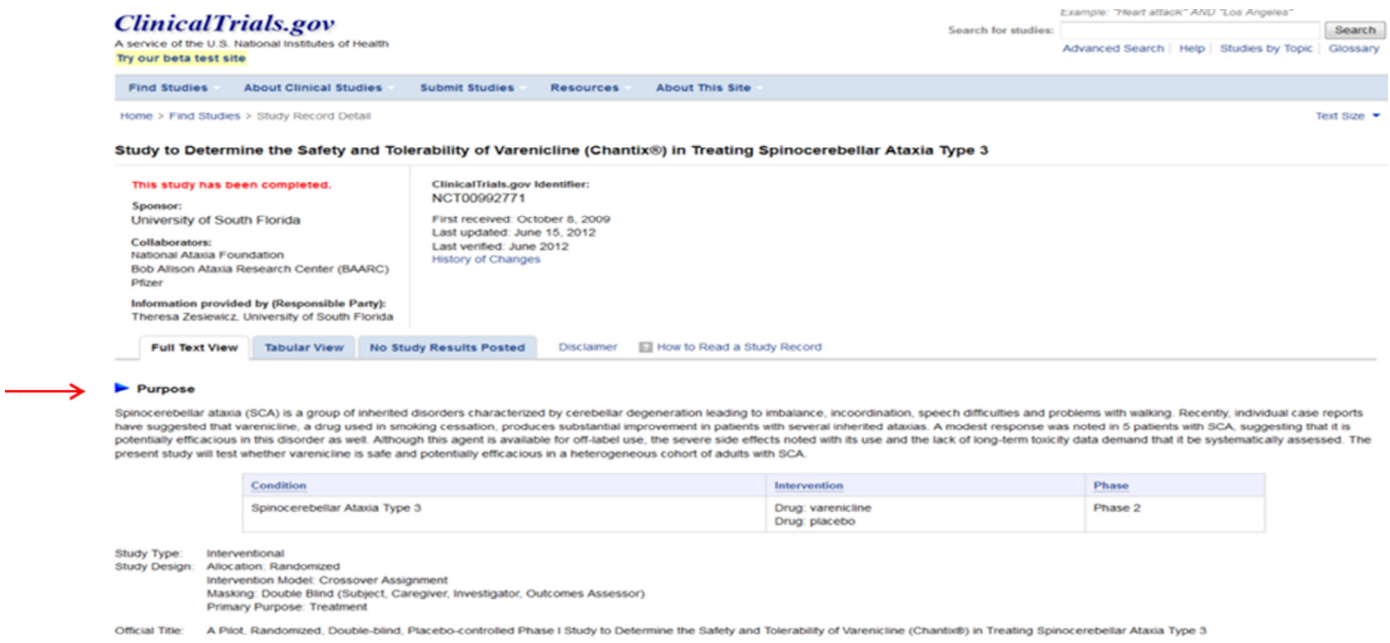
Let's quickly go check them out, right click on the entity and go to Properties, External Identifiers, PubChem CID:



This will usually give you a nice, succinct summary of a compound/drug's mechanism of action (MOA) and therapeutic indication.



In this case, varencline is partial agonist (activator) of the nicotinic receptor and is used for smoking cessation. The clinical trial record indicates that a modest response was noted in patients with SCA:



ClinicalTrials.gov
A service of the U.S. National Institutes of Health
Try our beta test site

Search for studies: Search
Advanced Search | Help | Studies by Topic | Glossary

Find Studies | About Clinical Studies | Submit Studies | Resources | About This Site

Home > Find Studies > Study Record Detail

Text Size

Study to Determine the Safety and Tolerability of Varenicline (Chantix®) in Treating Spinocerebellar Ataxia Type 3

This study has been completed.

Sponsor:
University of South Florida

Collaborators:
National Ataxia Foundation
Bob Allison Ataxia Research Center (BAARC)
Pfizer

Information provided by (Responsible Party):
Theresa Zesiewicz, University of South Florida

ClinicalTrials.gov Identifier:
NCT00992771

First received: October 8, 2009
Last updated: June 15, 2012
Last verified: June 2012
History of Changes

Full Text View | Tabular View | No Study Results Posted | Disclaimer | How to Read a Study Record

Purpose

Spinocerebellar ataxia (SCA) is a group of inherited disorders characterized by cerebellar degeneration leading to imbalance, incoordination, speech difficulties and problems with walking. Recently, individual case reports have suggested that varenicline, a drug used in smoking cessation, produces substantial improvement in patients with several inherited ataxias. A modest response was noted in 5 patients with SCA, suggesting that it is potentially efficacious in this disorder as well. Although this agent is available for off-label use, the severe side effects noted with its use and the lack of long-term toxicity data demand that it be systematically assessed. The present study will test whether varenicline is safe and potentially efficacious in a heterogeneous cohort of adults with SCA.

Condition	Intervention	Phase
Spinocerebellar Ataxia Type 3	Drug: varenicline Drug: placebo	Phase 2

Study Type: Interventional
Study Design: Allocation: Randomized
Intervention Model: Crossover Assignment
Masking: Double Blind (Subject, Caregiver, Investigator, Outcomes Assessor)
Primary Purpose: Treatment

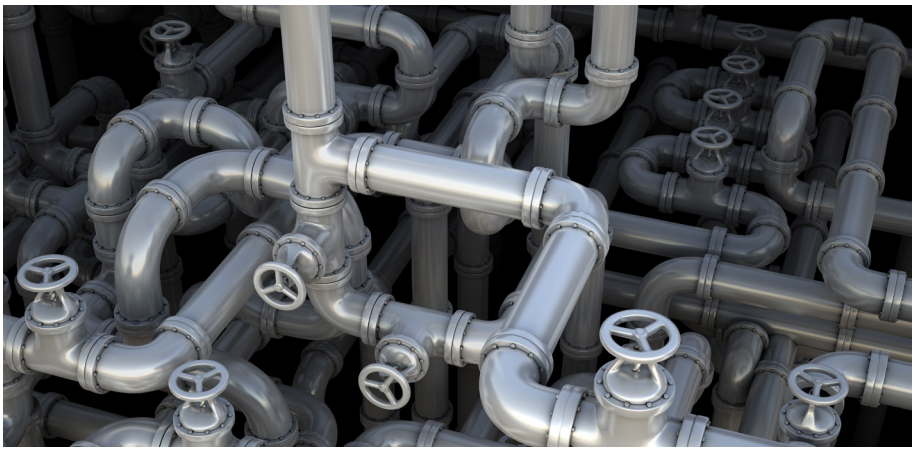
Official Title: A Pilot, Randomized, Double-blind, Placebo-controlled Phase I Study to Determine the Safety and Tolerability of Varenicline (Chantix®) in Treating Spinocerebellar Ataxia Type 3

A quick similar look at lithium carbonate and 4-phenylbutyrate shows that they are a mood stabilizing agent or have antineoplastic activity, respectively.

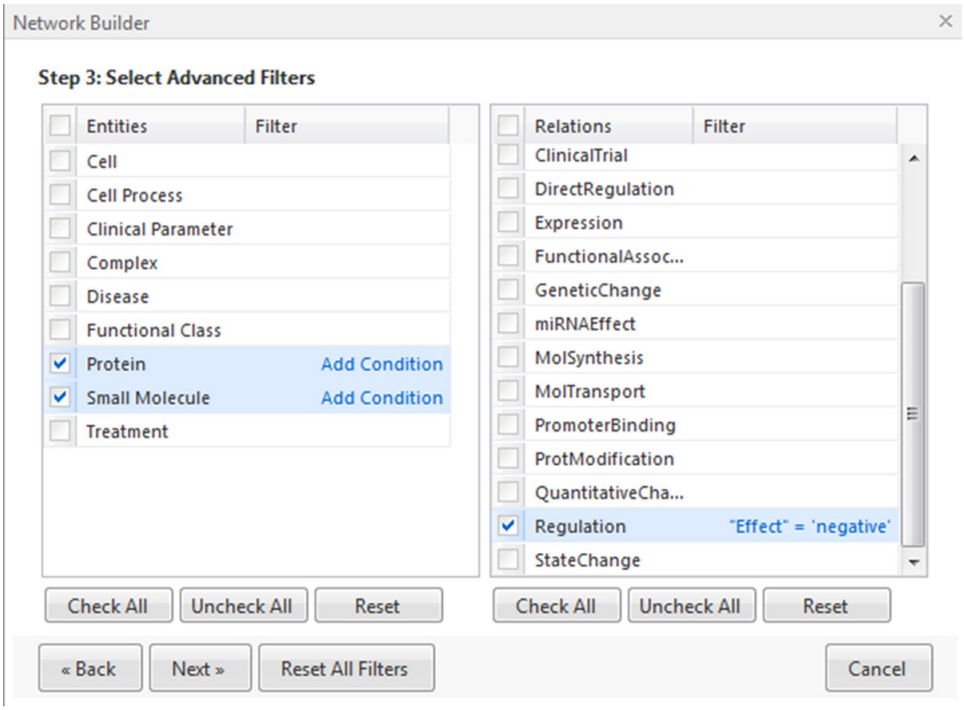
It clearly looks like these trials are about repurposing drugs developed for other indications and are also primarily directed at treating the symptoms of ataxia.

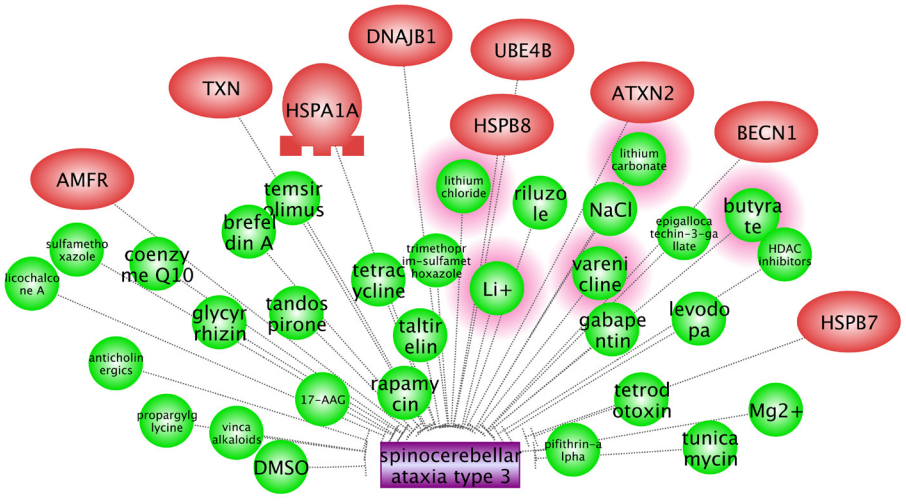
10.9 New Potential Targets

So what's in the pipeline?

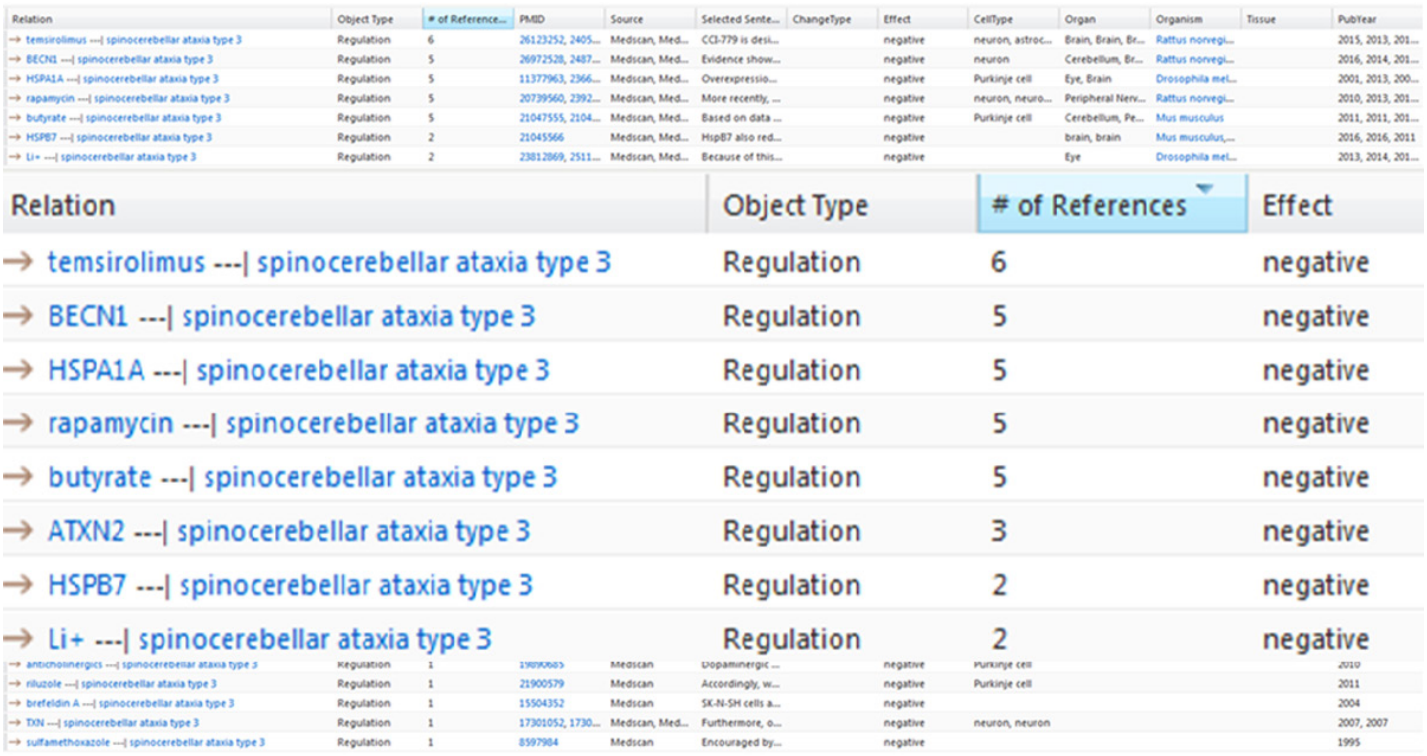


Let's find ALL protein and small molecule negative regulators discussed in connection with SCA3 in the scientific literature.





Sort results by the highest number of references (highlighted small molecules above are already in clinical trials).



Relation	Object Type	# of References	Effect
temsirolimus ---| spinocerebellar ataxia type 3	Regulation	6	negative
BECN1 ---| spinocerebellar ataxia type 3	Regulation	5	negative
HSPA1A ---| spinocerebellar ataxia type 3	Regulation	5	negative
rapamycin ---| spinocerebellar ataxia type 3	Regulation	5	negative
butyrate ---| spinocerebellar ataxia type 3	Regulation	5	negative
ATXN2 ---| spinocerebellar ataxia type 3	Regulation	3	negative
HSPB7 ---| spinocerebellar ataxia type 3	Regulation	2	negative
Li+ ---| spinocerebellar ataxia type 3	Regulation	2	negative

"Chemical activation of autophagy with **rapamycin** or its analogue CCI-779 [Temsirolimus] also reduces the levels of mutant ataxin-3, and ameliorates its toxicity in cell and mouse (expressing full-length ataxin-3-Q70) models of SCA3."[1]

"These data demonstrate that autophagy is a key degradation pathway, with beclin-1 playing a significant role in alleviating Machado-Joseph disease pathogenesis."[2]

"Over expression of **Hsp70** or specific Hsp40 chaperones suppressed neurotoxicity in the Drosophila spinocerebellar ataxia3 model, suggesting that modulation of protein folding affects the disease process."[3]

1. Sarkar S. Regulation of autophagy by mTOR-dependent and mTOR-independent pathways: autophagy dysfunction in neurodegenerative diseases and therapeutic application of autophagy enhancers. Biochem Soc Trans. 2013 Oct;41(5):1103-30. PubMed PMID: 24059496.

2. Nascimento-Ferreira I, Santos-Ferreira T, Sousa-Ferreira L, Auregan G, Onofre I, Alves S, Dufour N, Colomer Gould VF, Koeppen A, Déglon N, Pereira de Almeida L. Over expression of the autophagic beclin-1 protein clears mutant ataxin-3 and alleviates Machado-Joseph disease. Brain. 2011 May;134(Pt 5):1400-15.PubMed PMID: 21478185.

3. Bernards A, Hariharan IK. Of flies and men--studying human disease in Drosophila. Curr Opin Genet Dev. 2001 Jun;11(3):274-8. Review. PubMed PMID:11377963.

An emerging theme in potential therapeutic approaches to treatment of SCA3 disease is to go beyond just treating the symptoms of ataxia. Other treatments include proactively promoting the cellular process of autophagy (intracellular degradation system) and boosting proper protein folding.

Looks promising!

Before we go, what did we learn today?

- How to identify the most studied aspects of a rare disease.
- How to identify the gene mutation associated with SCA3 disease.
- Study the molecular genetics and pathogenesis of the SCA3 using the OMIM database link out from Pathway Studio.
- Data mining the scientific literature from Relations and References.
- Identification of small molecules involved in clinical trials for SCA3.
- Identification of potential small molecule and protein targets for therapeutic intervention for SCA3.

Study Questions 10

1. Find four proteins that are positively regulated by varenicline.

2. Find the top five diseases (by # of references) that are inhibited by varenicline.

3. Find cell processes affected by varenicline. How many of them? What are the top two cell processes (by # of references)?

4. How many total relations are in the Pathway Studio database for the rare disease "Evans Syndrome"?

5. What drug is currently in clinical trials for Evans Syndrome? What is the most studied drug (by # of references) that may be useful for the negative regulation of Evans Syndrome? At what target on what cell type is rituximab targeted?

6. What top three diseases (by # of references) are most associated with Evans Syndrome?

7. Based on information obtained in Questions 4 and 5, what is the most likely affected major tissue organ system in Evans Syndrome?

For Small Molecule Abundance, Clinical Trials, Functional Associations between Diseases, and Cell Processes, How Do I...

Exercise 10.1: What proteins regulate the synthesis or catabolism of a small molecule?

Identifies proteins involved in the translocation of a protein or small molecule target.

Step 1: Create New Pathway or within Pathway, select a protein(s)

Step 2: Select Add- > Network Builder

Step 3: Select Expand Pathway, Advanced Expand Pathway Tool

Step 4: Select Direction: Upstream

Step 5: Entities: Select “Protein and Small Molecules” Relations: Select "MolTransport"

Continue to Network Builder Step 5 and click Launch Interactive Network Builder or Finish

Exercise 10.2: What small molecules/drugs have been tested in clinical trials for a disease?

Identifies small molecules/drugs that have been involved in clinical trials. Drugs are included in ChemEffect Data. ClinicalTrial relations are included in DiseaseFx data. Monodonal antibodies are represented as small molecules in the ChemEffect database.

Step 1: Create New Pathway or within Pathway, select disease

Step 2: Select Add- > Network Builder

Step 3: Select Expand Pathway, Advanced Expand Pathway Tool

Step 4: Select Direction: Downstream

Step 5: Entities: Select “Small Molecule” Relations: Select "ClinicalTrial"

Continue to Network Builder Step 5 and click Launch Interactive Network Builder or Finish

Exercise 10.3: What cellular processes are associated with a disease?

Identifies associations between cellular processes and diseases (no Direction in the relations).

Step 1: Create New Pathway or within Pathway, select disease

Step 2: Select Add- > Network Builder

Step 3: Select Expand Pathway, Advanced Expand Pathway Tool

Step 4: Select Direction: All

Step 5: Entities: Select “Cell Process” Relations: Select "FunctionalAssociation"

Continue to Network Builder Step 5 and click Launch Interactive Network Builder or Finish

Further Study:

1. In this module, you identified three drugs already in clinical trials for SCA3. Of these, 4-phenyl butyrate was withdrawn prior to enrollment. The other two (lithium carbonate and varenicline) went to completion. Identify results in the literature for these two drugs. (Hint: Check the clinical trial record for the lithium publications and other Pathway Studio SCA3 relations for varenicline publications.)

Module 11

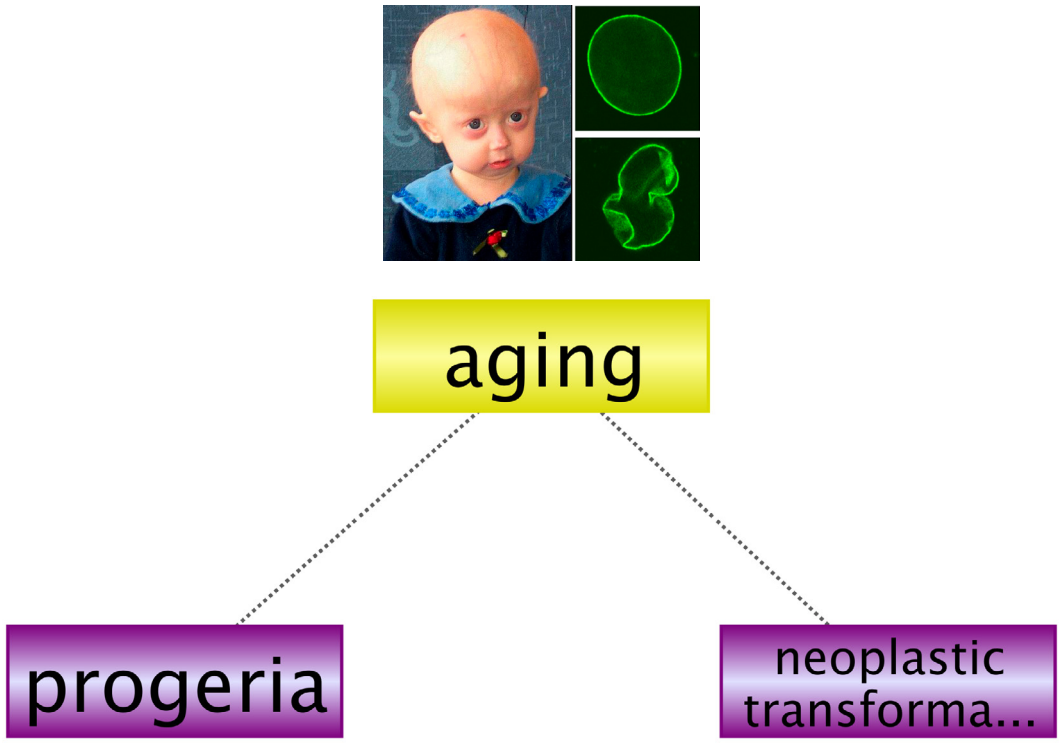
Progeria and Aging

Contents

11.1 Definition of a Rare Disease

Progeria, an abnormal congenital condition, is associated with defects in the LAMIN TYPE A gene and is characterized by premature aging in children.

What are the connections between Progeria, aging, and cancer ?



It is a sobering fact of genetic study that a significant amount of our understanding of normal physiological processes is derived from the study of diseases in which things go wrong.

Single gene (monogenic), so-called Mendelian (inherited) defects are a particularly useful source for study because the ultimate genetic causal factor is known. Examples of these types of diseases include cystic fibrosis (due to mutations in Cystic Fibrosis Transmembrane Conductance Regulator [CFTR] gene) and sickle cell anemia (mutations in hemoglobin S).

11.2 Hutchinson-Gilford Progeria Syndrome (HGPS)

Hutchinson-Gilford progeria syndrome (HGPS) is an extremely rare genetic disorder that causes premature, rapid aging (the rate of aging is accelerated up to seven times that of normal) **shortly after birth.**

Individuals with HGPS have mutations in their LMNA gene that encodes lamin A and C, the A-type lamins, which are an important structural component of the nuclear envelope.

Incomplete processing of mutant lamin A (also called progerin) results in nuclear lamina abnormalities resulting in an age-dependent, cumulative, and ultimately devastating effect on nuclear architecture and function leading to spatial disorganization and transcriptional dysregulation.

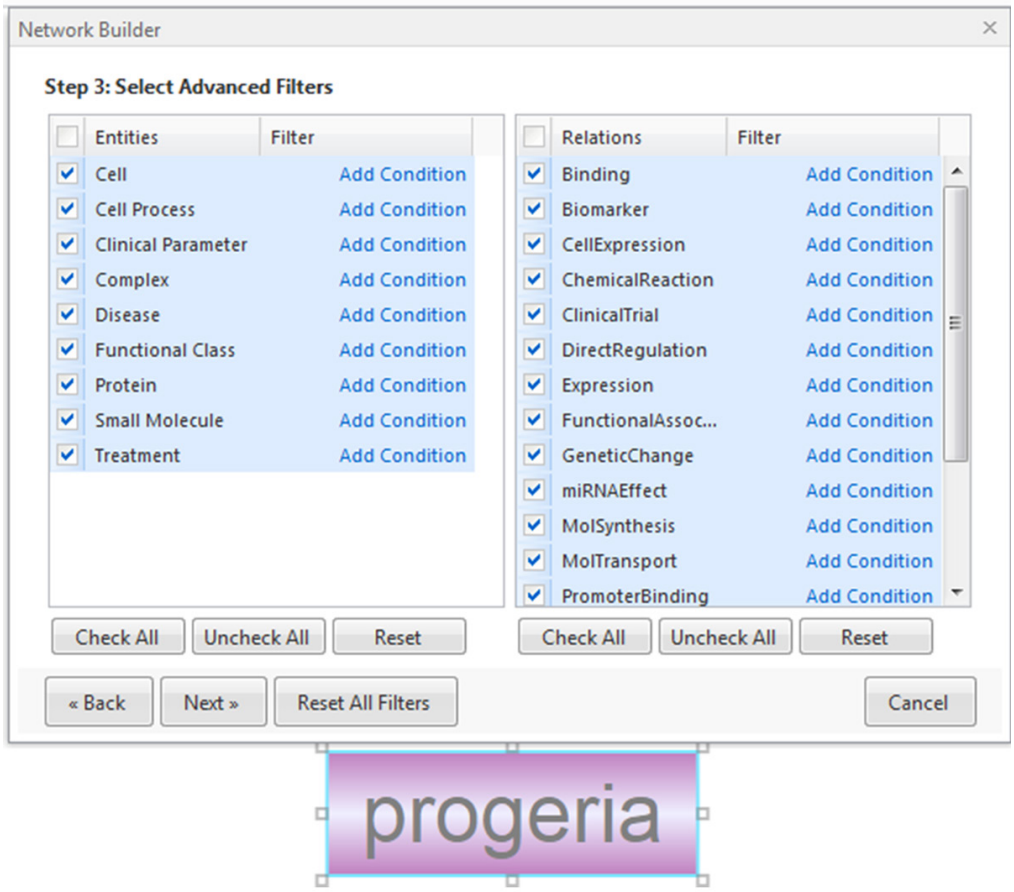
Nevertheless, the question still remains:

Why do LMNA mutant cells enter senescence earlier than normal cells?

11.3 Progeria, Cancer, and Aging

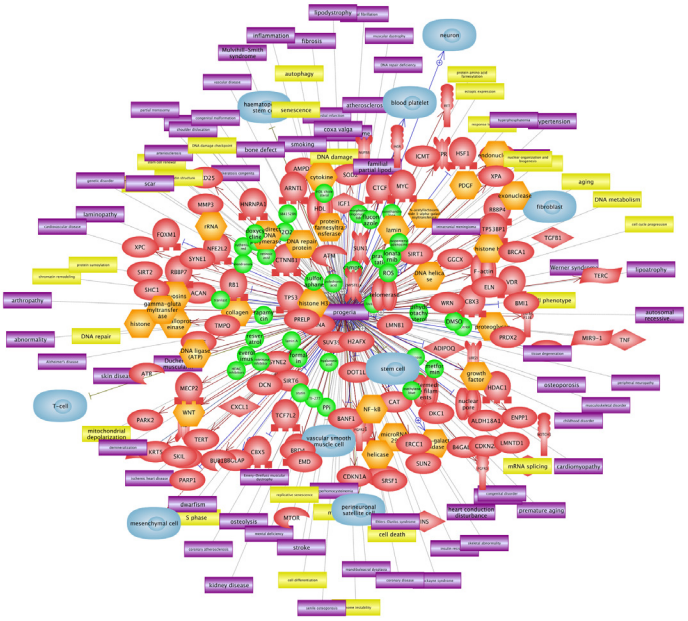
One approach to this question is to perform a comparative literature analysis between the disease of progeria, the cell process of aging, and the disease of neoplastic transformation.

First step is to add all database relations to progeria:



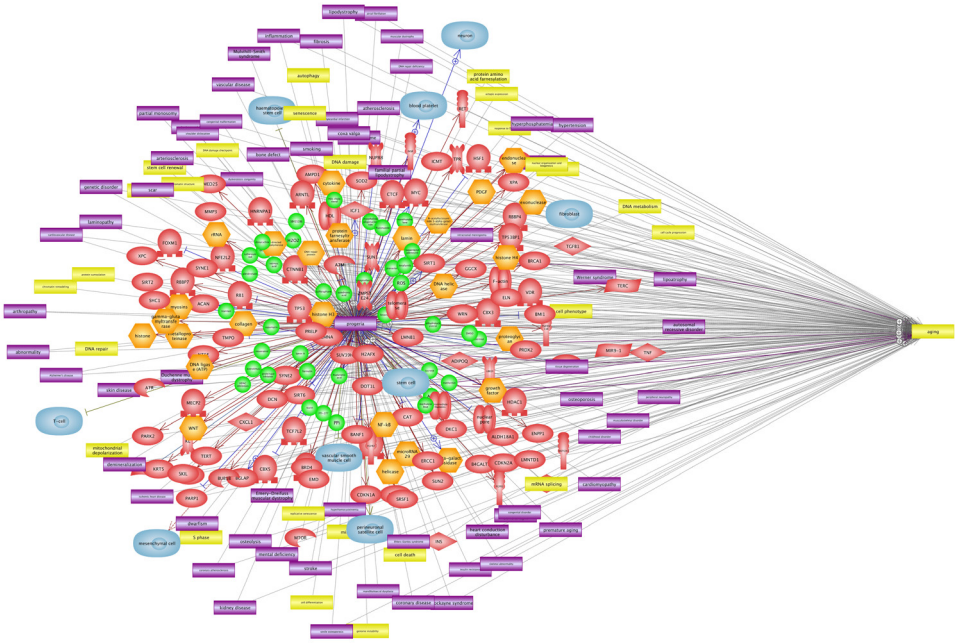
11.4 Progeria, All Relations

Progeria – All Relations

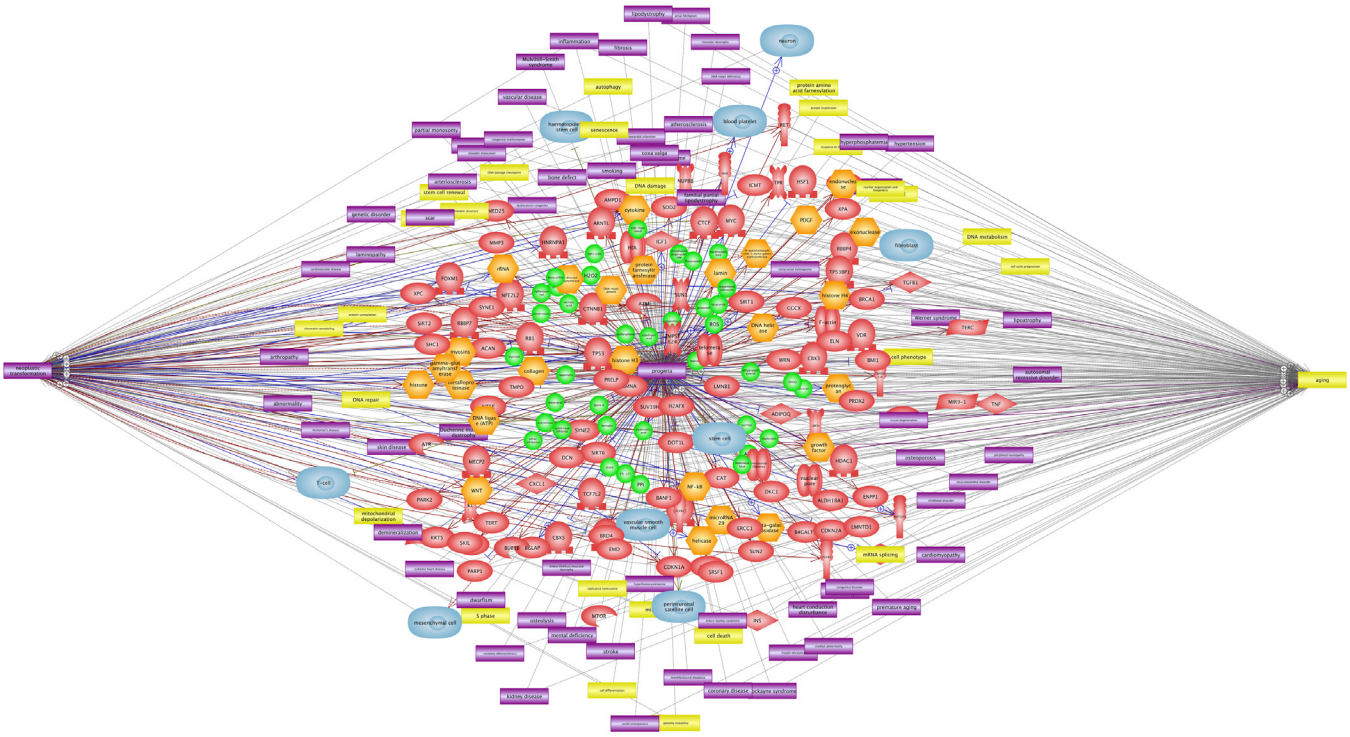


11.5 Progeria, Cancer, and Aging—Shared Relations

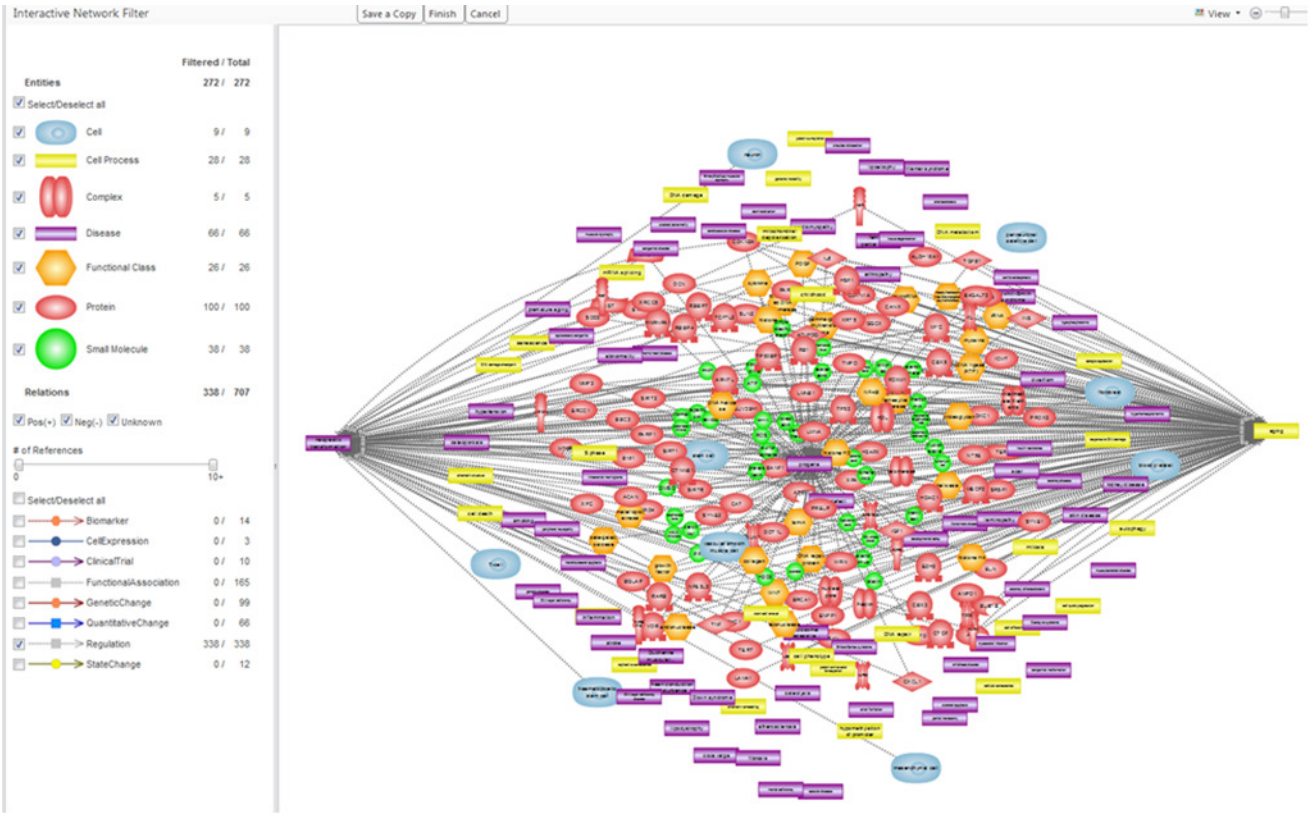
Next, in sequence, add in “aging” and “Relations between Selected and Unselected”:



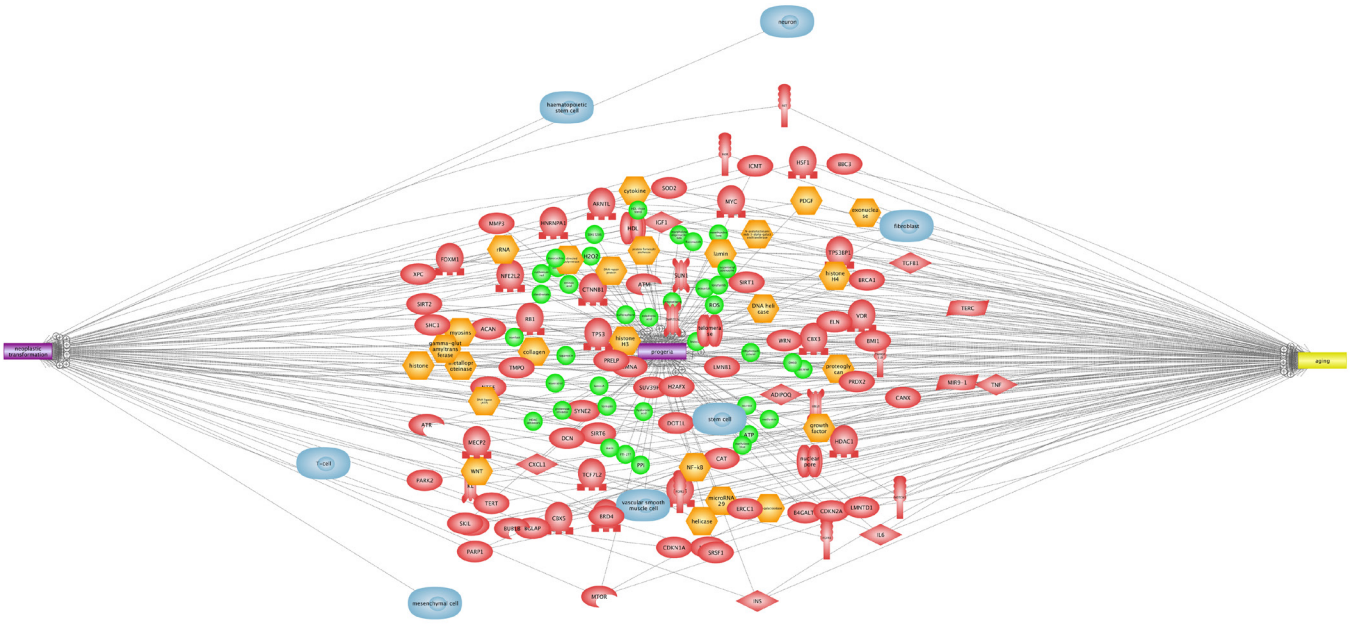
Then, add in “Neoplastic Transformation” and “Relations between Selected and Unselected”:



Filter for Regulations only:



Select “Unconnected Entities,” then edit “Remove”:

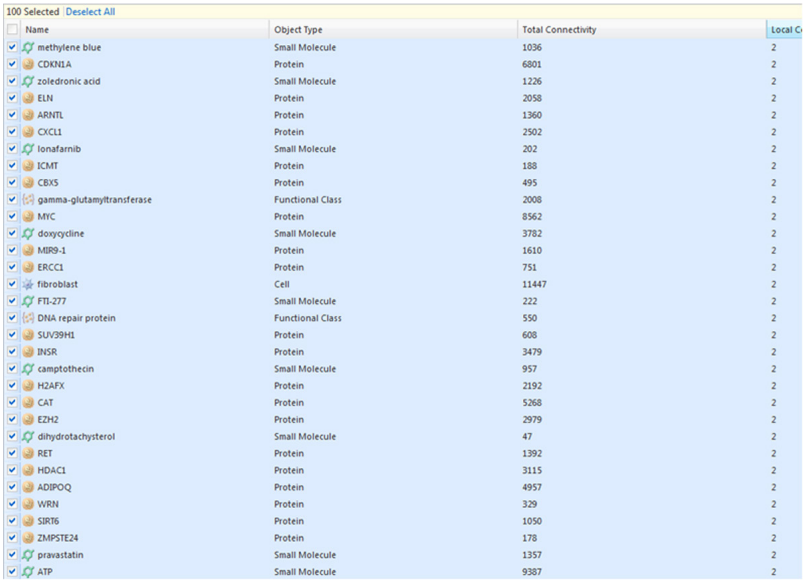


View “Entity Table View”, Sort “Local Connectivity”, Ascending

(If column is not in Entity Table View, click on tab on any column header, go to Customize Table and Add Column to display).

Select and remove all entities with less than 3 connections.

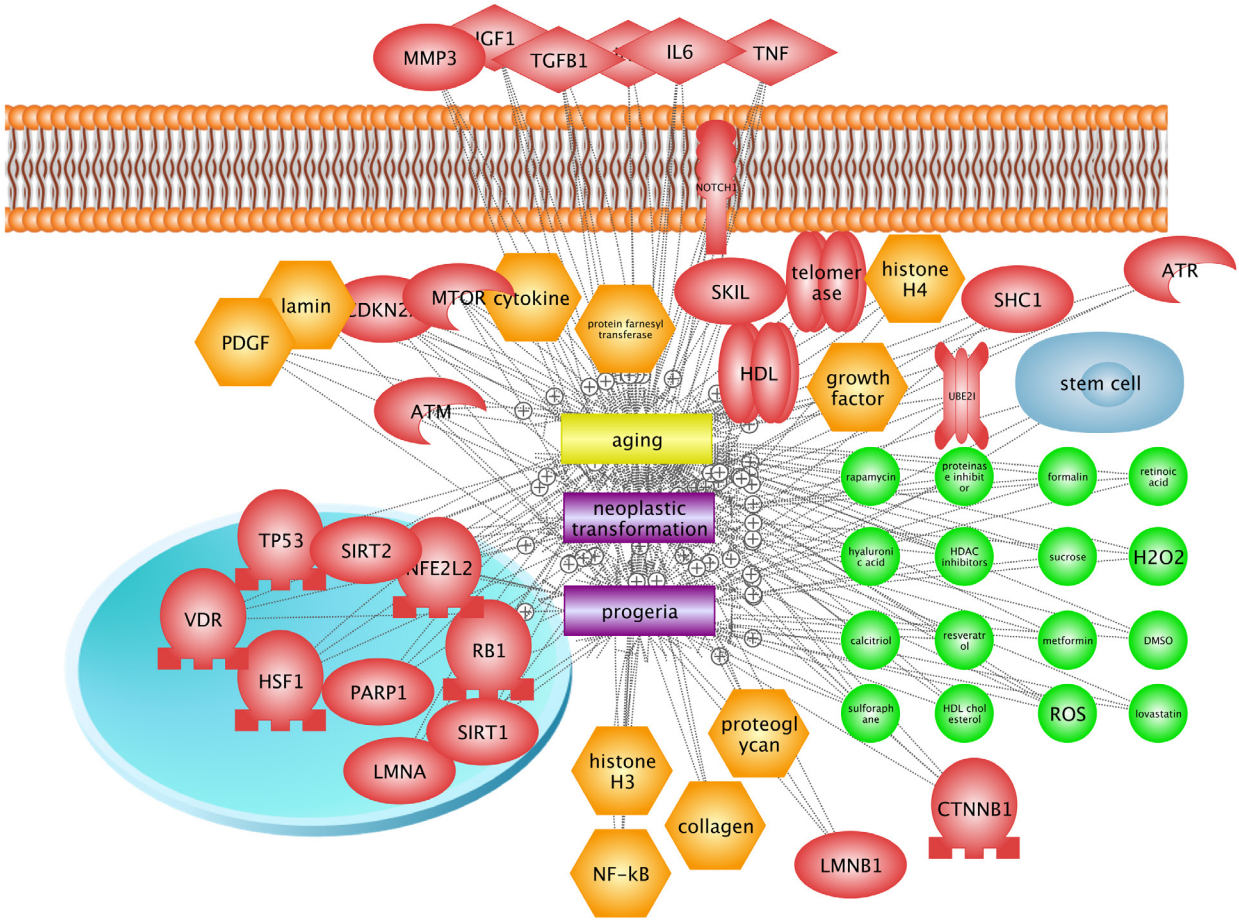
(The rationale here is that we want only entities that are connected to all three targets; progeria, aging, and neoplastic transformation. This is a pretty good approximate method; some entities will sneak in on the basis of multiple connections to a single entity. Check stem cell for an example of this - then remove!)



100 Selected | Deselect All

Name	Object Type	Total Connectivity	Local C
methylene blue	Small Molecule	1036	2
CDKN1A	Protein	6801	2
zoledronic acid	Small Molecule	1226	2
ELN	Protein	2058	2
ARNTL	Protein	1360	2
CXCL1	Protein	2502	2
lonafarnib	Small Molecule	202	2
ICMT	Protein	188	2
CBX5	Protein	495	2
gamma-glutamyltransferase	Functional Class	2008	2
MYC	Protein	8562	2
doxycycline	Small Molecule	3782	2
MIR9-1	Protein	1610	2
ERCC1	Protein	751	2
fibroblast	Cell	11447	2
FTI-277	Small Molecule	222	2
DNA repair protein	Functional Class	550	2
SUV39H1	Protein	608	2
INSR	Protein	3479	2
camptothecin	Small Molecule	957	2
H2AFX	Protein	2192	2
CAT	Protein	5268	2
EZH2	Protein	2979	2
dihydrotachysterol	Small Molecule	47	2
RET	Protein	1392	2
HDAC1	Protein	3115	2
ADIPOQ	Protein	4957	2
WRN	Protein	329	2
SIRT6	Protein	1050	2
ZMPSTE24	Protein	178	2
pravastatin	Small Molecule	1357	2
ATP	Small Molecule	9387	2

Return to Graph View and select "Layout by Localization, Plain Membrane."



Select All, Tools, Enrichment Analysis, SNEA, Cell Processes.

	Gene Set Seed	Overlap	Percent Overlap	Overlapping Entities	p-value
*	senescence	40	1	metformin, proteinase inhibit...	9.20903E-51
*	cellular senescence	37	2	metformin, CTNNB1, CDKN2A, ...	2.13369E-50
*	aging	40	1	metformin, proteinase inhibit...	5.49471E-49
	cell regeneration	33	2	metformin, CTNNB1, CDKN2A, ...	4.29845E-46
	mitotic entry	36	1	proteinase inhibitor, CTNNB1,...	1.58458E-44
	cell renewal	29	3	metformin, proteinase inhibit...	1.45328E-43
	protein processing	34	2	metformin, proteinase inhibit...	1.66965E-43
	cellular stress response	27	4	proteinase inhibitor, CTNNB1,...	6.27053E-43
	dedifferentiation	29	3	metformin, proteinase inhibit...	6.92878E-43
	adipocyte differentiation	33	2	metformin, proteinase inhibit...	2.26588E-42
	quiescence	29	3	metformin, proteinase inhibit...	4.17508E-42

Select All, Tools, Enrichment Analysis, SNEA, Diseases.

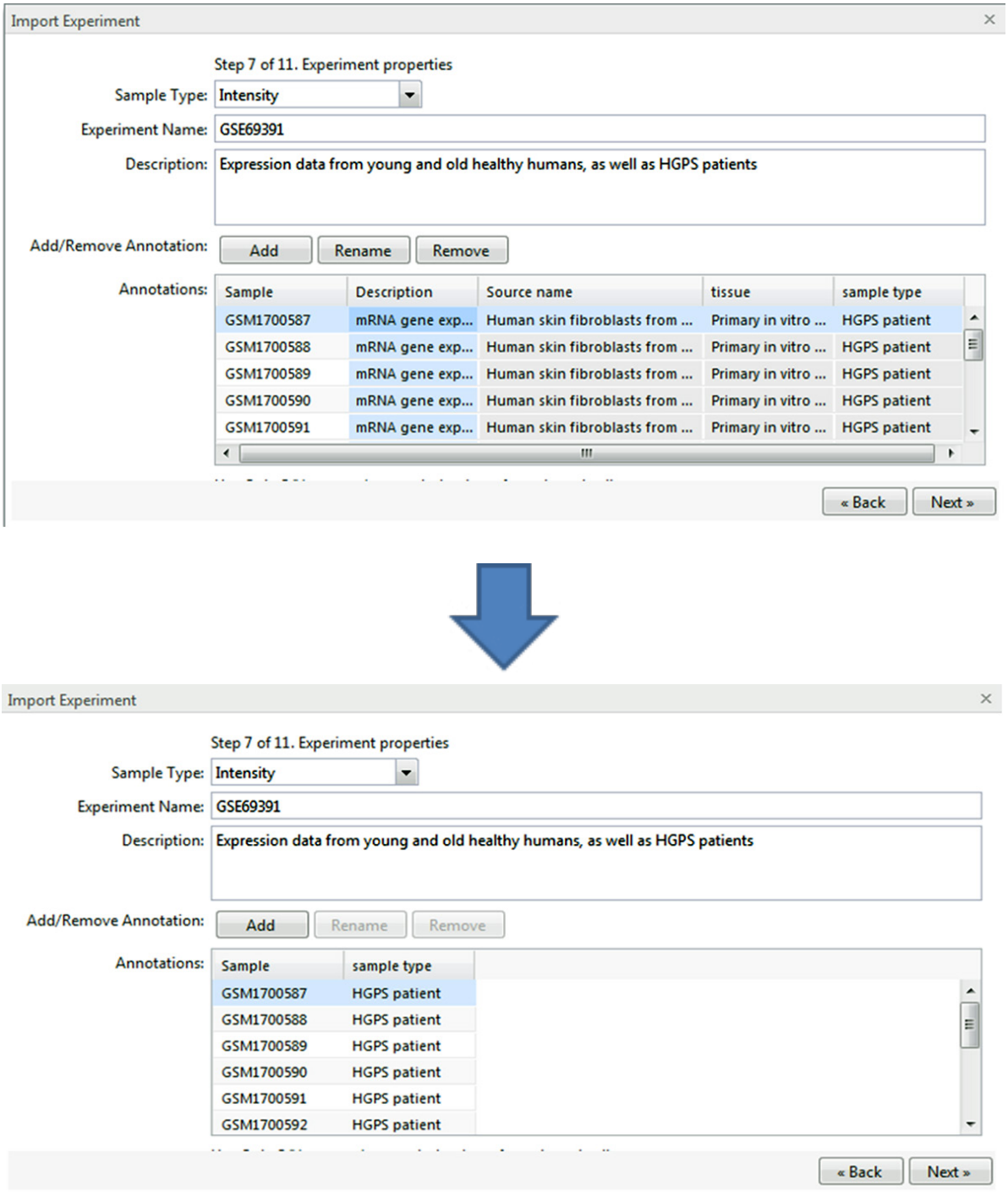
Gene Set Seed	Overlap	Percent Overl...	Overlapping Entities	p-value
progeria	37	50	metformin, proteinase inhibit...	1.51866E-103
neoplastic transformation	37	2	metformin, proteinase inhibit...	6.77341E-48
breast neoplasm	35	2	metformin, proteinase inhibit...	1.25166E-43
skin cancer	29	3	metformin, proteinase inhibit...	2.73584E-43
virus infection	35	1	metformin, proteinase inhibit...	5.55773E-43
diet-induced obesity	31	3	metformin, CTNNB1, INS, RB1, ...	5.64549E-43
malignant transformation	33	2	metformin, proteinase inhibit...	3.61593E-42
pancreatic neoplasm	25	4	metformin, proteinase inhibit...	1.00071E-38
diabetic nephropathy	29	2	metformin, CTNNB1, INS, TGFB...	1.94865E-38
hypertrophy	35	1	metformin, proteinase inhibit...	6.67571E-38
atherosclerosis	36	1	metformin, proteinase inhibit...	8.36620E-38
hyperplasia	33	1	metformin, proteinase inhibit...	1.59165E-37

Clearly this gene set is critical for progeria, aging, and neoplastic transformation as we would expect from its construction. Now let's go get some gene expression data in order to explore this question further.

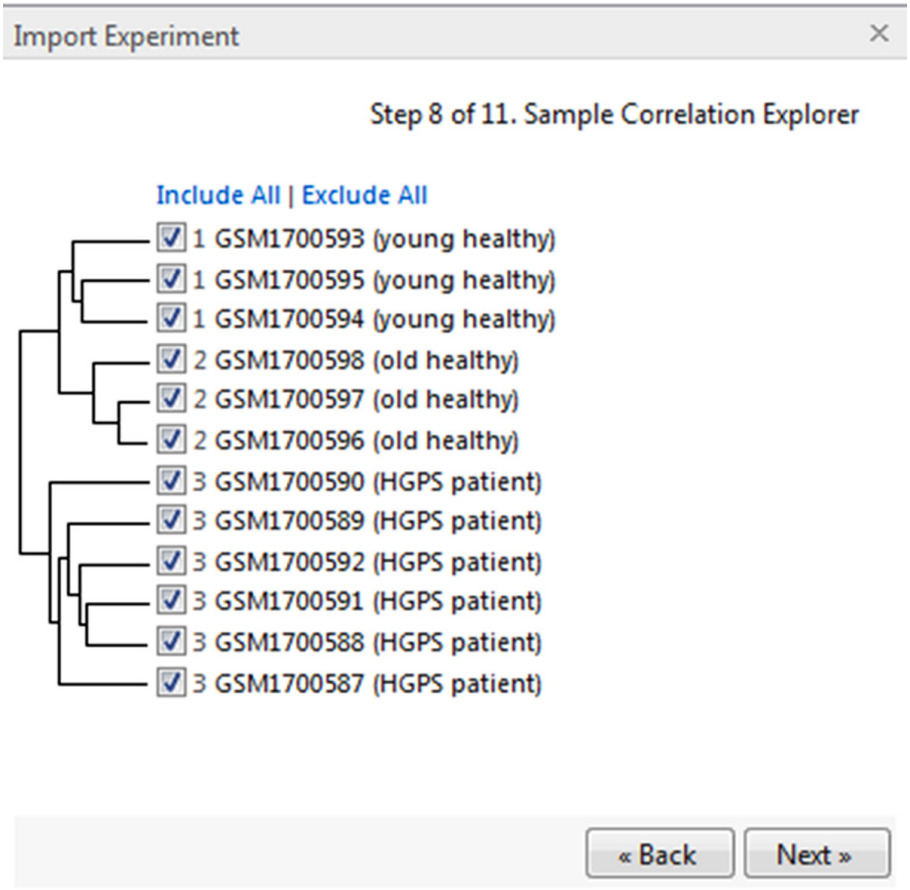
11.6 Progeria Gene Expression Data

Download the expression data set GSE69391, either directly from the GEO website or from Dropbox (data for Module 11). See Appendix 1.A for full import instructions.

“Remove” all columns except “Sample” and “Sample Type” at step 7 of the import process:



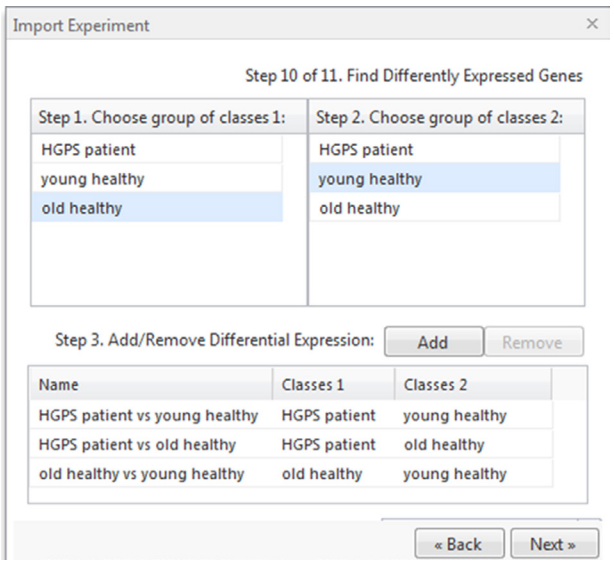
The next step indicates that the samples are nicely clustered by phenotype:

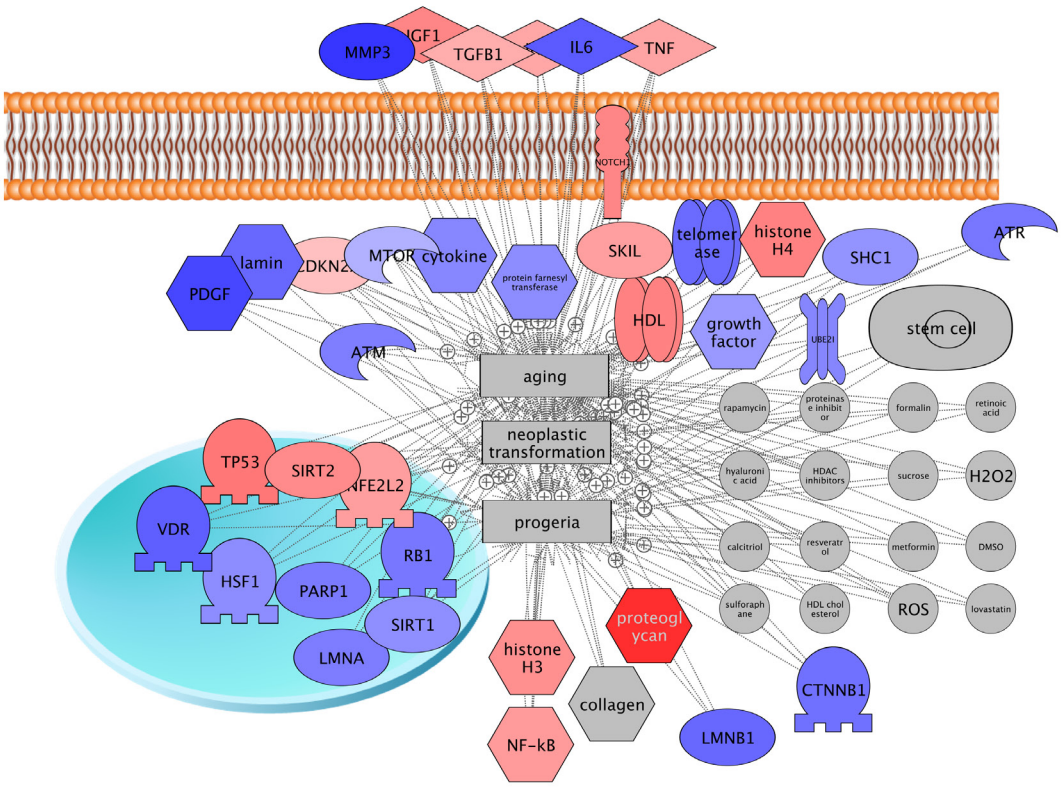


Primary skin fibroblasts were harvested from young and old healthy humans, as well as HGPS patients.

We will be interested in the contrasts between HGPS and healthy cells (both young and old) as well as between old and young healthy cells.

Now when you open up the gene expression data set alongside our last saved pathway you will see this!





You'll notice right away that the network for progeria, aging, and neoplastic transformation is now color-coded to the experimental data. It's using the data as selected in the data view column (in this case, the progeria versus young healthy fold changes). This feature can be toggled on/off using the "Link" button on the left-hand side of the data view ruler.

Now, go analyze your data!

Follow these steps!

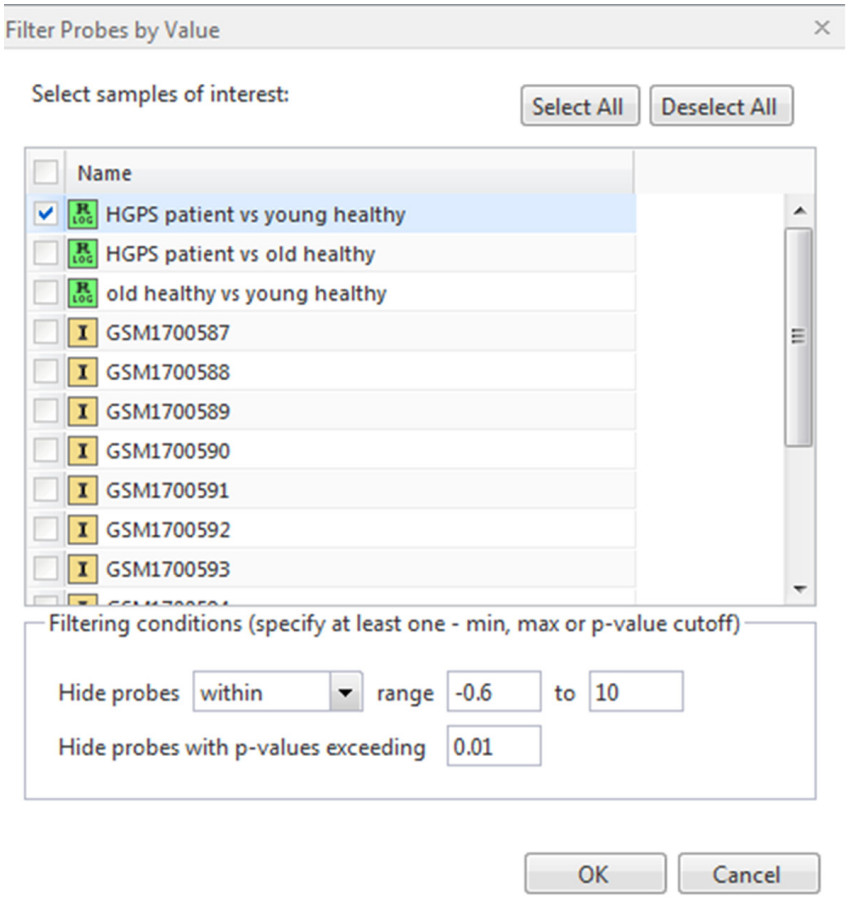
1. Sort data in column 1, ascending, this will put downregulated genes at the top.

2. Under the View tab, "Hide Unmapped Probes".

3. Filter, "Probes by Value", select first column, hide probes within -0.6 to 10 (this will, in effect, select probes with a logFC <

-0.6 or -1.5 fold or less),

hide probes with p-values exceeding 0.01 (this will in effect select probes with p-values < 0.01 or an FDR of 10%).



4. Select "Probes Remaining After Filtering", indicate in the pop-up box that the selection should apply to the "HGPS patient vs young healthy" column.

5. All this work should result in the selection of approximately 1400 progeria downregulated genes (1408 to be exact).

Now we'll do some quick **SNEA** for diseases and cell processes to characterize the downregulated genes (access SNEA via the Tools dropdown box on the right side of the Experimental Data View ruler).

And the Answer is...!

For diseases:

Gene Set Seed	Overl...	Percent Over...	Overlapping Entities	p-value
carcinogenesis	272	4	SIRPA, TRIM59, AGFG1, BCCI...	1.13352E-71
malignant transformation	95	6	USP24, MLLT10, MGMT, RAP...	1.24809E-31
metastasis	189	3	SIRPA, MGMT, JMJD6, DDX4...	6.77140E-31
neoplastic transformation	102	5	RBM39, MGMT, TRRAP, PRK...	3.68008E-30
neoplasm	325	2	SIRPA, LAMA4, TRIM59, BCCI...	5.24672E-29

For cell processes:

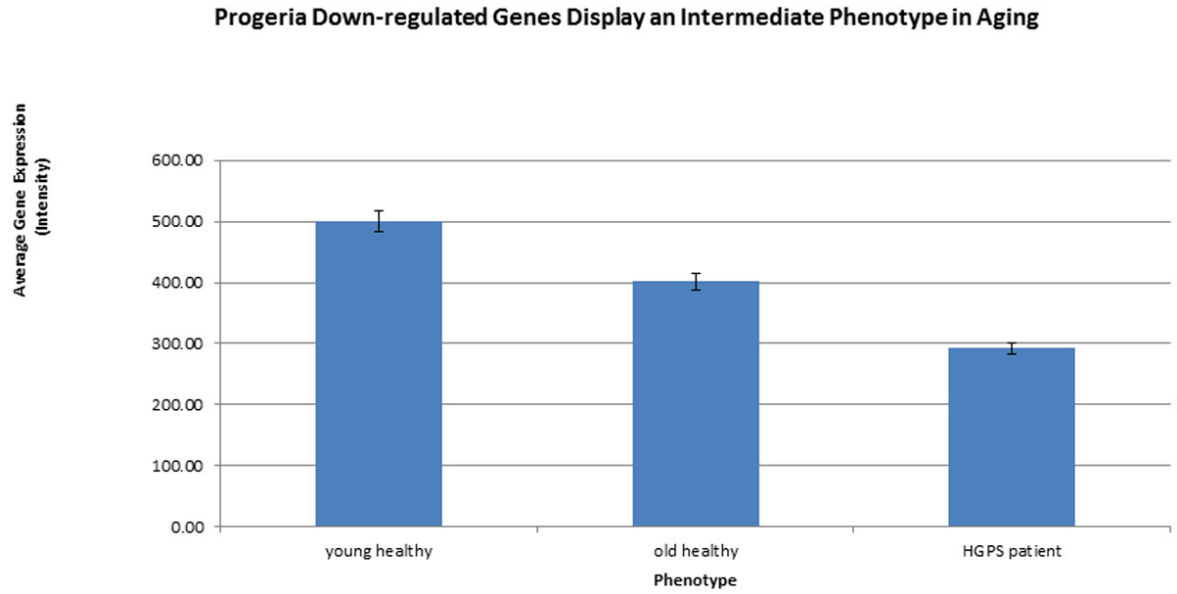
Gene Set Seed	Overlap	Percent Overl...	Overlapping Entities	p-value
chromatin remodeling	134	9	CLOCK, LBR, NIPBL, CHD2, T...	1.41367E-52
cell cycle regulation	134	8	TRIM59, CLOCK, MGMT, BCC...	8.85798E-47

So, the downregulated genes from this progeria data set are characterized by cancer genes, as well as genes involved in chromatin remodeling and cell cycle regulation.

As you inspect this data set, you will notice a curious thing. It appears that many of the Progeria downregulated genes are also downregulated in aging but just not as much.

In order to test this hypothesis computationally, we must first export the data for these genes for analysis using Excel. You don't have to do this yourself (although you are welcome to try!), you can simply download the results from Dropbox (GSE69391 allSIG DR progression genes 04-20-17.xlsx).

First, you'll average the gene expression for each phenotype group:



It looks like you were right! It appears there is a trend of decreasing expression from young healthy -> old healthy -> Progeria for those genes (1,408), which are statistically significantly downregulated between Progeria and young healthy.

This is an exciting finding as it confirms a direct transcriptional link between Progeria and aging.

11.7 Progeria Progression Genes

It would really be interesting to examine these genes directly, but first, you need to separate out only those genes (out of the total 1,408 downregulated genes) that contribute to this pattern.

You can call them the "progression" genes.

You can identify them using a very simple computational algorithm that requires that the average expression for each gene be greater from young healthy to old healthy to Progeria (see formula in Excel spreadsheet).

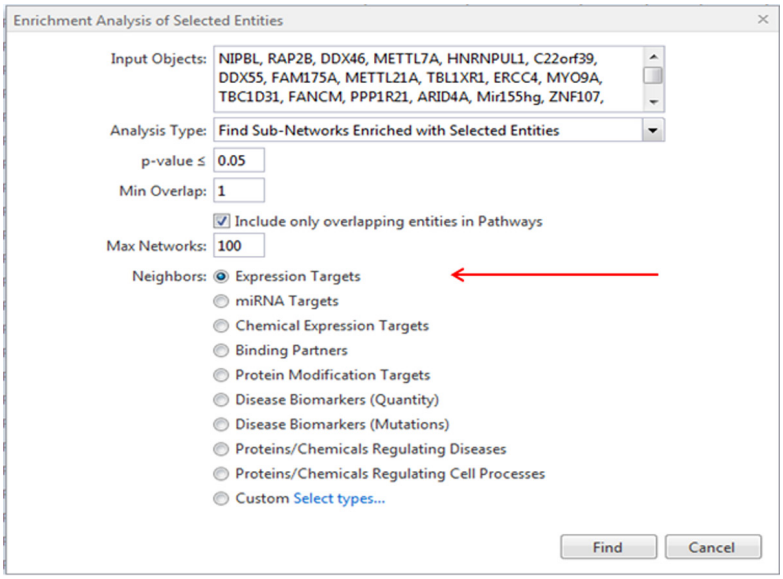
Filter for just those genes and transfer them to the second sheet. You can see that fully 1,088/1,408 (or 77%) of the original downregulated genes conform to this pattern. Clearly, this represents an important trend in the data!

Now all we have to do is Copy/Paste (or ctrl C/ctrl V) these genes back into Pathway Studio (Import Entity List) for further analysis.

Now, let's do something a little bit different!

11.8 Progression Genes, SNEA, Expression Regulators

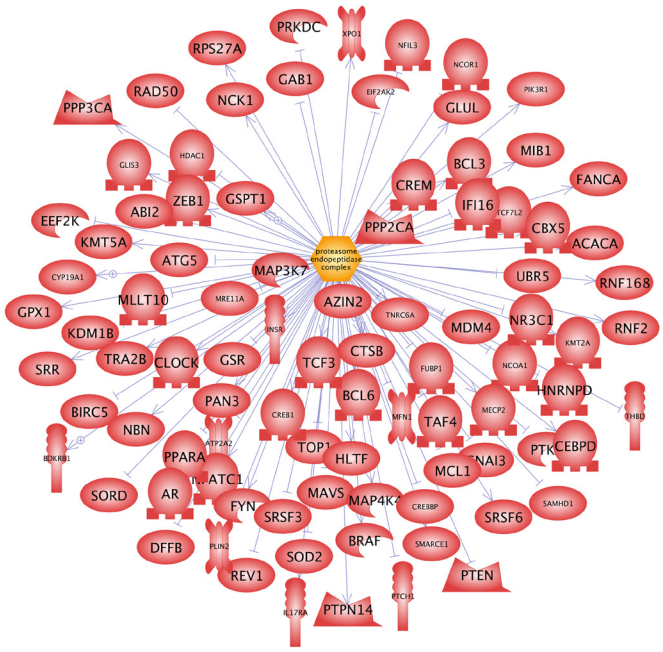
We'll perform a SNEA which identifies upstream expression regulators for these progression genes:



The expression of a large number (84) of the progression genes appears to be influenced by the proteasome endopeptidase complex (p-value < E-7):

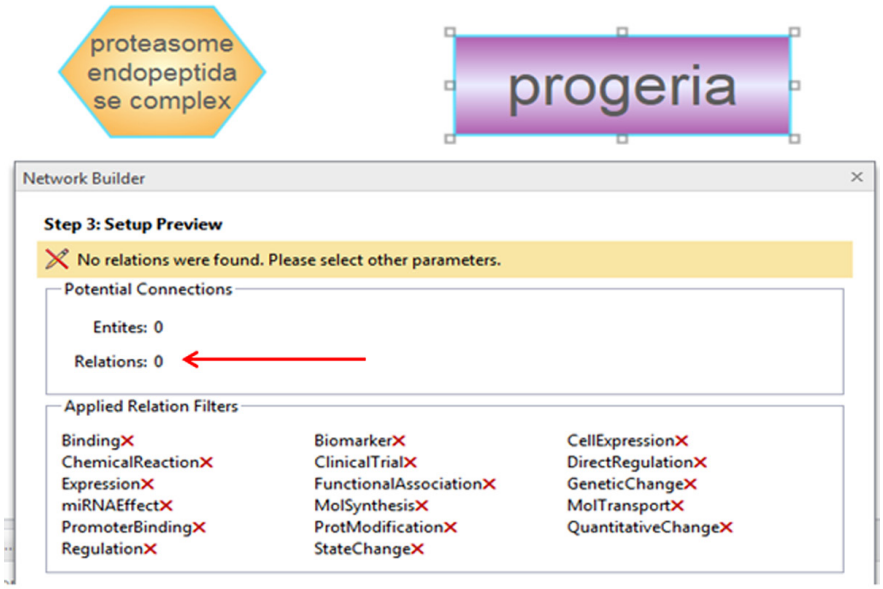
Gene Set Seed	Overlap	Percent Over...	Overlapping Entities	p-value
proteasome endopeptidase...	84	6	MLLT10, PAN3, CLOCK, SOR...	6.96636E-6
KCNN2	3	75	NRF1, TFAM, AXL	2.40281E-4
ubiquitin	52	6	CLOCK, MGMT, BIRC5, BMP...	2.92039E-4
TP53	71	5	MGMT, RAP2B, HNRNPUL1, ...	3.89700E-4
ubiquitin-protein ligase	31	7	PPP2CA, SAV1, NR3C1, HDA...	4.71147E-4

Progression genes whose expression is influenced by the proteasome endopeptidase complex:



11.9 The Proteasome Endopeptidase Complex and Progeria

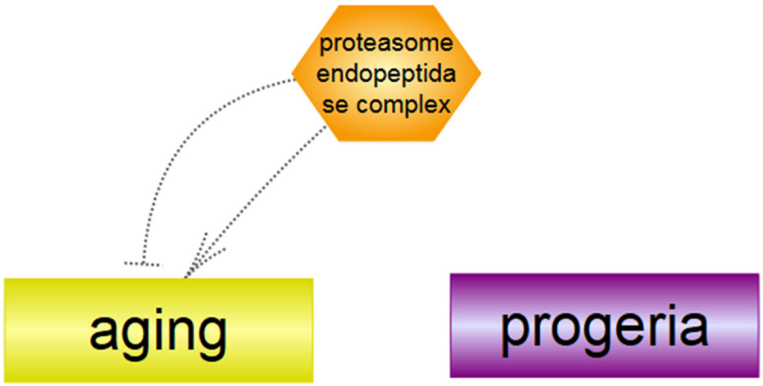
Is there a direct connection between this complex and Progeria?



Test this by selecting both entities (in a new pathway) and then, **"Add, Direct Interactions, All"**

And the answer is that we can find none at this time in the Pathway Studio database.

What about between the proteasome endopeptidase complex and aging?



→ Regulation (negative) **proteasome endopeptidase complex ---| aging**

121 references

Because the proteasome plays an essential role in cellular processes, an age-associated decline in proteasome function is assumed to contribute to the development of age-related pathology and to the aging process itself.

An age-related decrease in proteasomal activity has been assumed to be involved in the aging process and the development of age-related pathology.

It has previously been reported that the proteasome exhibits declines in function during cardiac ischemia/reperfusion and aging.

Well, actually there is substantial evidence for a connection between a loss of proteasome function and the "development of age-related pathology".

And now we've identified a similar new connection between proteasome function and progeria directly from gene expression data.

And that's not bad for a day's work!

Before we go, what did we learn today?

- The importance of single gene (monogenic) inherited defects for studying normal physiological function.
- How to combine entities and relations between diseases (progeria and cancer) and a cell process (aging).
- How to integrate gene expression data into a disease model for progeria.
- Identification of progression genes between young healthy, old healthy, and progeria patients.
- How to use a SNEA for the identification of enriched upstream regulators of gene expression.
- Identification of a new connection between deteriorating proteasome function, aging, and progeria.

Study Questions 11

Before answering the questions:

Read in the gene expression data set provided, calculate downregulated significant genes for HGPS versus young/healthy and old/healthy versus young/healthy. Save each pathway separately. Use significance criteria as outlined in the module to answer the following questions:

1. Combine the two pathways above and find the intersection. Save the resulting overlapping genes. Run an SNEA for diseases on this group. What is the top enriched disease (by p-value)? Run an SNEA for cell processes on this group. What is the top enriched cell process (by p-value)?

2. Subtract the "old/ healthy versus young/healthy" pathway from the "HGPS versus young/healthy" pathway and save the result. Run an SNEA for diseases on this group. What is the top enriched disease (by p-value)? Run an SNEA for cell processes on this group. What is the top enriched cell process (by p-value)?

3. For the SNEA cell process "cell survival" group from Question 2, what are the top three (by p-value) expression regulators? (Hint: SNEAS expression regulators)

4. How many genes/proteins are in the top group from Question 4? How many of these genes are also connected to the cell process of aging? (Hint: Check local connectivity for genes/proteins linked to both aging and cell survival.) What are the top five genes/proteins (by # of references) in this group for aging?

5. A mutation in which the five genes from Question 4 can directly cause Progeria? (Hint: Reassociate Progeria with the five genes from Question 4.)

For Cell Interactions, How Do I...

Exercise 11.1: What is the role of a cell in disease/clinical parameter?

Finds what role cells(s) have in Clinic Parameters and Diseases.

Step 1: Create New Pathway or within Pathway, select cell(s)

Step 2: Select Add- > Network Builder

Step 3: Select Expand Pathway, Advanced Expand Pathway Tool

Step 4: Select Direction: All

Step 5: Entities: Select “Clinic Parameters and Disease ” Relations: Select "Regulation"

Continue to Network Builder Step 5 and click Launch Interactive Network Builder or Finish

Exercise 11.2: How does a disease affect a certain cell type?

Finds effects of a disease on cell(s).

Step 1: Create New Pathway or within Pathway, select a disease(s)

Step 2: Select Add- > Network Builder

Step 3: Select Expand Pathway, Advanced Expand Pathway Tool

Step 4: Select Direction: All

Step 5: Entities: Select “Cell" Relations: Select "Biomarket and QuantitativeChange and Regulation and StateChange"

Continue to Network Builder Step 5 and click Launch Interactive Network Builder or Finish

Exercise 11.3: What proteins affect cell function?

Finds how proteins regulate cell functions.

Step 1: Create New Pathway or within Pathway, select cell(s)

Step 2: Select Add- > Network Builder

Step 3: Select Expand Pathway, Advanced Expand Pathway Tool

Step 4: Select Direction: Upstream

Step 5: Entities: Select “Protein, Complex, and Functional Class ” Relations: Select "Regulation"

Continue to Network Builder Step 5 and click Launch Interactive Network Builder or Finish

Exercise 11.4: What small molecules affect cell function?

Finds how small molecules regulate cell functions.

Step 1: Create New Pathway or within Pathway, select a cell(s) and or cell process(es)

Step 2: Select Add- > Network Builder

Step 3: Select Expand Pathway, Advanced Expand Pathway Tool

Step 4: Select Direction:Upstream

Step 5: Entities: Select “Small Molecule Class " Relations: Select ""Regulation"

Continue to Network Builder Step 5 and click Launch Interactive Network Builder or Finish

Exercise 11.5: What proteins are expressed in a cell? What proteins are exposed at the surface of the cell?

Finds what proteins have cell expersions relations or are exposed at the surface of a cell.

Step 1: Create New Pathway or within Pathway, select a cell(s)

Step 2: Select Add- > Network Builder

Step 3: Select Expand Pathway, Advanced Expand Pathway Tool

Step 4: Select Direction: All

Step 5: Entities: Select “Protein” Relations: "CellExpression"

Add condition: Mechanism “is equal to” surface

Continue to Network Builder Step 5 and click Launch Interactive Network Builder or Finish

Exercise 11.6: What proteins/small molecules are secreted from the cell?

Finds what proteins secrete a specific proteins and/or small molecules

Step 1: Create New Pathway or within Pathway, select a cell(s)

Step 2: Select Add- > Network Builder

Step 3: Select Expand Pathway, Advanced Expand Pathway Tool

Step 4: Select Direction: Downstream

Step 5: Entities: Select “Protein or SmallMolecule” Relations: "MolTransport"

Continue to Network Builder Step 5 and click Launch Interactive Network Builder or Finish

Exercise 11.7: What proteins are secreted from the tumor cell?

Identifies what proteins are secreted from cancers.

Step 1: Create New Pathway or within Pathway, select Disease

Step 2: Select Add- > Network Builder

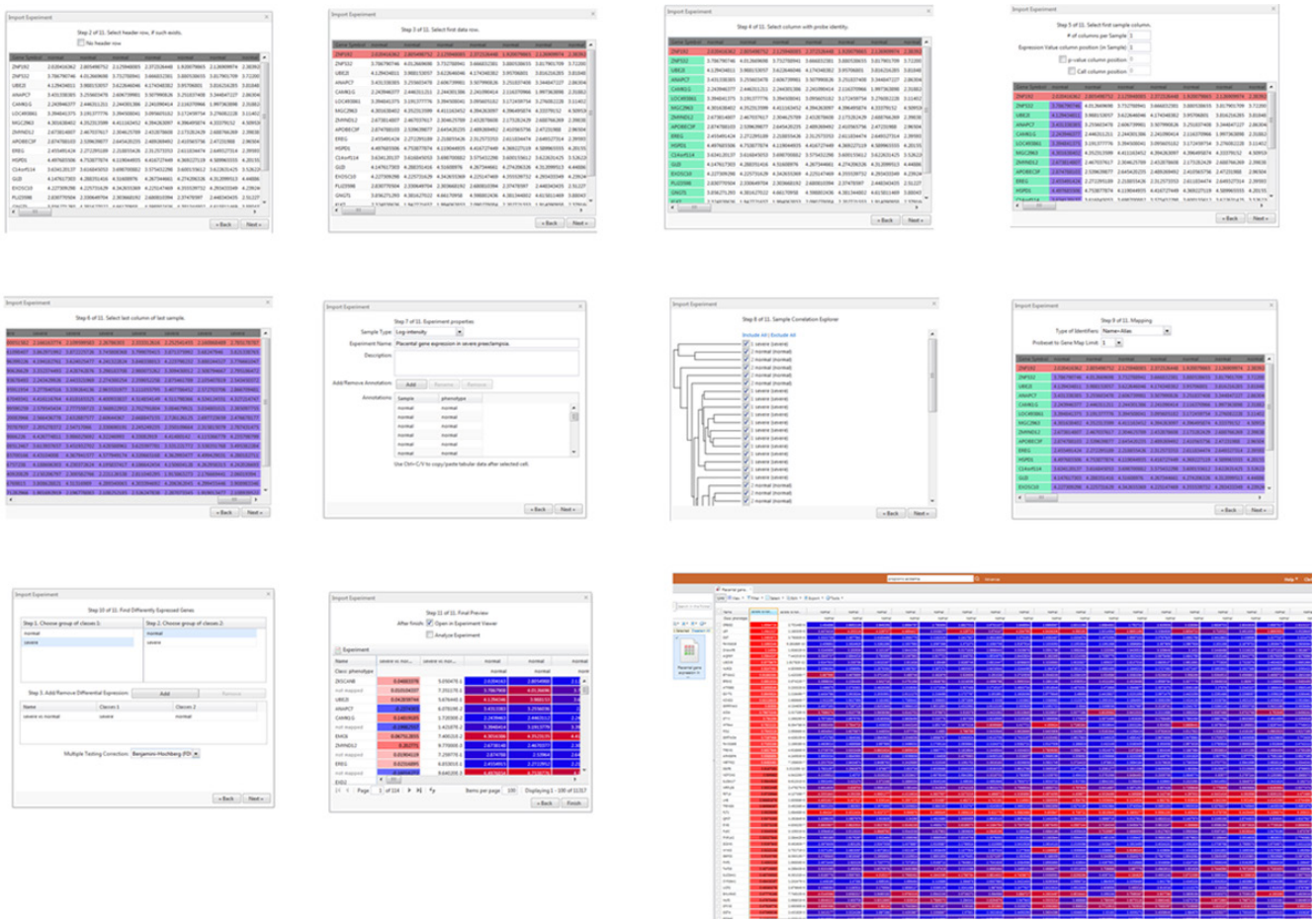
Step 3: Select Expand Pathway, Advanced Expand Pathway Tool

Step 4: Select Direction: Downstream

Step 5: Entities: Select “Protein, Complex, and Functional Class ” Relations: Select "QuantativeType"

Add condition: QuantitativeType “is equal to” secretion

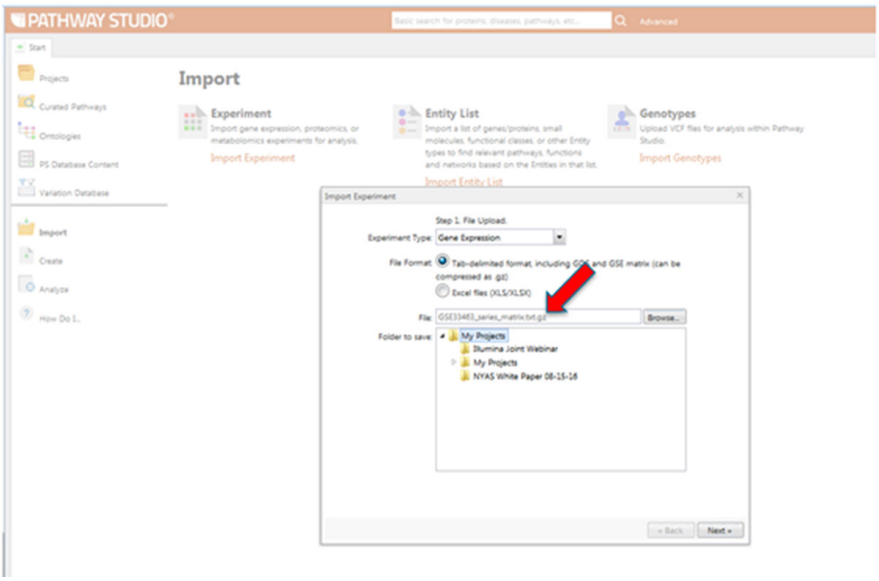
Continue to Network Builder Step 5 and click Launch Interactive Network Builder or Finish



The simplest way to get data into Pathway Studio is by direct import from GEO.

GSE33463

Download file either directly from the GEO website or from Dropbox (GSE33463_series_matrix.txt). Import to your home directory and browse to file in Pathway Studio "Import Experiment."



Appendix 1.A

Data Import from the Gene Expression Omnibus (GEO)

Getting gene expression data (microarray and RNASeq data) successfully into Pathway Studio.

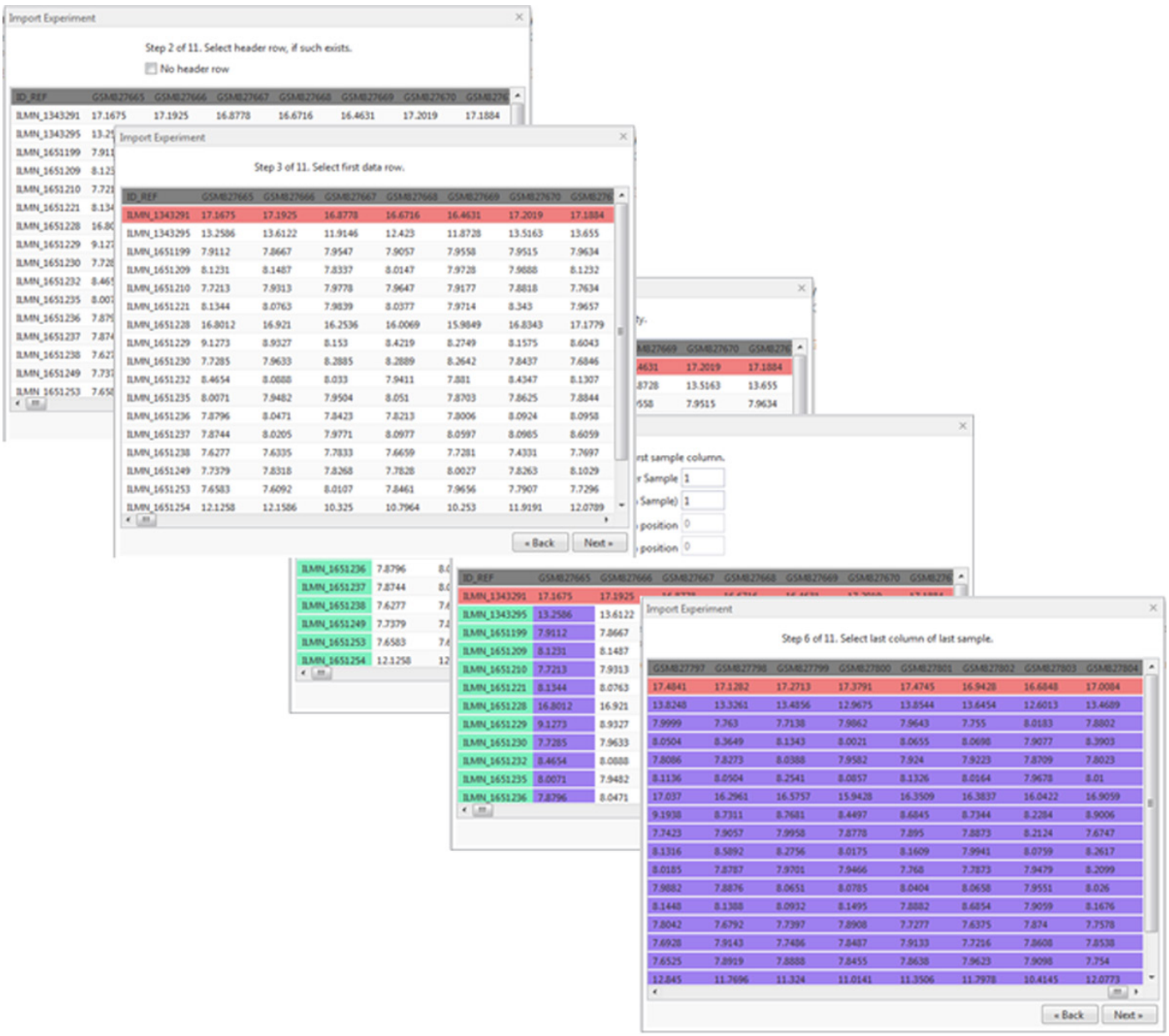
Pathway Studio can import gene expression data directly from GEO or from a tab-delimited spreadsheet file. Import is designed to be highly flexible, allowing user-defined location of key fields and column designations.

So you can.....

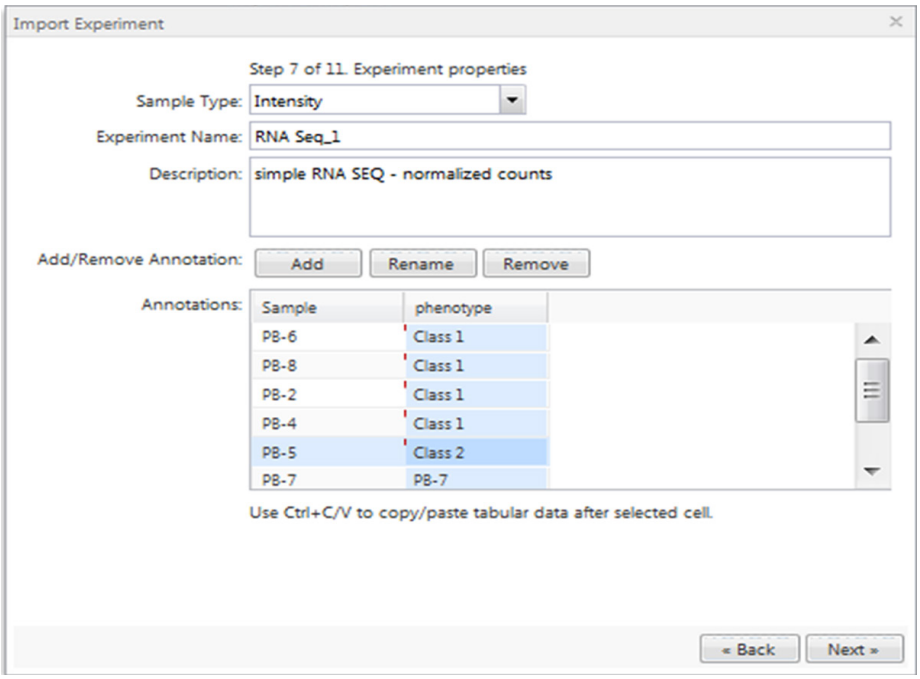
And Follow the Yellow Brick Road !

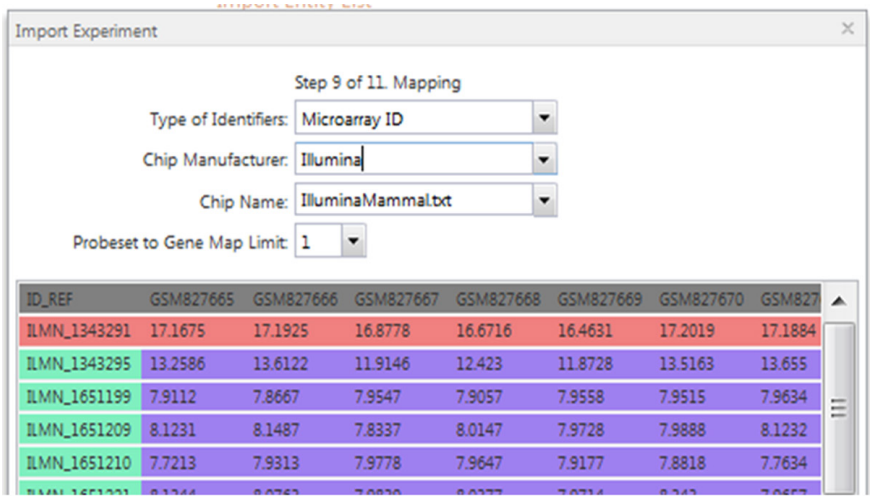
(the 11-step program)

And Follow the Yellow Brick Road!



This works particularly well because it will automatically import the phenotype data.





Just make sure to specify the correct identifier type (Hint: Affymetrix and Illumina are among the most popular.).

Appendix 1.B

Data Import from a Tab Delimited File

Example 1:

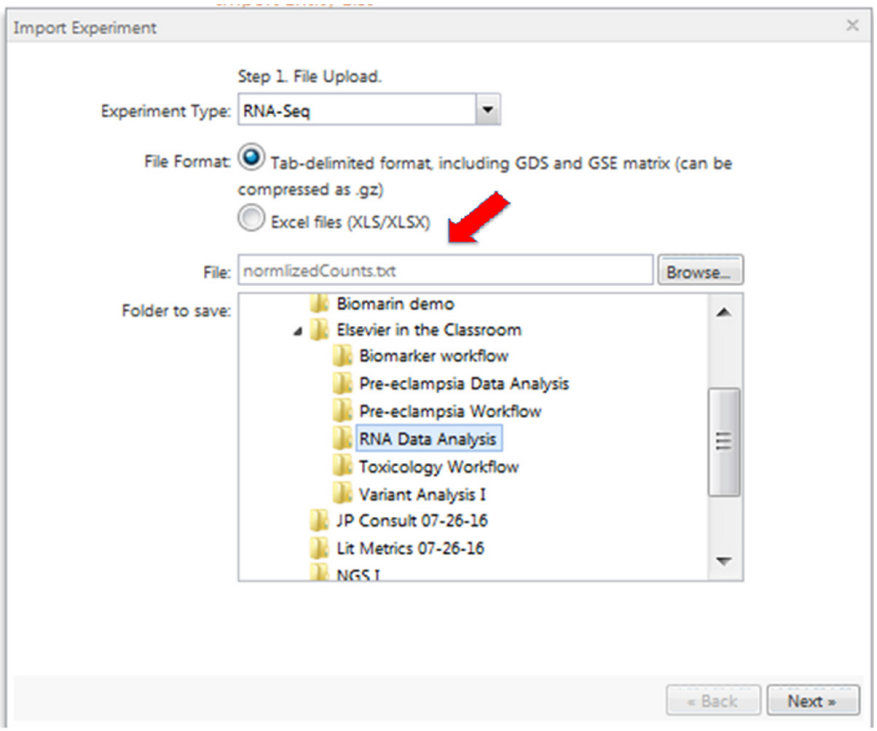
Class 1 = CC - 6,8,2,4
Class 2 = CC -1,4,5,7

Gene identifier = Entrez GeneID.

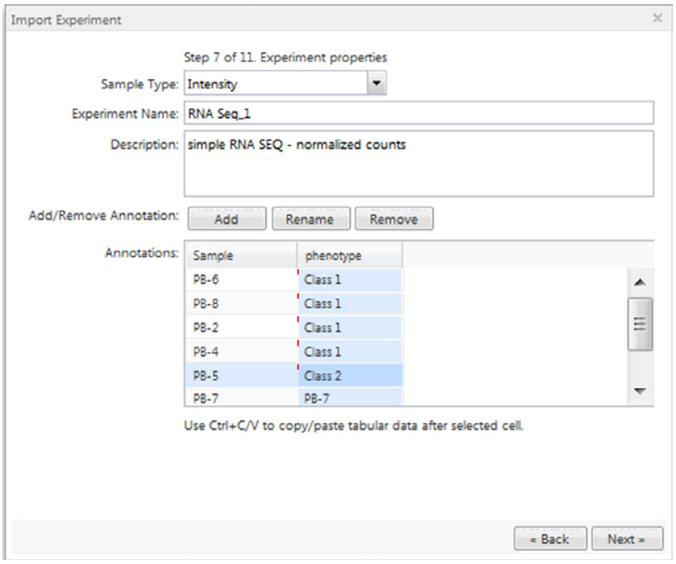
EntrezID	CC-6	CC-8	CC-2	CC-4	CC-5	CC-7	CC-1	CC-3
497097	31.75776	34.09936	25.46663	30.25447	26.30409	18.32842	25.24533	17.16282
100503874	5.60431	11.36645	15.67178	9.454523	1.878864	1.832842	5.049066	0
100038431	0	0	0	0	0	1.832842	0	0
19888	0	0	0	0	0	1.832842	0	0
20671	1.868103	3.788818	3.917944	9.454523	3.757727	3.665685	5.049066	1.716282
27395	44.83448	30.31055	41.13841	35.92719	35.69841	38.48969	31.97742	44.62334
18777	50.43879	45.46582	45.05635	49.16352	26.30409	40.32253	28.61137	36.04193
100503730	1.868103	0	1.958972	0	0	1.832842	0	0
21399	99.00948	77.67077	105.7845	54.83624	67.63909	111.8034	119.4946	97.8281
58175	3.736207	7.577636	1.958972	5.672714	3.757727	9.164212	8.41511	5.148847
108664	160.6569	168.6024	178.2664	151.2724	114.6107	159.4573	186.8154	157.898
18387	0	0	0	0	0	1.832842	0	0
12421	226.0405	212.1738	248.7894	206.1086	191.6441	214.4426	255.8193	293.4843
240690	192.4146	215.9626	188.0613	240.1449	217.9482	201.6127	203.6457	207.6702
319263	246.5896	234.9067	276.215	264.7267	223.5848	285.9234	316.4081	305.4983
71096	18.68103	9.472046	21.54869	7.563619	7.515455	20.16127	13.46418	22.31167
59014	11.20862	17.04968	3.917944	11.34543	15.03091	12.8299	11.78115	5.148847
76187	9.340517	5.683227	9.79486	5.672714	1.878864	7.33137	6.732088	1.716282
72481	0	0	1.958972	0	0	3.665685	0	1.716282
76982	13.07672	17.04968	11.75383	7.563619	3.757727	10.99705	11.78115	17.16282
17864	7.472413	11.36645	17.63075	18.90905	18.78864	14.66274	21.87929	25.74424
70675	69.11982	75.77636	60.72813	62.39985	41.335	62.31664	70.68692	53.20476
170755	20.54914	17.04968	9.79486	17.01814	13.15205	25.65979	13.46418	18.87911
620986	1.868103	1.894409	0	1.890905	1.878864	0	0	0
240697	1.868103	3.788818	0	1.890905	5.636591	3.665685	1.683022	6.86513
73824	1.868103	7.577636	9.79486	9.454523	5.636591	14.66274	18.51324	15.44654
266793	0	1.894409	0	0	0	0	0	0
100038398	0	1.894409	0	0	0	0	0	0
69312	0	0	0	0	0	1.832842	0	0

Find data at:
PS tab delimited data set for import – example1_10-06-16.txt

Locate the data in your files.

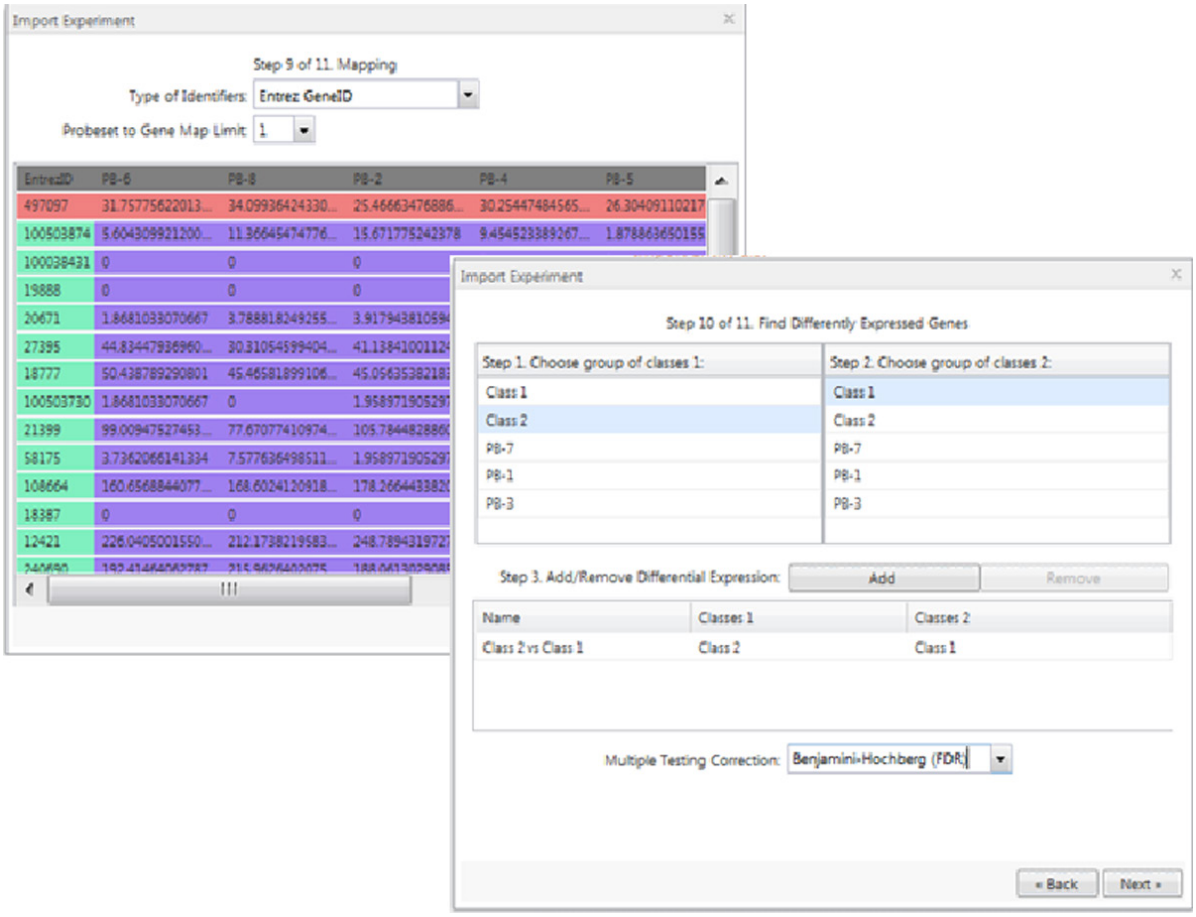


Manually input the phenotype class names.

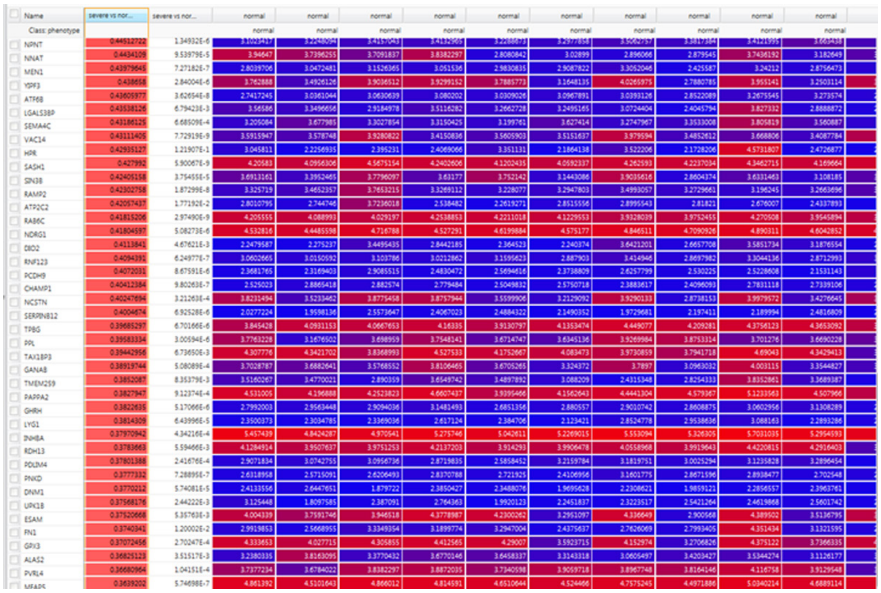


Hint: Copy and Paste (using Ctrl C and Ctrl V) the sample and phenotype columns into Excel. Rename as needed and then copy and paste back into Pathway Studio (this is especially helpful for large data sets).

Locate the data in your files.



Now....analyze your data!



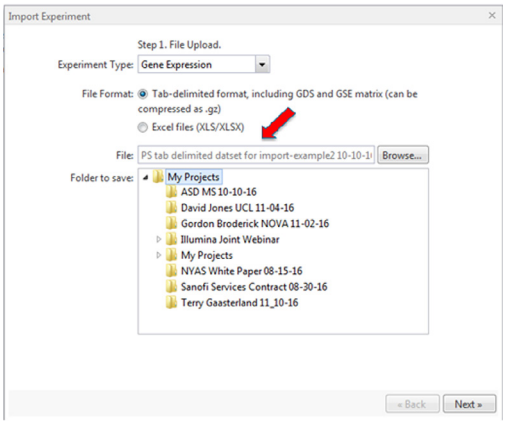
Example 2. Import Fold Change and P-value Data

	A	B	C	D	E	F	G	H	I	J	K
1	EntrezID	Symbol	ki_vs_wt::adjust_p_value	ki_vs_wt::foldChange	GeneName	Chr	Start	End	RefSeqID	EnsemblID	UNIGENE
2	231130	Tnip2	0.46295	5.1541	TNFAIP3 interact	5	-34496096	-34513979	NM_139064;N	ENSMUSG00(	Mm.28615
3	330222	Sdk1	0.46864	4.6492	sidekick homolog	5	141241534	142213791	NM_177879;N	ENSMUSG00(	Mm.151931;M
4	50787	Hs6st3	0.46864	4.1528	heparan sulfate (	14	119138265	119869815	NM_015820;N	ENSMUSG00(	Mm.445777
5	21405	Hnf1a	0.46864	4.1521	HNF1 homeobox	5	-114948980	-114971062	NM_009327;N	ENSMUSG00(	Mm.332607
6	66589	Ube2v1	0.31656	2.9121	ubiquitin-conjuga	2	-167607639	-167632005	NM_023230;N	ENSMUSG00(	Mm.278783;M
7	19791	Rn18s	0.21073	2.3094	18S ribosomal RN	6	NA	NA	NR_003278	NA	NA
8	102436	Lars2	0.23964	1.9412	leucyl-tRNA syntl	9	123366940	123462664	NM_153168;N	ENSMUSG00(	Mm.276076
9	72003	Synpr	0.20006	1.8036	synaptoporin	14	13284780;1345	13615469;1361	NM_0011630:	ENSMUSG00(	Mm.317515
10	17263	Meg3	0.20006	1.4666	maternally expre	12	109545398;109	109568600;109	NM_144513;N	ENSMUSG00(	Mm.289645;M
11	100861531	Rn45s	0.20006	1.3918	45S pre-ribosoma	17	39842997	39848829	NR_046233	NA	NA
12	16438	Itpr1	0.20006	-1.2386	inositol 1,4,5-tris	6	108213096	108551116	NM_010585;N	ENSMUSG00(	Mm.227912
13	12307	Calb1	0.23948	-1.3468	calbindin 1	4	15881264	15906709	NM_009788;N	ENSMUSG00(	Mm.277665
14	57295	Icmt	0.25878	-1.429	isoprenylcysteine	4	152297214	152307126	NM_133788;N	ENSMUSG00(	Mm.277464
15	22629	Ywhah	0.46864	-1.4367	tyrosine 3-monoc	5	33018816	33027966	NM_011738;N	ENSMUSG00(	Mm.332314
16	11676	Aldoc	0.16688	-1.4774	aldolase C, fructc	11	78324198	78326760	NM_0013034:	ENSMUSG00(	Mm.7729
17	56298	Atl2	0.46864	-1.4953	atlastin GTPase 2	17	-79848392;-79	-79896028;-79	NM_0012866	ENSMUSG00(	Mm.175403
18	20623	Snrk	0.20006	-1.5039	SNF related kinas	9	122117266	122169702	NM_0011645	ENSMUSG00(	Mm.257989
19	242202	Pde5a	0.23964	-1.5084	phosphodiestera	3	122729158	122859374	NM_153422;N	ENSMUSG00(	Mm.134911
20	67792	Rgs8	0.094407	-1.512	regulator of G-pr	1	153653037	153697665	NM_026380;N	ENSMUSG00(	Mm.379143
21	104175	Sbk1	0.20006	-1.5168	SH3-binding kina	7	126272619	126294999	NM_145587;N	ENSMUSG00(	Mm.29660
22	20513	Slc1a6	0.46295	-1.5268	solute carrier far	10	78780496	78814825	NM_009200;N	ENSMUSG00(	Mm.6257
23	18546	Pcp4	0.20006	-1.5802	Purkinje cell prot	16	96467606	96525793	NM_008791;N	ENSMUSG00(	Mm.5023
24	239217	Kctd12	0.20006	-1.6075	potassium chann	14	-102976581	-102982637	NM_177715;N	ENSMUSG00(	Mm.246466
25	66540	Fam107b	0.20006	-1.75	family with seque	2	3713458	3782134	NM_025626;N	ENSMUSG00(	Mm.277864
26	98758	Hnrnpf	0.46295	-1.7649	heterogeneous n	6	117906782;117	117925622;117	NM_0011664:	ENSMUSG00(	Mm.422979;M
27	110834	Chrna3	0.46864	-4.1142	cholinergic recep	9	-55011343	-55026559	NM_145129;N	ENSMUSG00(	Mm.63569
28	30937	Lmcd1	0.23964	-4.8776	LIM and cysteine	6	112273758	112330423	NM_144799;N	ENSMUSG00(	Mm.234441

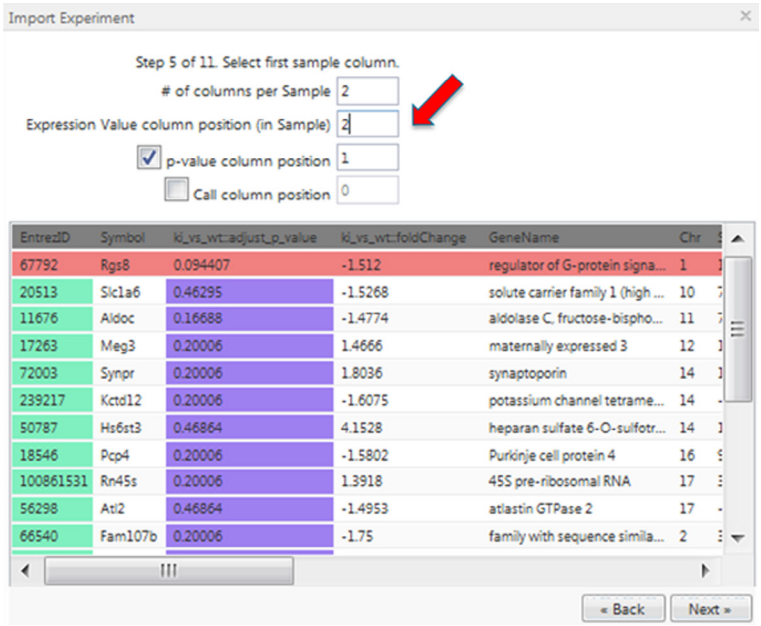
Find data at:

PS tab delimited data set for import – example2_10-10-16.txt

Locate the data in your files.

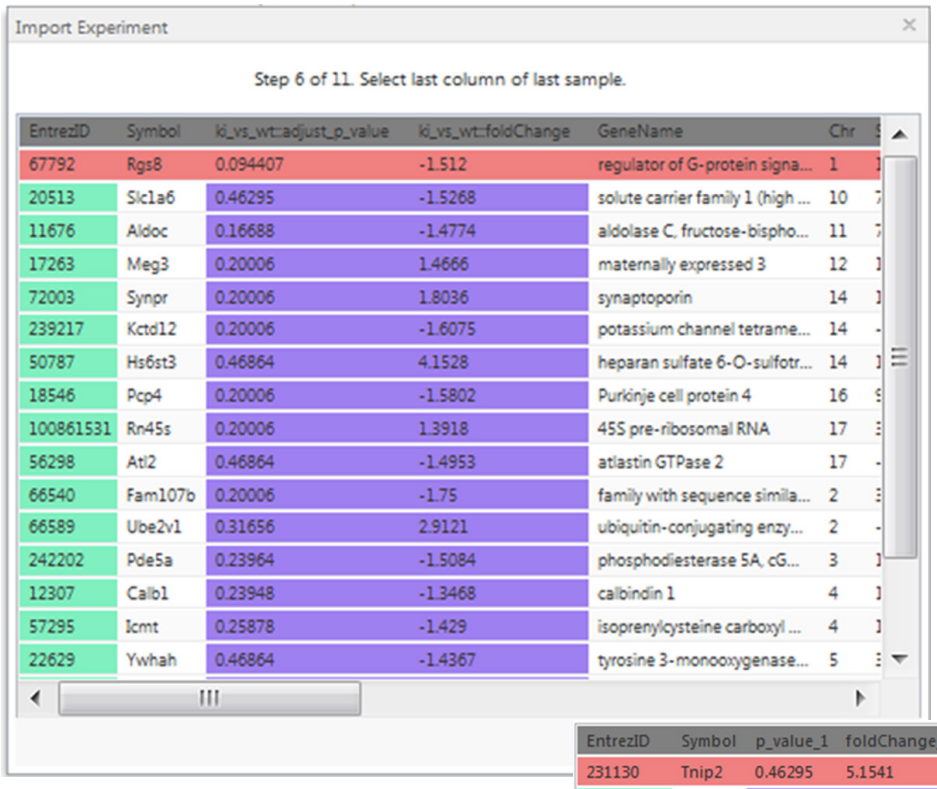


Gene identifier = Entrez GeneID



Indicate data location; fold change and p-value.

Limit data to these two columns (i.e. p-value and fold change).

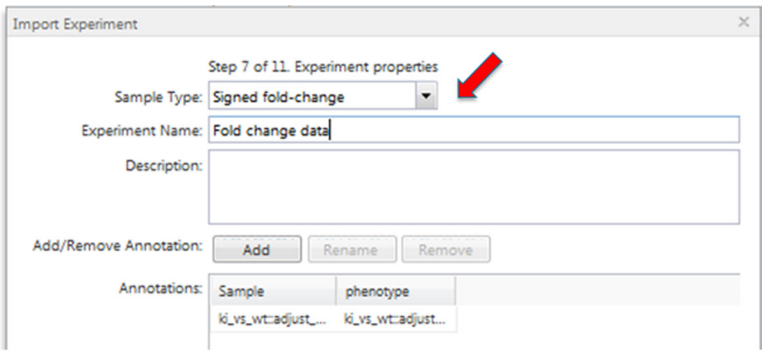


EntrezID	Symbol	ki_vs_wt::adjust_p_value	ki_vs_wt::foldChange	GeneName	Chr
67792	Rgs8	0.094407	-1.512	regulator of G-protein signa...	1
20513	Slc1a6	0.46295	-1.5268	solute carrier family 1 (high ...	10
11676	Aldoc	0.16688	-1.4774	aldolase C, fructose-bispho...	11
17263	Meg3	0.20006	1.4666	maternally expressed 3	12
72003	Synpr	0.20006	1.8036	synaptoporin	14
239217	Kctd12	0.20006	-1.6075	potassium channel tetrame...	14
50787	Hs6st3	0.46864	4.1528	heparan sulfate 6-O-sulfotr...	14
18546	Pcp4	0.20006	-1.5802	Purkinje cell protein 4	16
100861531	Rn45s	0.20006	1.3918	45S pre-ribosomal RNA	17
56298	Atl2	0.46864	-1.4953	atlastin GTPase 2	17
66540	Fam107b	0.20006	-1.75	family with sequence simila...	2
66589	Ube2v1	0.31656	2.9121	ubiquitin-conjugating enzy...	2
242202	Pde5a	0.23964	-1.5084	phosphodiesterase 5A, cG...	3
12307	Calb1	0.23948	-1.3468	calbindin 1	4
57295	Icmt	0.25878	-1.429	isoprenylcysteine carboxyl ...	4
22629	Ywhah	0.46864	-1.4367	tyrosine 3-monooxygenase...	5

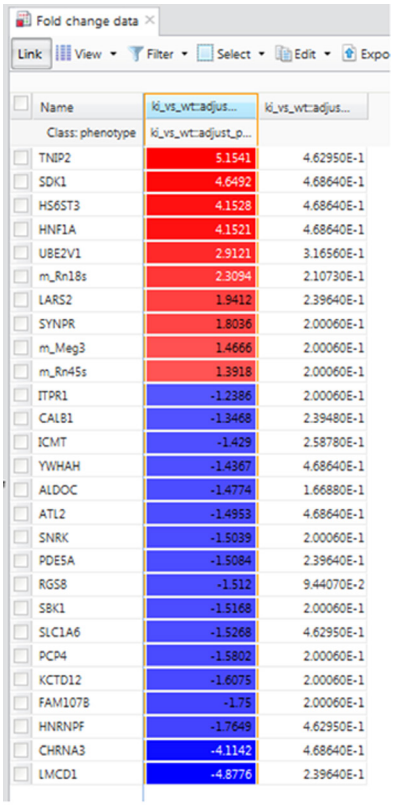
Note: in this example, for simplicity's sake we only have one comparison. In fact, this method could be adopted for any number of different, separate comparisons. Simply adjust the # of columns highlighted (all have to be in the same p-value, fold change order and also have to be grouped together (see below).

EntrezID	Symbol	p_value_1	foldChange_1	p_value_2	foldChange_2	p_value_3	foldChange_3
231130	Tnip2	0.46295	5.1541	0.46295	5.1541	0.46295	5.1541
330222	Sdk1	0.46864	4.6492	0.46864	4.6492	0.46864	4.6492
50787	Hs6st3	0.46864	4.1528	0.46864	4.1528	0.46864	4.1528
21405	Hnf1a	0.46864	4.1521	0.46864	4.1521	0.46864	4.1521

Indicate data type.



Now....analyze your data!



Appendix 1.c

Data Import for RNASeq Data

Complex RNASeq data (from the CuffLinks program)

Find data at:

LungNormal_vs_LungTumor.gene_exp.diff_10-10-16.txt

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	gene	locus	sample_1	sample_2	status	value_1	value_2	FC_T/N	log2(fold_change)	test_stat	p_value	FC_T/N	log2(fold_change)	q_value	significant
2	42430	chr4:1644	LungNorm	LungTumor	OK	0.888877	1.63541	1.839860858	0.879596665	3.10178	0.00015	1.839860858	0.879596665	0.000336887	yes
3	42430	chr1:2209	LungNorm	LungTumor	OK	18.2479	7.90096	0.432979137	-1.207630583	-5.79906	0.00005	0.432979137	-1.207630583	0.00011798	yes
4	42431	chr1:2209	LungNorm	LungTumor	OK	4.59102	9.55956	2.082230093	1.0581295	4.40671	0.00005	2.082230093	1.0581295	0.00011798	yes
5	42431	chr19:847	LungNorm	LungTumor	OK	4.61475	8.07907	1.750705889	0.807936737	2.8992	0.00005	1.750705889	0.807936737	0.00011798	yes
6	42432	chr5:1262	LungNorm	LungTumor	OK	0.962022	0.407359	0.42344042	-1.239769104	-3.10235	0.0001	0.42344042	-1.239769104	0.000228857	yes
7	42433	chr2:2171	LungNorm	LungTumor	NOTEST	0.0110425	0.0189432	1.715481096	0.778613228	0	1	1.715481096	0.778613228	1	no
8	42434	chr10:940	LungNorm	LungTumor	OK	8.41669	3.9645	0.471028397	-1.086114056	-4.99175	0.00005	0.471028397	-1.086114056	0.00011798	yes
9	42435	chr5:1035	LungNorm	LungTumor	OK	47.5576	32.7109	0.687816458	-0.539904457	-2.82576	0.00005	0.687816458	-0.539904457	0.00011798	yes
10	42436	chr2:1605	LungNorm	LungTumor	OK	40.1367	21.811	0.543417869	-0.879866088	-4.64003	0.00005	0.543417869	-0.879866088	0.00011798	yes
11	42437	chr10:459	LungNorm	LungTumor	OK	8.24987	10.7157	1.298893195	0.377282806	1.62279	0.0205	1.298893195	0.377282806	0.0323616	yes
12	42438	chr12:581	LungNorm	LungTumor	OK	0.946702	0.561338	0.592940545	-0.754040644	-1.81427	0.0172	0.592940545	-0.754040644	0.0275972	yes
13	42439	chr17:607	LungNorm	LungTumor	OK	0.536472	0.652436	1.216160396	0.282333514	0.187988	0.8053	1.216160396	0.282333514	0.836655	no
14	42614	chr16:303	LungNorm	LungTumor	OK	1.46039	2.53332	1.73468731	0.79467563	2.00533	0.01285	1.73468731	0.79467563	0.0211307	yes
15	42615	chr2:2421	LungNorm	LungTumor	OK	37.9996	61.7796	1.625796061	0.701146298	1.85005	0.009	1.625796061	0.701146298	0.015243	yes
16	42616	chr22:423	LungNorm	LungTumor	OK	1.04544	1.09603	1.048391108	0.068177023	0.176019	0.80155	1.048391108	0.068177023	0.833478	no
17	42617	chr17:565	LungNorm	LungTumor	OK	3.06323	15.6396	5.105591157	2.352078015	8.52392	0.00005	5.105591157	2.352078015	0.00011798	yes
18	42618	chr22:197	LungNorm	LungTumor	OK	12.3691	3.63608	0.293964799	-1.766284684	-5.7979	0.00005	0.293964799	-1.766284684	0.00011798	yes
19	42619	chrX:1187	LungNorm	LungTumor	OK	2.84457	5.64163	1.983298003	0.987901468	3.76105	0.00005	1.983298003	0.987901468	0.00011798	yes
20	42620	chr7:3584	LungNorm	LungTumor	OK	46.1913	56.5115	1.223423026	0.290923334	1.50383	0.0314	1.223423026	0.290923334	0.0477865	yes
21	42621	chr5:1320	LungNorm	LungTumor	OK	4.00653	4.6348	1.156811505	0.210153806	0.84747	0.2275	1.156811505	0.210153806	0.281308	no
22	42622	chr17:752	LungNorm	LungTumor	OK	106.781	110.778	1.037431753	0.053016432	0.278004	0.68845	1.037431753	0.053016432	0.73395	no
23	42623	chr2:1103	LungNorm	LungTumor	OK	57.1868	36.4956	0.638182238	-0.647959639	-3.31384	0.00005	0.638182238	-0.647959639	0.00011798	yes
24	42624	chr4:7787	LungNorm	LungTumor	OK	15.9228	26.6012	1.670635818	0.740397274	3.86171	0.00005	1.670635818	0.740397274	0.00011798	yes
25	42627	chr7:5586	LungNorm	LungTumor	NOTEST	0.0527852	0.0305913	0.57954313	-0.787012065	0	1	0.57954313	-0.787012065	1	no
26	42628	chr1:8732	LungNorm	LungTumor	OK	17.5671	19.1205	1.088426661	0.122244201	0.573455	0.4174	1.088426661	0.122244201	0.476824	no
27	A1BG	chr19:588	LungNorm	LungTumor	OK	0.274634	0.480508	1.749630417	0.807050207	1.04905	0.1744	1.749630417	0.807050207	0.222972	no
28	A1BG-AS1	chr19:588	LungNorm	LungTumor	OK	0.120942	0.333523	2.757710307	1.463470912	1.41657	0.1016	2.757710307	1.463470912	0.137601	no
29	A2M	chr12:921	LungNorm	LungTumor	OK	53.9448	1066.04	19.76168231	4.304633864	22.4442	0.00005	19.76168231	4.304633864	0.00011798	yes
30	A2M-AS1	chr12:921	LungNorm	LungTumor	OK	0.27323	0.482627	1.766376313	0.820792731	0.0886351	0.654	1.766376313	0.820792731	0.702531	no
31	A2ML1	chr12:897	LungNorm	LungTumor	OK	9.77056	0.0549773	0.005626832	-7.473461386	-9.22745	0.00005	0.005626832	-7.473461386	0.00011798	yes
32	A4GALT	chr22:430	LungNorm	LungTumor	OK	18.9259	5.0193	0.265207995	-1.914803825	-8.70986	0.00005	0.265207995	-1.914803825	0.00011798	yes

Step 1. Identify columns to be imported, in this case, log2 fold change and q-value. Hint: you may need to change the original position of the fold change and p-value columns so that they are located side-by-side in the final spreadsheet import.

The CuffDiff output provides a significance call for each calculated estimate of differential expression (see last column above).

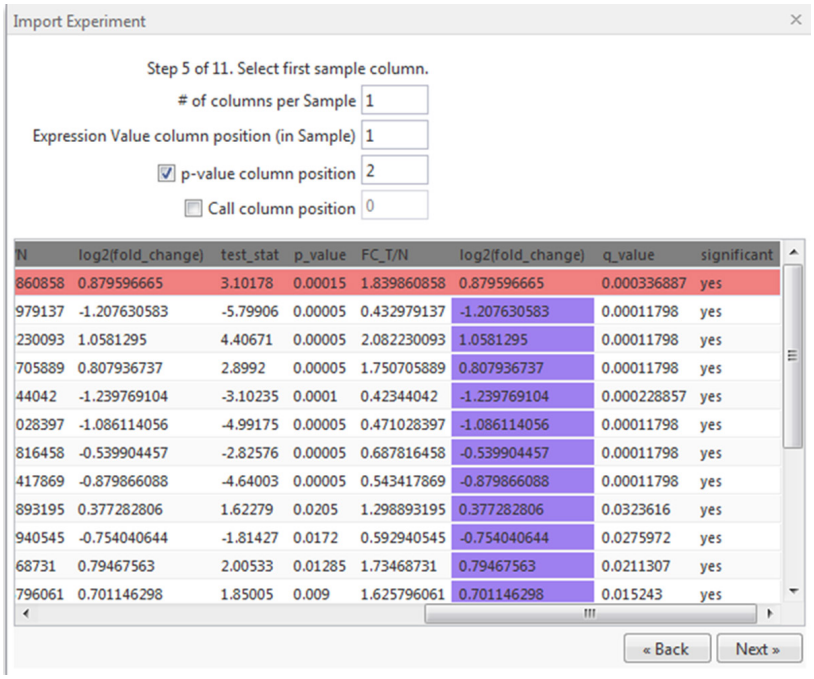
It is strongly recommended that the user pre-filter their data to include only significant calls. This will eliminate, among other things, high fold changes measured at or below background levels on both sides. This will NOT be evident after the data has been imported into Pathway Studio and could seriously (and silently) impact the resulting analysis. The data below reflects this pre-filtering step.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	gene	locus	sample_1	sample_2	status	value_1	value_2	FC_T/N	log2(fold_	test_stat	p_value	FC_T/N	log2(fold_	q_value	significant
2	42430	chr4:1644	LungNorm	LungTumc	OK	0.888877	1.63541	1.839861	0.879597	3.10178	0.00015	1.839861	0.879597	0.000337	yes
3	42430	chr1:2209	LungNorm	LungTumc	OK	18.2479	7.90096	0.432979	-1.20763	-5.79906	0.00005	0.432979	-1.20763	0.000118	yes
4	42431	chr1:2209	LungNorm	LungTumc	OK	4.59102	9.55956	2.08223	1.05813	4.40671	0.00005	2.08223	1.05813	0.000118	yes
5	42431	chr19:847	LungNorm	LungTumc	OK	4.61475	8.07907	1.750706	0.807937	2.8992	0.00005	1.750706	0.807937	0.000118	yes
6	42432	chr5:1262	LungNorm	LungTumc	OK	0.962022	0.407359	0.42344	-1.23977	-3.10235	0.0001	0.42344	-1.23977	0.000229	yes
7	42434	chr10:940	LungNorm	LungTumc	OK	8.41669	3.9645	0.471028	-1.08611	-4.99175	0.00005	0.471028	-1.08611	0.000118	yes
8	42435	chr5:1035	LungNorm	LungTumc	OK	47.5576	32.7109	0.687816	-0.5399	-2.82576	0.00005	0.687816	-0.5399	0.000118	yes
9	42436	chr2:1605	LungNorm	LungTumc	OK	40.1367	21.811	0.543418	-0.87987	-4.64003	0.00005	0.543418	-0.87987	0.000118	yes
10	42437	chr10:459	LungNorm	LungTumc	OK	8.24987	10.7157	1.298893	0.377283	1.62279	0.0205	1.298893	0.377283	0.032362	yes
11	42438	chr12:581	LungNorm	LungTumc	OK	0.946702	0.561338	0.592941	-0.75404	-1.81427	0.0172	0.592941	-0.75404	0.027597	yes
12	42614	chr16:303	LungNorm	LungTumc	OK	1.46039	2.53332	1.734687	0.794676	2.00533	0.01285	1.734687	0.794676	0.021131	yes
13	42615	chr2:2421	LungNorm	LungTumc	OK	37.9996	61.7796	1.625796	0.701146	1.85005	0.009	1.625796	0.701146	0.015243	yes
14	42617	chr17:565	LungNorm	LungTumc	OK	3.06323	15.6396	5.105591	2.352078	8.52392	0.00005	5.105591	2.352078	0.000118	yes
15	42618	chr22:197	LungNorm	LungTumc	OK	12.3691	3.63608	0.293965	-1.76628	-5.7979	0.00005	0.293965	-1.76628	0.000118	yes
16	42619	chrX:1187	LungNorm	LungTumc	OK	2.84457	5.64163	1.983298	0.987901	3.76105	0.00005	1.983298	0.987901	0.000118	yes
17	42620	chr7:3584	LungNorm	LungTumc	OK	46.1913	56.5115	1.223423	0.290923	1.50383	0.0314	1.223423	0.290923	0.047787	yes
18	42623	chr2:1103	LungNorm	LungTumc	OK	57.1868	36.4956	0.638182	-0.64796	-3.31384	0.00005	0.638182	-0.64796	0.000118	yes
19	42624	chr4:7787	LungNorm	LungTumc	OK	15.9228	26.6012	1.670636	0.740397	3.86171	0.00005	1.670636	0.740397	0.000118	yes
20	A2M	chr12:921	LungNorm	LungTumc	OK	53.9448	1066.04	19.76168	4.304634	22.4442	0.00005	19.76168	4.304634	0.000118	yes
21	A2ML1	chr12:897	LungNorm	LungTumc	OK	9.77056	0.054977	0.005627	-7.47346	-9.22745	0.00005	0.005627	-7.47346	0.000118	yes
22	A4GALT	chr22:430	LungNorm	LungTumc	OK	18.9259	5.0193	0.265208	-1.9148	-8.70986	0.00005	0.265208	-1.9148	0.000118	yes
23	A4GNT	chr3:1378	LungNorm	LungTumc	OK	0.146681	0.598347	4.07924	2.0283	2.40889	0.01205	4.07924	2.0283	0.019945	yes
24	AACS	chr12:125	LungNorm	LungTumc	OK	5.90326	3.5885	0.607884	-0.71813	-3.13482	0.00005	0.607884	-0.71813	0.000118	yes
25	AADAC	chr3:1513	LungNorm	LungTumc	OK	0.829139	1.74367	2.102989	1.072441	2.00692	0.00735	2.102989	1.072441	0.012659	yes
26	AADACL2	chr3:1513	LungNorm	LungTumc	OK	2.06816	0.364689	0.176335	-2.50361	-2.55421	0.0222	0.176335	-2.50361	0.034796	yes
27	AADAT	chr4:1709	LungNorm	LungTumc	OK	9.23837	6.97247	0.754729	-0.40597	-1.73537	0.0139	0.754729	-0.40597	0.02271	yes
28	AAGAB	chr15:674	LungNorm	LungTumc	OK	12.972	9.93491	0.765873	-0.38482	-1.75333	0.01895	0.765873	-0.38482	0.03013	yes
29	AAMDC	chr11:775	LungNorm	LungTumc	OK	6.15364	23.9798	3.896848	1.962308	7.90551	0.00005	3.896848	1.962308	0.000118	yes
30	AAR2	chr20:348	LungNorm	LungTumc	OK	9.21531	6.73842	0.73122	-0.45162	-1.96335	0.0045	0.73122	-0.45162	0.008046	yes
31	AARS	chr16:702	LungNorm	LungTumc	OK	68.7434	39.2857	0.571483	-0.80722	-4.34398	0.00005	0.571483	-0.80722	0.000118	yes
32	AARS2	chr6:4426	LungNorm	LungTumc	OK	5.13432	1.64905	0.321182	-1.63854	-6.77488	0.00005	0.321182	-1.63854	0.000118	yes
33	AASDH	chr4:5720	LungNorm	LungTumc	OK	6.86677	9.32449	1.357915	0.441393	2.10929	0.00375	1.357915	0.441393	0.006785	yes

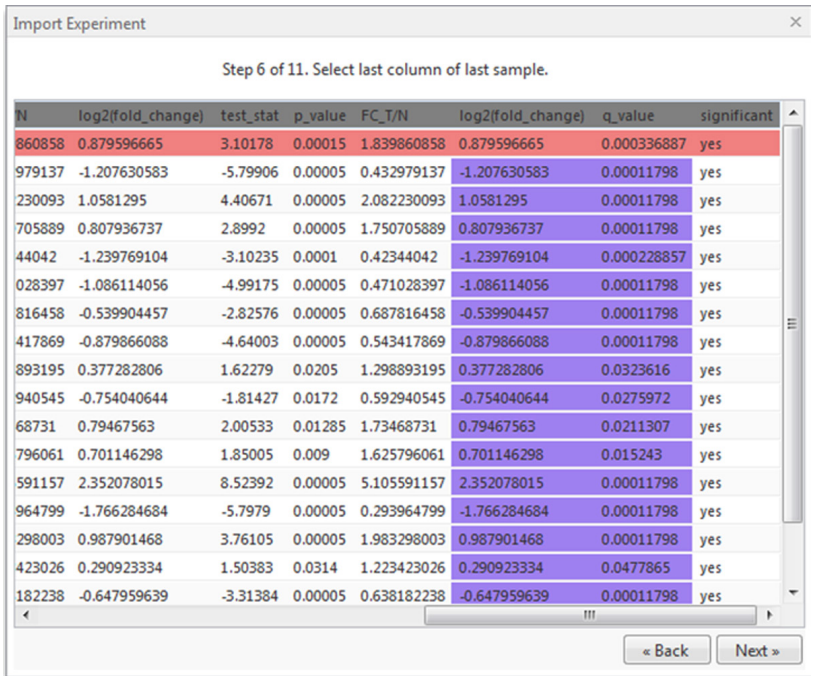
Find data at:

LungNormal_vs_LungTumor.gene_exp.diff2_10-10-16.txt

Import fold change and q-value data the same as in Appendix I.B, Example 2.

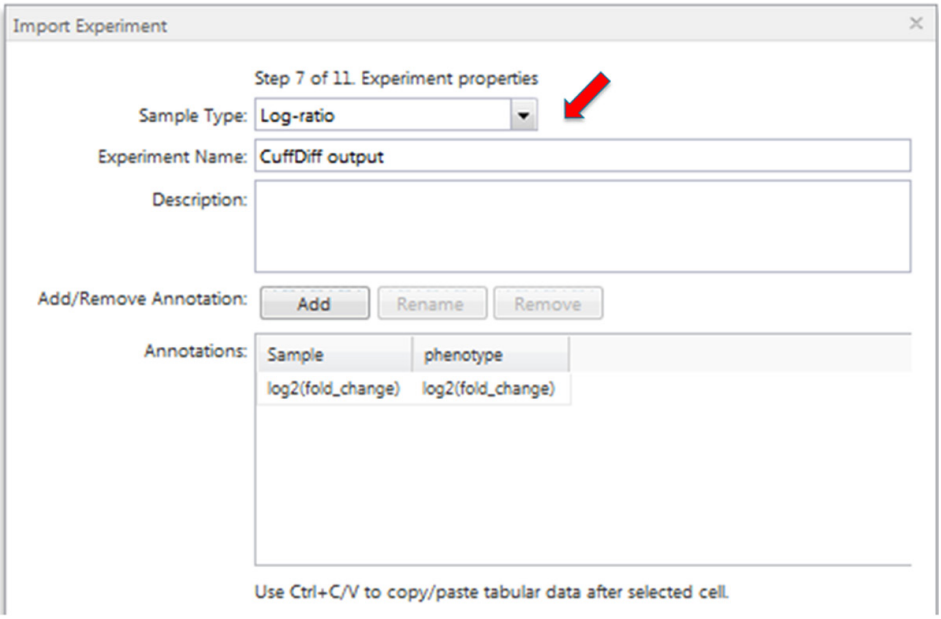


N	log2(fold_change)	test_stat	p_value	FC_T/N	log2(fold_change)	q_value	significant
860858	0.879596665	3.10178	0.00015	1.839860858	0.879596665	0.000336887	yes
979137	-1.207630583	-5.79906	0.00005	0.432979137	-1.207630583	0.00011798	yes
230093	1.0581295	4.40671	0.00005	2.082230093	1.0581295	0.00011798	yes
705889	0.807936737	2.8992	0.00005	1.750705889	0.807936737	0.00011798	yes
44042	-1.239769104	-3.10235	0.0001	0.42344042	-1.239769104	0.000228857	yes
028397	-1.086114056	-4.99175	0.00005	0.471028397	-1.086114056	0.00011798	yes
816458	-0.539904457	-2.82576	0.00005	0.687816458	-0.539904457	0.00011798	yes
417869	-0.879866088	-4.64003	0.00005	0.543417869	-0.879866088	0.00011798	yes
893195	0.377282806	1.62279	0.0205	1.298893195	0.377282806	0.0323616	yes
940545	-0.754040644	-1.81427	0.0172	0.592940545	-0.754040644	0.0275972	yes
68731	0.79467563	2.00533	0.01285	1.73468731	0.79467563	0.0211307	yes
796061	0.701146298	1.85005	0.009	1.625796061	0.701146298	0.015243	yes

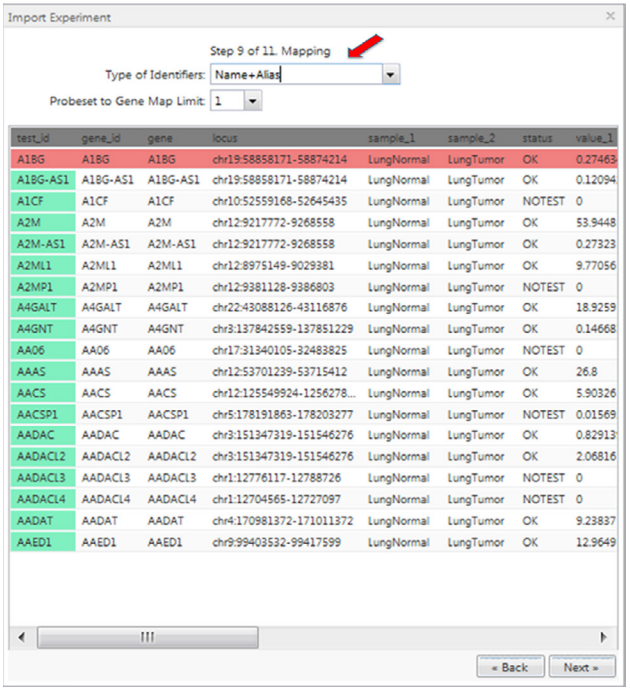


N	log2(fold_change)	test_stat	p_value	FC_T/N	log2(fold_change)	q_value	significant
860858	0.879596665	3.10178	0.00015	1.839860858	0.879596665	0.000336887	yes
979137	-1.207630583	-5.79906	0.00005	0.432979137	-1.207630583	0.00011798	yes
230093	1.0581295	4.40671	0.00005	2.082230093	1.0581295	0.00011798	yes
705889	0.807936737	2.8992	0.00005	1.750705889	0.807936737	0.00011798	yes
44042	-1.239769104	-3.10235	0.0001	0.42344042	-1.239769104	0.000228857	yes
028397	-1.086114056	-4.99175	0.00005	0.471028397	-1.086114056	0.00011798	yes
816458	-0.539904457	-2.82576	0.00005	0.687816458	-0.539904457	0.00011798	yes
417869	-0.879866088	-4.64003	0.00005	0.543417869	-0.879866088	0.00011798	yes
893195	0.377282806	1.62279	0.0205	1.298893195	0.377282806	0.0323616	yes
940545	-0.754040644	-1.81427	0.0172	0.592940545	-0.754040644	0.0275972	yes
68731	0.79467563	2.00533	0.01285	1.73468731	0.79467563	0.0211307	yes
796061	0.701146298	1.85005	0.009	1.625796061	0.701146298	0.015243	yes
591157	2.352078015	8.52392	0.00005	5.105591157	2.352078015	0.00011798	yes
964799	-1.766284684	-5.7979	0.00005	0.293964799	-1.766284684	0.00011798	yes
298003	0.987901468	3.76105	0.00005	1.983298003	0.987901468	0.00011798	yes
423026	0.290923334	1.50383	0.0314	1.223423026	0.290923334	0.0477865	yes
182238	-0.647959639	-3.31384	0.00005	0.638182238	-0.647959639	0.00011798	yes

Indicate data type (in this case, log2 of fold change = Log-ratio).



Indicate probe identity type.



Hint: when the probe identity uses gene symbol (as in this case) it's often best to use the Name+Alias designation as this will capture both current and past gene symbol assignments for the same gene/protein entity.

Now....analyze your data!

LungNormal_vs_...

Link | View | Filter | Select

Name	log2(fold_ch...	q_value
Class: phenotype	log2(fold_chan...	q_value
SFTPC	13.147715	1.1798E-4
SFTPB	12.085524	1.1798E-4
SFTPA1	12.040313	1.1798E-4
SFTPA2	10.691367	1.1798E-4
NAPSA	10.454914	7.48607E-4
PGC	10.358013	0.00703833
SCGB3A2	9.754293	1.1798E-4
AQP4	9.564383	0.0169025
AGTR2	9.491557	0.00152403
SFTPD	9.351456	0.0140808
CACNA2D2	8.7339525	1.1798E-4
CLDN18	8.6818905	0.00199503
TMEM100	8.601007	0.0154719
SCN1A	8.503489	1.1798E-4
HHIP	8.458812	1.1798E-4
AGBL1	8.294703	0.0198685
VEPH1	8.231634	1.1798E-4
AGER	8.11477	1.1798E-4
CA4	7.9945545	0.00862657
NKX2-1	7.986033	9.46723E-4
TRHDE	7.9664173	0.0350243
SLC6A4	7.7363644	1.1798E-4
RASGRF1	7.7332873	1.1798E-4
FCN3	7.617541	1.1798E-4
WIF1	7.583697	1.1798E-4
GRIA1	7.403018	1.1798E-4
FOXA2	7.364509	0.0047396
NDNF	7.3066463	1.1798E-4
C16orf89	7.2883987	6.47053E-4
ANKRD1	7.232523	1.1798E-4
ADH1B	7.1854696	1.1798E-4
CRTAC1	7.0702586	0.00152403
MCEMP1	7.055917	1.1798E-4
PKHD1L1	7.0241814	1.1798E-4
COL6A6	6.9559903	1.1798E-4
TYRP1	6.9388504	2.28857E-4

Appendix 2

Pathway Studio Data Model

The type of entities you want to connect in pathway view will define a set of allowed relations; the choice of relation type defines additional parameters such as Mechanism, BiomarkerType, ChangeType, and QuantitativeType.

Relation type menu option	RNEF Control type	Effect	Mechanism	Biomarker Type	Change Type	Quantitative Type	Is Direct Physical	Allowed types for Entity 1	Allowed types for Entity 2
Binds and activates	Direct Regulation	positive					Direct Physical	Protein, FunctionalClass, Complex, SmallMol	Protein, FunctionalClass, Complex, SmallMol
Binds and inhibits	Direct Regulation	negative					Direct Physical	Protein, FunctionalClass, Complex, SmallMol	Protein, FunctionalClass, Complex, SmallMol
Binds	Binding						Direct Physical	Protein, FunctionalClass, Complex, SmallMol	Protein, FunctionalClass, Complex, SmallMol
Biomarker for	Biomarker							Protein, FunctionalClass, Complex, SmallMol	CellProcess, Disease, ClinicalParamter
Diagnostic Biomarker for	Biomarker			diagnostic				Protein, FunctionalClass, Complex, SmallMol	CellProcess, Disease, ClinicalParamter
Prognostic Biomarker for	Biomarker			prognostic				Protein, FunctionalClass, Complex, SmallMol	CellProcess, Disease, ClinicalParamter
Expressed in	Cell Expression							Protein, FunctionalClass, Complex	CellType
Catalyzes	Chem Reaction						Direct Physical	Protein, FunctionalClass, Complex	SmallMol
Activates expression	Expression	positive						Protein, FunctionalClass, Complex, SmallMol	Protein, FunctionalClass, Complex
Inhibits expression	Expression	negative						Protein, FunctionalClass, Complex, SmallMol	Protein, FunctionalClass, Complex
Genetically linked	Genetic Change							Protein, FunctionalClass, Complex	Disease

Epigenetically controlled in	GeneticChange				epigenetic			Protein, FunctionalClass, Complex	Disease
Gene mutated in	GeneticChange				mutation			Protein, FunctionalClass, Complex	Disease
Gene deleted in	GeneticChange				gene deletion			Protein, FunctionalClass, Complex	Disease
Gene amplified in	GeneticChange				gene amplification			Protein, FunctionalClass, Complex	Disease
Activate synthesis	MolSynthesis	positive						Protein, FunctionalClass, Complex, SmallMol	SmallMol
Activate degradation	MolSynthesis	negative						Protein, FunctionalClass, Complex, SmallMo	SmallMol
imports	MolTransport		import				Direct Physical	Protein, FunctionalClass, Complex	Protein, FunctionalClass, Complex, SmallMol
Exports	MolTransport		export				Direct Physical	Protein, FunctionalClass, Complex	Protein, FunctionalClass, Complex, SmallMol
Activates transport	MolTransport	positive						Protein, FunctionalClass, Complex	Protein, FunctionalClass, Complex, SmallMol
Inhibits transport	MolTransport	negative						Protein, FunctionalClass, Complex	Protein, FunctionalClass, Complex, SmallMol
Binds gene promoter	Promoter Binding						Direct Physical	Protein, FunctionalClass, Complex	Protein
Binds gene promoter to activate gene expression	Promoter Binding	positive					Direct Physical	Protein, FunctionalClass, Complex	Protein

Binds gene promoter to inhibit gene expression	Promoter Binding	negative					Direct Physical	Protein, FunctionalClass, Complex	Protein
Increase activity in	Quantitative Change	positive				activity		Protein, FunctionalClass, Complex	CellProcess, Disease
Decrease activity in	Quantitative Change	negative				activity		Protein, FunctionalClass, Complex	CellProcess, Disease
Increase abundance in	Quantitative Change	positive				abundance		Protein, FunctionalClass, Complex, SmallM	CellProcess, Disease
Decrease abundance in	Quantitative Change	negative				abundance		Protein, FunctionalClass, Complex, SmallMol	CellProcess, Disease
Increase expression in	Quantitative Change	positive				expression		Protein, FunctionalClass, Complex	CellProcess, Disease
Decrease expression in	Quantitative Change	negative				expression		Protein, FunctionalClass, Complex, Treatment	CellProcess, Disease
Activates or Induces	Regulation	positive						Protein, FunctionalClass, Complex, SmallMol, Treatment	Protein, FunctionalClass, Complex, CellProcess, Disease, ClinicalParamter
Inhibits or Diminishes	Regulation	negative						Protein, FunctionalClass, Complex, SmallMol, Treatment	Protein, FunctionalClass, Complex, CellProcess, Disease, ClinicalParamter
Change phosphorylation status in	StateChange				phospho rylation			Protein, FunctionalClass, Complex	CellProcess, Disease, ClinicalParamter
Change splicing pattern in	StateChange				alternative splcing			Protein	CellProcess, Disease, ClinicalParamter

Altered in	StateChange							Protein, FunctionalClass, Complex	CellProcess, Disease, ClinicalParamter
Phospho rylates	ProtModification		phospho rylation				Direct Physical	Protein, FunctionalClass, Complex	Protein, FunctionalClass, Complex
Phospho rylates to activate	Prot Modification	positive	phospho rylation				Direct Physical	Protein, FunctionalClass, Complex	Protein, FunctionalClass, Complex
Phosphorylates to inhibit	Prot Modification	negative	phosphorylation				Direct Physica	Protein, FunctionalClass, Complex	Protein, FunctionalClass, Complex
Ubiquitinates	Prot Modification	negative	ubiquitination				Direct Physical	Protein, FunctionalClass, Complex	Protein, FunctionalClass, Complex
Acetylates	Prot Modification		acetylation				Direct Physical	Protein, FunctionalClass, Complex	Protein, FunctionalClass, Complex
Acetylates to activate	Prot Modification	positive	acetylation				Direct Physical	Protein, FunctionalClass, Complex	Protein, FunctionalClass, Complex
Acetylates to inhibit	Prot Modification	negative	acetylation				Direct Physical	Protein, FunctionalClass, Complex	Protein, FunctionalClass, Complex
Methylates	Prot Modification		methylation				Direct Physical	Protein, FunctionalClass, Complex	Protein, FunctionalClass, Complex
Methylates to activate	Prot Modification	positive	methylation				Direct Physical	Protein, FunctionalClass, Complex	Protein, FunctionalClass, Complex
Methylates to inhibit	Prot Modification	negative	methylation				Direct Physical	Protein, FunctionalClass, Complex	Protein, FunctionalClass, Complex
Glycosylates	Prot Modification		glycosylation				Direct Physical	Protein, FunctionalClass, Complex	Protein, FunctionalClass, Complex

Glycosylates to activate	ProtModification	positive	glycosylation				Direct Physical	Protein, FunctionalClass, Complex	Protein, FunctionalClass, Complex
Glycosylates to inhibit	ProtModification	negative	glycosylation				Direct Physical	Protein, FunctionalClass, Complex	Protein, FunctionalClass, Complex
Activates Expression	Expression	positive						Protein, FunctionalClass, Complex, SmallMol, Treatment	Protein, FunctionalClass, Complex
Inhibits Expression	Expression	negative						Protein, FunctionalClass, Complex, SmallMol, Treatment	Protein, FunctionalClass, Complex
Cleaves to activate	ProtModification	positive						Protein, FunctionalClass, Complex	Protein, FunctionalClass, Complex
Cleaves to inhibit	ProtModification	negative						Protein, FunctionalClass, Complex	Protein, FunctionalClass, Complex
Induce cleavage or degradation	Expression	negative						Protein, FunctionalClass, Complex, SmallMol, Treatment	Protein, FunctionalClass, Complex
Inhibit cleavage or degradation	Expression							Protein, FunctionalClass, Complex, SmallMol, Treatment	Protein, FunctionalClass, Complex