

PREFACE

This tutorial is designed to demonstrate several tools built into MetaCore™, including data management, enrichment analysis, and data mining plus creating custom maps with MapEditor™. Each section starts with an introduction followed by step-by-step guides and examples of how to interpret the results. Each section also includes exercises to encourage your own analysis. Throughout the manual take advantage of highlighted notes including:

- a) The Applications Box at the beginning of each section as a “when to use” guide
- b) Helpful Hints: designed to highlight key features of the tool
- c) Objectives and Analytical Observations: to demonstrate analysis goals and an example of how to interpret the results.

Introduction to Training Goals

We are now in an era where generation of high-throughput molecular data and its analysis have become an integral part of disease biology research. The “OMICs” field has expanded to include proteomics, metabolomics, chemogenomics and genotyping, allowing scientists to easily interrogate mechanisms of diseases on multiple levels concurrently: from miRNA, DNA, to gene expression, to protein behavior and metabolic or chemical disturbances. This novel, integrative approach to analysis, otherwise known as **systems biology**, is an emerging field with emerging demands where data clustering and statistical analysis are no longer sufficient for meaningful data mining in the context of disease as they are devoid of any functional assessment (1). The demand for **functional assessment** of OMICs science requires the ability to combine mining of knowledge bases, statistical analysis, graph theory, and mathematical modeling. *GeneGo has developed the tools that envelop these demands to provide an efficient and thorough analysis of OMICs data and more.*

Introduction Course Case Study: The role of gamma-secretase in T-cell lymphoma

To demonstrate GeneGo’s MetaCore™ platform we have chosen a gene expression microarray profile from a T-cell line treated with a gamma (γ)-secretase inhibitor, as an example. Gamma -secretase is a multi-subunit protease complex that cleaves single-pass transmembrane proteins at residues within the transmembrane domain (2, 3). One example of a gamma secretase substrate is the amyloid precursor protein that when cleaved produces a short 39-42 amino acid peptide called β amyloid. Abnormally folded fibril forms of amyloid peptides ($A\beta$) are the primary components of amyloid plaques found in the brains of Alzheimer's disease patients. Therefore, inhibitors of γ -secretase have been extensively investigated for the treatment of Alzheimer’s disease.

γ -Secretase is also critical for the processing of Notch proteins. It specifically cleaves Notch, which leads to nuclear translocation of the intracellular domain of this receptor causing changes in the expression of downstream genes that are critical for cellular differentiation during development, and adult tissue homeostasis (2-4). Notch is a known oncogene, first discovered in human T-cell acute lymphoblastic leukemia (T-ALL) (5, 6). Molecular and genetic alterations at the Notch receptor, such as chromosomal translocation or amplification and point mutations have been observed in solid tumors from cervical, head and neck, endometrial, renal, lung, pancreatic, ovarian, breast and prostate carcinomas, osteosarcoma, mesothelioma, gliomas, and medulloblastomas (7). Therefore, inhibition of γ -secretase has also become a strategy for cancer treatment by inhibiting downstream Notch events.

Materials:

- 1) Gene expression data from T-ALL, MOLT4 cells following the administration of a γ -secretase inhibitor: DAPT (**notated as GSI**)
<http://www.ncbi.nlm.nih.gov/sites/GDSbrowser?acc=GDS2794>
DAPT (7) has been shown to lower A β -levels in both cerebral spinal fluid and plasma in APP transgenic mice (7).

Note: Data has been previously statistically analyzed with Bioconductor (<http://www.bioconductor.org/>) and only statistically-significant expression changes will be analyzed.

- 2) MetaCore Software

Objective:

During this course we will determine elements of signaling in general and on the Notch Signaling pathway that are affected by the inhibition of γ -secretase activity.

Methods:

- 1) Data mine to extract genes/proteins associated with T-cell lymphoma to analyze in conjunction with data from high thorough-put analysis of a T-cell lymphoma cell line treated with γ -secretase inhibitors.
- 2) Perform functional enrichment analysis of data to determine representative biology.
- 3) Assess connectivity of the data using building network algorithms and interactome to determine if any genes from the data are directly connected or can be regulated by central receptors/transcription factors.
- 4) Personalize a network with MapEditor™.

Additional Training tools:

Visit Genego.com for:

- 1) Sample publications (under science and technology)
- 2) Case studies (under science and technology)
- 3) Application Scientist tool tips (under the training section)
- 4) More training webinars and events (under the training section)

From the MetaCore interface (portal.genego.com) help section, use:

- 1) Searchable index
- 2) Additional manuals
- 3) Recorded training webinars

TABLE OF CONTENTS

PREFACE	1
Introduction to Training Goals.....	1
Introduction Course Case Study: The role of gamma-secretase in T-cell lymphoma	1
EUREKA! KNOWLEDGE MINING SEARCH TOOLS	4
DATA MINING WITH EZ SEARCH	5
EZ DATA MANAGEMENT.....	8
<i>File Types</i>	8
<i>Activating Data</i>	8
<i>Sharing Data</i>	8
UPLOADING AND ACTIVATING DATA.....	9
<i>Uploading data</i>	9
<i>Activating data</i>	11
<i>Verifying data properties</i>	12
<i>Determine expression/signal distribution</i>	13
<i>Changing Thresholds</i>	14
Breakout Session I: Using the Data Manager interface and uploading data.....	15
ANALYSIS OVERVIEW	16
FUNCTIONAL ENRICHMENT ANALYSIS	18
<i>Getting more from your enrichment results</i>	21
Breakout Session II: Determining the biological relevance of your gene list	26
ASSESSING CONNECTIVITY USING NETWORK ALGORITHMS	27
Break Out Group Session III.....	36
ASSESSING CONNECTIVITY USING INTERACTOME ANALYSIS	37
CREATING A CUSTOM MAP WITH MAPEDITOR™	40

EUREKA! KNOWLEDGE MINING SEARCH TOOLS

The benefits of using Eureka! search tools are the following:

- Search results are linked to the MetaCore™ objects via hyperlinks
- Search results can be used directly to build a network
- A list of results can be saved as a gene list or an experiment within MetaCore™ for further analysis such as:
 - Logical gene list operations (finding unions and intersections)
 - Enrichment analysis
 - Building networks
 - Interactome analysis
 - Drug lookup
- A list of results can be exported to Excel.

There are 3 ways to search, browse and export GeneGo content:

- 1) EZ Search: search your favorite gene, protein, miRNA, disease, drug in a Google-like fashion
- 2) Browse Content: search ontology trees in tabular form
- 3) MetaSearch™: perform Boolean-based queries from the GeneGo database without the need to use special query languages. Queries can be simple lists of terms separated by logical operators or be nested, essentially providing several levels of queries. A Boolean search, in particular, refers to a search where the terms exist in a logical relationship within three logical operations (AND, OR, NOT). An example of a Boolean-based query is as follows:

“Determine the list that includes all genes associated with diabetes and not psoriasis”

This course will focus on the EZ Search capability. Please refer to online help for guidance on MetaSearch and Browse Content features.

DATA MINING WITH EZ SEARCH

Applications of EZ Search

One-stop shop for all information associated with a search item (gene, protein, disease, compound)

Use categorized search results to direct further data mining

Have no data? Use search results to create lists and start an analysis workflow

Objectives:

- To determine a list of biomarkers for T cell lymphoma.
- Learn how to work with each search category.

- 1) To use EZ Search enter your search term in the top tool bar (a) in the interface or select the “search and browse tab” to find the “EZ Search” option (b).

The figure illustrates the EZ Search interface and its search results. The top panel shows the main interface with a search bar (a) and a 'Search & Browse Content' button (b). The search bar contains the text 't cell lymphoma'. The search results are organized into categories listed in the top left panel, including Genes (39), Proteins (37), Drugs (80), Small Molecule Drugs (57), Biologics (23), Network Objects (20), Diseases (5), GeneGo Maps (23), and Articles (196). The search results are displayed in a list format, showing the top 5 results for 't cell lymphoma'. The results are organized into categories listed in the top left panel, including Genes (39), Proteins (37), Drugs (80), Small Molecule Drugs (57), Biologics (23), Network Objects (20), Diseases (5), GeneGo Maps (23), and Articles (196). The search results are displayed in a list format, showing the top 5 results for 't cell lymphoma'. The results are organized into categories listed in the top left panel, including Genes (39), Proteins (37), Drugs (80), Small Molecule Drugs (57), Biologics (23), Network Objects (20), Diseases (5), GeneGo Maps (23), and Articles (196). The search results are displayed in a list format, showing the top 5 results for 't cell lymphoma'.

Search Results:

- 1. **Lymphoma, T-Cell**
Mesh Reference: D016399
Synonyms: Lymphoma, T Cell, Lymphoma, T-Cell, Lymphomas, T-Cell, T Cell Lymphoma, T-Cell Lymphoma, T-Cell Lymphomas
- 2. **Lymphoma, T-Cell, Peripheral**
Mesh Reference: D016411
Synonyms: Lymphoma, Peripheral T-Cell, Lymphoma, T Cell, Peripheral, Lymphoma, T-Cell, Peripheral, Lymphomas, Peripheral T-Cell, Peripheral T Cell Lymphoma, Peripheral
- 3. **Lymphoma, T-Cell, Cutaneous**
Mesh Reference: D016411
Synonyms: Lymphoma, Cutaneous T-Cell, Lymphoma, T Cell, Cutaneous, Cutaneous T-Cell Lymphoma, Cutaneous T-Cell Lymphomas, Granulomatous Slack Skin, Lymphoma, Cutaneous T-Cell, Lymphoma, T
- 4. **Leukemia, T-Cell, Acute, HTLV-I-Associated**
Mesh Reference: D015460
Synonyms: ATLL, ATLLs, Adult T-Cell Leukemia, Adult T-Cell Leukemia-Lymphoma, Adult T-Cell Leukemia-Lymphomas, Adult T-Cell Leukemias, HTLV Associated Leukemia
- 5. **Lymphoma, Lymphoblastic**
Mesh Reference: D016401
Synonyms: Convoluted T-Cell Lymphoma, Convoluted T-Cell Lymphomas, Diffuse Lymphoblastic Lymphosarcoma, Diffuse Lymphoblastic Lymphosarcomas, Large Cell

Figure 1: Select EZ Search and search results organized in categories listed in top left panel

Helpful Hint: EZ Search is also a good method to search the GeneGo database for pre-built canonical maps and networks. [Go live to see more.....](#)

- 2) Use the top left panel to view results in each category. Each category lists search results in the right hand EZ Search panel, where each item is a link to more information summarized according to gene, network object details etc. The table below lists some of the information found for each category of organization. All categories also include general information such as synonyms and molecular weight etc. (not listed below).

Table 1: Types of categories and details in GeneGo MetaDiscovery platform tables.

Detailed Page	Links to Public Domain Analysis Tools and Information	Links to GeneGo Analysis Tools and Information
Gene	Chromosome location, Species, Protein, Locus link, HUGO id, OMIM, Unigene, RefSeq id's, SNP-rs id's	Diseases, GeneGo Maps, Protein orthologs, Gene orthologs,
Protein	Enzyme id, GO molecular functions, GO processes, GeneGo Maps, References (Swiss Prot, OMIM, EMBL, PIR, PDB)	Gene, Orthologs, Build interactions network options, Protein complex details, tissue/cell species/expression
Network Object	GO Processes PubMed References	Corresponding database objects, GeneGo Maps, Targeting drugs, Diseases, Annotated interactions with links to PubMed
Disease	MESH, OMIM, ICD, MedlinePlus, eMedicine, Mutation DataBase, Wikipedia, SNPs, Genes, RNA, Proteins	Annotation tree, Export gene list options, SNPs, Genes, RNA, Proteins, Drugs, Disease Maps, GeneGo maps, Networks,
Drug	Kegg, PubChem, Mesh, DrugBank, ChEBI, PharmGKB, FDA, Drug References, Contradictions	Therapeutic disease indications, Drug-Drug interactions, targets (to network objects), option to build target network, Biological activity assays, ADMET, mechanism of action

- 3) For the purpose of this exercise we will export all the biomarkers for T-ALL. From the search panel select the “disease” category and check the box next to “Lymphoma, T-cell” (a). Additional options will appear in the bottom left panel.
- 4) Choose the option to ”save/export” (b) , name the file in the export box as “T-cell lymphoma GeneGo biomarkers” and select the “Gene” option under the list category (c) and press “export” . **The file will be saved to the data manger interface under the “Gene Lists” folder.** Note you can also export to Excel.

EZ Search

Name ☐ Exact match

Objects Found

- Genes (39)
- Proteins (37)
- Drugs (80)
 - Small Molecule Drugs (57)
- Biologics (23)
- Network Objects (20)
- Diseases (5)**
- GeneGo Maps (23)
- Articles (196)

Selected Diseases

-
-
-
-

1 ☒ **Lymphoma, T-Cell** b

Result pages: 1 (Showing results 1 to 5 of 5)

[Show tree with all Diseases](#)

- ☒ **Lymphoma, T-Cell** a
Mesh Reference: D016399
Synonyms: Lymphoma, T Cell, Lymphoma, T-Cell, Lymphomas, T-Cell, T Cell Lymphoma
- ☐ **Lymphoma, T-Cell, Peripheral**
Mesh Reference: D016411
Synonyms: Lymphoma, Peripheral T-Cell, Lymphoma, T Cell, Peripheral, Lymphoma
- ☐ **Lymphoma, T-Cell, Cutaneous**
Mesh Reference: D016410
Synonyms: Cutaneous T Cell Lymphoma, Cutaneous T-Cell Lymphoma, Cutaneous
- ☐ **Leukemia-Lymphoma, T-Cell, Acute, HTLV-I-Associated**
Mesh Reference: D015460
Synonyms: ATLL, ATLLs, Adult T-Cell Leukemia, Adult T-Cell Leukemia-Lymphoma,
- ☐ **Lymphoma, Lymphoblastic**
Mesh Reference: D016401
Synonyms: Convoluted T-Cell Lymphoma, Convoluted T-Cell Lymphomas, Diffuse L

Export c

List name

Export as ☒ Genes ☐ Network objects

Lists

Species ☒ Human ☐ Mouse ☐ Rat

Export to ☒ List ☐ Excel

☐ Resolver ☐ GeneSpring GX ☐ DecisionSite

Figure 2: Select items of interest in genes, proteins or network objects for further analysis options listed in bottom left panel such as export

Analytical Observations: Using EZ Search, we determined that T-cell lymphoma had 5 disease category matches, 39 genes, 37 proteins, 80 drugs, 23 pre-built GeneGo maps and 20 network objects.

EZ DATA MANAGEMENT

Applications of the Data Manager

Unlimited Data storage

EZ organization – create, move, share data and folders

Data activating system to allow analysis of multiple data sets and types

The Data Manager portion of the interface is used to organize, share and store your data and any outputs generated during your analysis. Since the Data Manager is in a multi-user environment, there is a separate workplace for every username. Any uploaded data will be stored in an “Experiment Folder” and items generated by MetaCore™ analysis, such as networks, netshots, maps or gene lists will be stored in their respective folders. There is no limit to the amount of data that can be uploaded. Additionally, you can create your own folders, move, rename, delete and copy data between folders.

File Types

There are several file types that you can work with in the Data Manager. For instance, when experimental data is uploaded using gene or protein identifiers, the general parser is used and each column of data is given a GX suffix. Each type of file has different functionality listed in the table below.

Table 2: Summary of data parsers type and functionality

Extension	List Description	List Functions
NL	Network objects list	Build Network, Look up drugs, Export, Share
GL	Gene list	Build Network, Look up drugs, Export, Add to Experiments, Share
IL	Experimental data uploaded via interactions parser	Build Network with or without GX files, Share
GX	Experimental data uploaded via general parser	ALL ANALYSIS OPTIONS, Export as Gene List, Share
MX	Experimental data uploaded via metabolic parser	Build Network, Enrichment analysis, Share

Helpful Hint 1: While all analysis options only exist with GX files, you can convert a GL file to a GX file:
right-click on the GL file → add to experiments option and create a GX file

Helpful Hint 2: While there are some export and intersection options with a GX file, to have a full range of export and intersection capabilities convert your GX file to a GL file:



right-click on the GX file → Export genes option to create a GL file

Activating Data

In addition to data management, the Data Manager harbors the ability to direct the type of data files you wish to analyze. Analysis functions are performed on experiments or structures located in the “Active Data” folder. You can delete items from the active folder and the original experiments will still be retained in the “experiments” folder. **Details on how to activate data will be described in the upcoming exercises.**

Sharing Data

With the Data Manager, you can share experiments, gene lists, and saved networks with other users or groups. To access the Share tool, right click on file and select “Share”. Enter a username in the search field and once found, click “Add selected to list” then choose a share level. For more information on the different levels of sharing refer to the help section on *portal.genego.com*.

Experiments you own, that are shared with other users, will be marked with a blue hand icon in the Data Manager (). Experiments owned by another user, that are shared with you, can be found in the Shared Data folder and will be marked with a green hand icon ().

UPLOADING AND ACTIVATING DATA

If you are using MetaCore to analyze high-throughput data, there are a series of pre-analysis steps you may want to consider:

- Statistical analysis to eliminate outliers
- Eliminating background noise
- Data normalization
- Data clustering and/or other statistical analyses

The software can be integrated with a variety of microarray data processing tools including Array Studio, Rosetta, GeneData, GeneSpring and Spotfire. For information on how to integrate these tools refer to the Integrations from the Help section in MetaCore™.

To begin an analysis, a list of identifiers is all you need. Types of data that can be used include a variety of high-throughput data including arrays of miRNA, genes, proteins, metabolites, chemical structures and new interaction data. You can also use a list of your favorite miRNAs, genes, proteins etc. without experimental values.

Objectives: Upload given data, activate, check data properties signal distribution and learn how to change threshold settings

Uploading data

- 1) In an internet browser, open <https://portal.genego.com> and enter username and password
- 2) A window will open that has 2 sections:
 - a. DATA MANAGER WINDOW on the left and
 - b. An analysis guide/EZ START WINDOW on the right.
- 3) To begin uploading data select the “Upload Data Tab” from the EZ Start window (a). You will be directed to the first step of the Data Analysis Wizard.

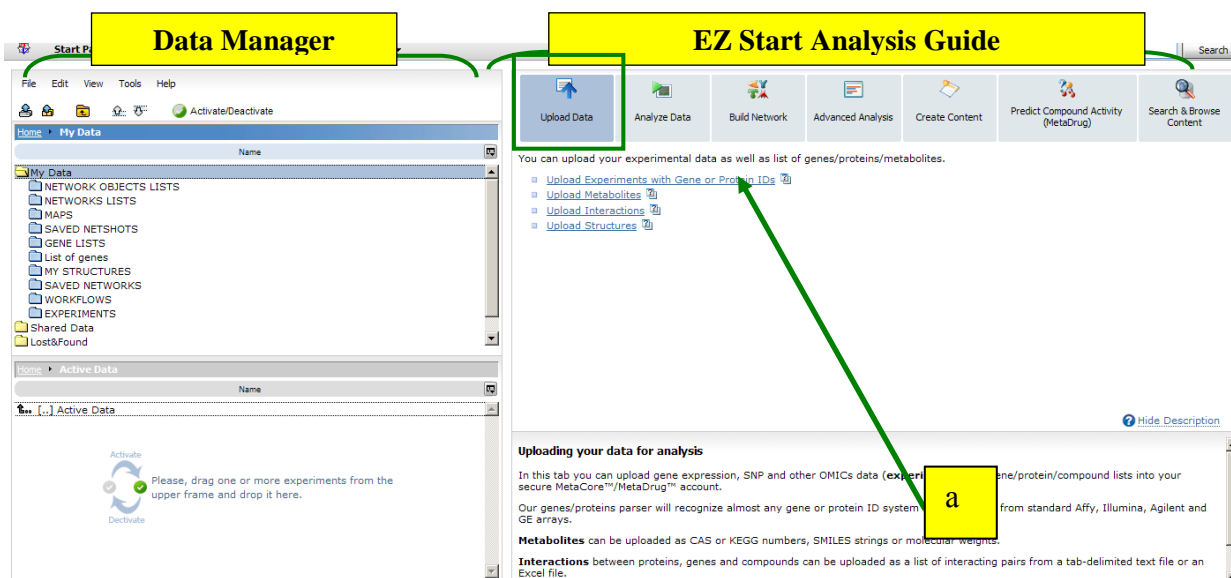


Figure 3: a. Select “Upload experiments with gene or protein ID”

Helpful Hint 1: It is a good idea to preview your data file (text or Excel) to ensure the Data Wizard correctly parses the file. Also, if you have p-values, the text/Excel data file should be organized so that the column of p-values is after the corresponding column of intensity the p-value is associated with.

Helpful Hint 2: The first interface of the Data Wizard lists all possible gene identifiers that will be recognizable. A complete list is also accessible from the Help section. You only need one column of gene identifier and it is important not to mix gene identifiers in one column. **If you have several columns of identifiers, use only one and choose to “ignore” the rest (see below).**

- 4) Browse for your file, in this case an Excel file called “**Gamma secretase inhibition**”. The Data Analysis Wizard will auto-parse the columns of information from your file and will detect gene identifier columns or columns with a numerical value as “intensity”.
- 5) The second step of the Data Analysis Wizard offers the opportunity to verify and/or change each column assignment. The Wizard should correctly identify the column of gene IDs and “intensities” from your data file. For this exercise use the drop down menus to change the following (a):
 - Second column to Fold change
 - Regulation columns to ignore
 - All p-value columns to p-value
 - Gene symbol column to ignore

Data Analysis Wizard (General parser)

Step 2

Only first 10 lines of your file are shown. Use horizontal scrolling if needed.

Use checkboxes against each row to specify table header lines

Specify the column types in your file:

Type	Name	Intensity	P-value	Intensity	P-value	Gene Symbol
<input checked="" type="checkbox"/>	Probe Set ID					
<input type="checkbox"/>	200001_at					
<input type="checkbox"/>	200083_at					
<input type="checkbox"/>	200605_s_at					
<input type="checkbox"/>	200608_s_at					
<input type="checkbox"/>	200618_at					
<input type="checkbox"/>	200627_at					
<input type="checkbox"/>	200655_s_at					
<input type="checkbox"/>	200681_at					
<input type="checkbox"/>	200720_s_at					

Figure 4: a. Select downward arrow to change column assignment

- When the assigned columns are correct, select “next” then choose human for species followed by clicking “next” again. A “Background processing status” window will appear. Once the upload is complete, the status will read:”100% Complete” and “FINISHED”.

Activating data

- At this point, the data should automatically be placed in a folder entitled “Gamma secretase inhibition” in the “experiments folder” in the data manger section and in the active window. Notice every column of data is assigned a GX suffix (p-values are wrapped into the respective data it is associated with).

Helpful Hint: To ensure your data upload is complete, refresh the DATA MANAGER window and search for your file in the Experiments folder.

NOTE: In the current example, the data automatically “activated”. This occurs when you upload new data.

- To activate pre-existing data from the Experiments folder, select your item of choice (folder or individual GX file) and select “Activate/Deactivate or double click and the item will appear in the active data window (a).

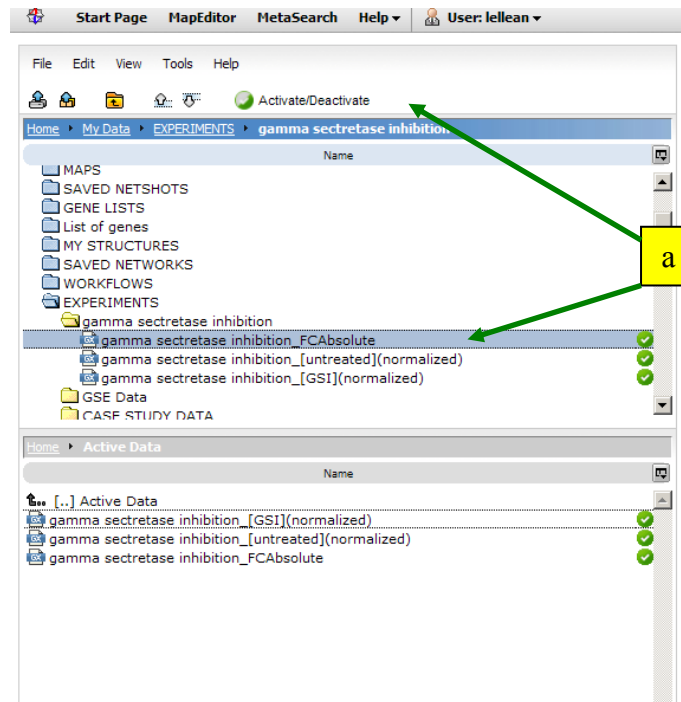


Figure 5: Highlight item to be activated and select “Activate Selected”

- 3) For this exercise **we will be analyzing according to the normalized and untreated files only**. If these files are not activated, select and drag to the active window. If the “FC absolute” file is in the active window select, right click and deactivate.

Verifying data properties

There are several useful steps to follow BEFORE you begin your analysis. One is to check your dataset properties.

- 1) To determine if your file is uploaded correctly, right click on the file called “**Gamma secretase inhibition FC absolute**” in the DATA MANAGER window. A new window with “Dataset Info” summary should appear. The summary will include a list of numbers in ratios. The first number of each ratio includes:
 - The number of IDs on file
 - The number of IDs recognized during the upload process
 - The number of IDs placed on different ontologies (categories) including Maps, disease and GO processes.

The second number of the ratio refers to the number of IDs that are used according to the current threshold settings. We will become more familiar with these options in the first exercise below.

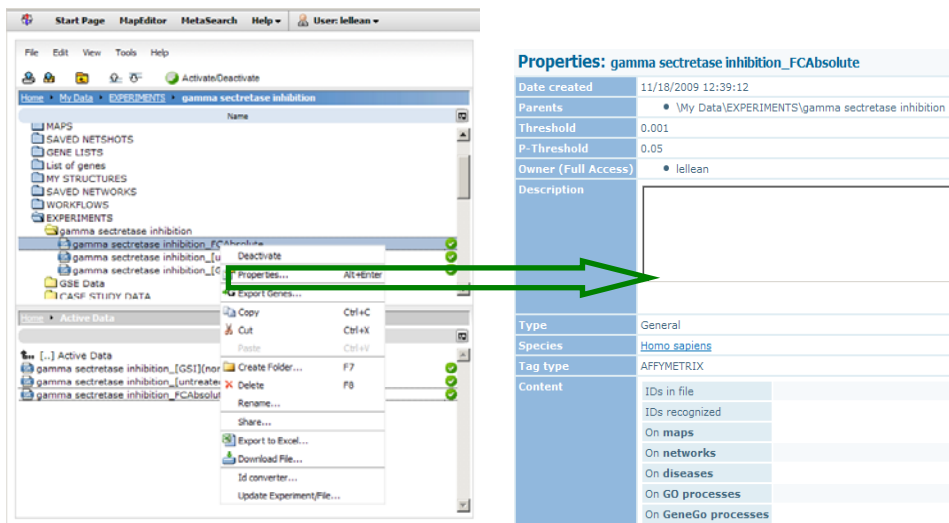


Figure 6

Determine expression/signal distribution:

Often, with a large amount of data, there can be a large range in expression of each gene ID between samples. With MetaCore™, there are the added abilities to analyze data in the negative and positive range, and to change thresholds to suit your analytic needs. It is always a good idea to check the signal distribution of the data set. If you upload a file with gene identifiers and expression values, the signal distribution will depict the range of expression values from your data set. This option can be used to determine the distribution of one of the GX files in the active window. To check the signal distribution of your data set:

- 1) Activate “ γ -secretase inhibition” and “normalized untreated” GSI files. **If the FC-absolute file is activated, select and deactivate.**
- 2) From the Data Manager tool bar, select **View** → “**Signal Distribution**” from the drop down window. For the current example, at a threshold of 0.001, with the “ γ -secretase inhibition” for untreated and GSI - GX files (Figure 7a).
- 3) Double click on the blue bar representing the number of genes that have a value between 0.001 and 0.5 to determine a better range of distribution (Figure 7b).

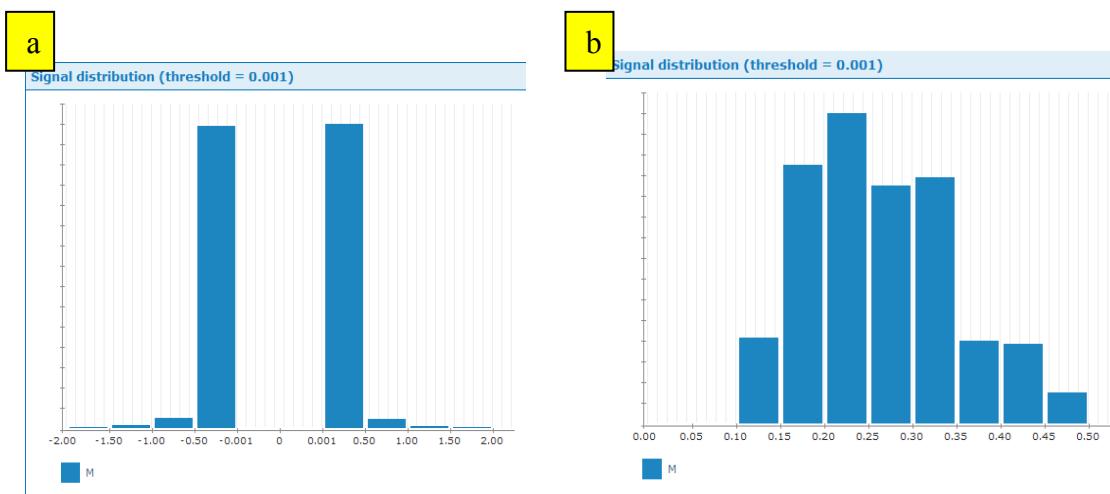


Figure 7

Analytical Observations: With reference to our data set, the signal distribution describes the following:

- At threshold intensity 0.001, the majority of the gene IDs from the data set have values falling within ± 0.001 and 0.5. Within this range we determine that the majority of the gene IDs have a value of 0.2 and 0.3. Since there is a significant change in the distribution at 0.25 (tallest bars compared to 0.15), this is a good indication of where to set thresholds.

Changing Thresholds

When analyzing data, make sure that the threshold is set appropriately to reflect the data values in your file. The threshold works as an absolute value; so, for example, if the threshold is set to 3, any genes with data values less than 3 and greater than -3 will be filtered out. The p-value threshold filters out any genes with a p-value equal to or less than, until it approaches zero. **The effect of the thresholds can be seen in the experiments properties window.**

- 1) To change the current threshold, select "Background List and Threshold" from the Tools menu (a).
- 2) For this data, we will set the threshold to 0.25 (b).

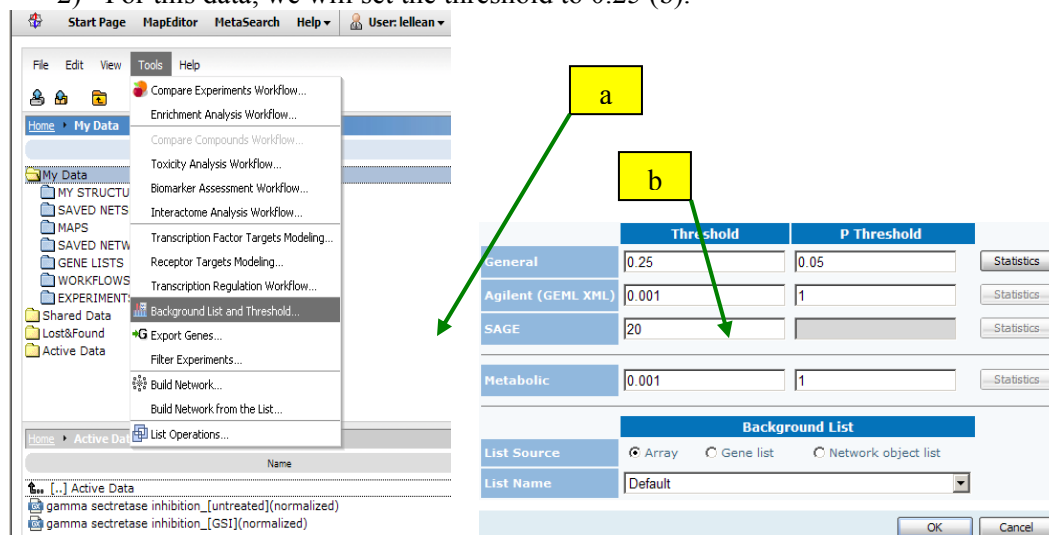


Figure 8: a. Experiment activated. b. Click here to set the thresholds.

Breakout Session I: Using the Data Manager interface and uploading data

Exercise 1: Uploading data

Follow the steps outlined above to upload the “**Gamma secretase inhibitor**” file located on the USB drive given to you with your training package. Once the file is uploaded, ***remember to refresh the data manager*** and locate the newly uploaded data in the experiments folder.

Exercise 2: Checking dataset properties

Once the file is uploaded, activate the “**Gamma secretase inhibitor FC absolute**” GX and right click to access the properties. Complete the following table (in Figure 9) with the information obtained from the dataset properties window, including content ratios.

Type															
Date created															
Species															
Tag type															
Parents															
Threshold															
P-Threshold															
Content	<table><tr><td>IDs in file</td><td></td></tr><tr><td>IDs recognized</td><td></td></tr><tr><td>On maps</td><td></td></tr><tr><td>On networks</td><td></td></tr><tr><td>On diseases</td><td></td></tr><tr><td>On GO processes</td><td></td></tr><tr><td>On GeneGo processes</td><td></td></tr></table>	IDs in file		IDs recognized		On maps		On networks		On diseases		On GO processes		On GeneGo processes	
IDs in file															
IDs recognized															
On maps															
On networks															
On diseases															
On GO processes															
On GeneGo processes															

Figure 9

Question 1: What do the ratios of the content list mean?

Exercise 3: Changing thresholds and checking dataset properties.

Close the dataset properties window. Go to the Data Manager menu, select “Tools”, and from the drop down window choose “Background List and Threshold” (as shown in Figure 8 above).

Change the threshold to 1, then close the threshold option window. Remember to **refresh the Data Manager Window**.

Right Click on “**Gamma secretase inhibitor FC absolute**”. Complete the following table with the information obtained from the data set window.

Question 2: How do the ratios change compared to those in Exercise 2? (Hint: reflective of changing threshold)?

ANALYSIS OVERVIEW

In general MetaCore™ users approach the tool with a list of genes, proteins, RNA or DNA. Throughout this course you will learn how to use MetaCore™ to generate a list from searching the database, enabling a user to approach the tool *even without any experimental data on hand*.

The most common question is “What are the pathways?”. Before we begin it is important to clarify GeneGo’s definition of pathways and what we define as identifying pathways. First you can identify pre-built pathways and create your own. When your question is:

“What are the most represented pathways?”

we understand this as

“What pre-built pathways are represented by the collective list of genes?”

and this is called a functional analysis using the enrichment ontologies (Figure 10). Using this function you can extend this question to what specific types of pathways are represented (biological function, disease, drug target etc.) depending on the ontology you use. Each ontology has a set of pre-built pathways in the form of a network or canonical map.

Alternatively, you can ask:

“How do the items on my list connect with each other and what pathway does this create?”

We see this as a connectivity question that can be addressed using the build Network tools that contain 10 different algorithms, giving you the utmost control and flexibility in creating specific interaction networks.

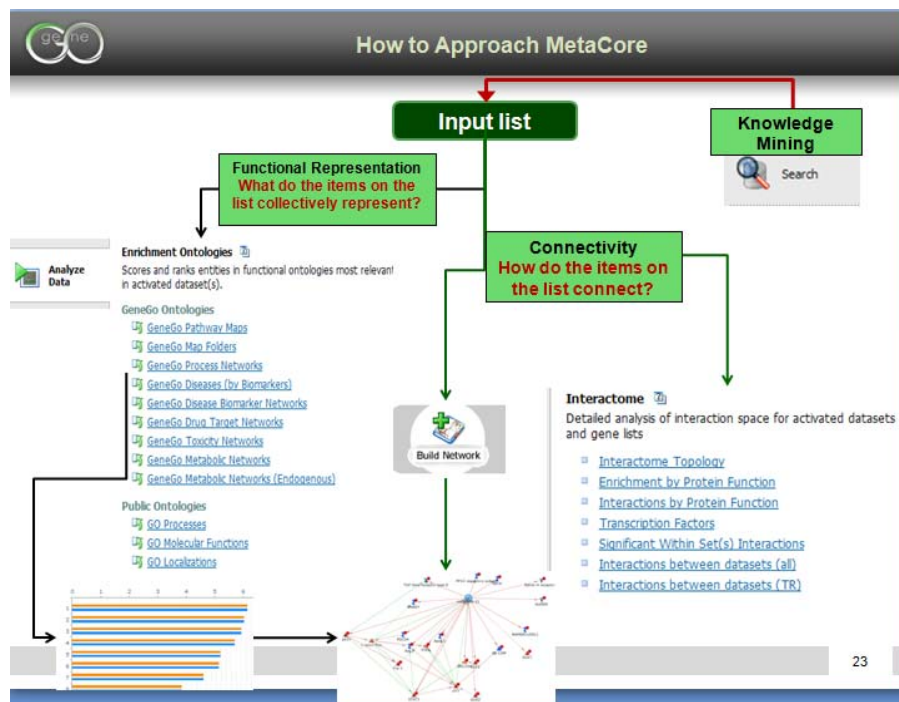


Figure 10

When you use the build network options, you are creating your own networks that have interactions with one or multi-step associations. This leads us to defining the difference between

pathway maps and networks. At GeneGo, a pathway map is equivalent to a canonical map (also called GeneGo Map) that has multiple sequential steps of interactions, defining a signaling mechanism. Each step is also well-defined and experientially validated and accepted by the field. They are similar to pathways you would see in a review article describing a particular function. A pathway in the form of a network can have a 1-step association that does not lead to further up or downstream events and multi-step interactions. The networks are descriptive of a biological function but contain more information than a map and may have newer results on them. You can think of a map as a snapshot from a section of a network. Figure 11 describes the differences in a tabular form.

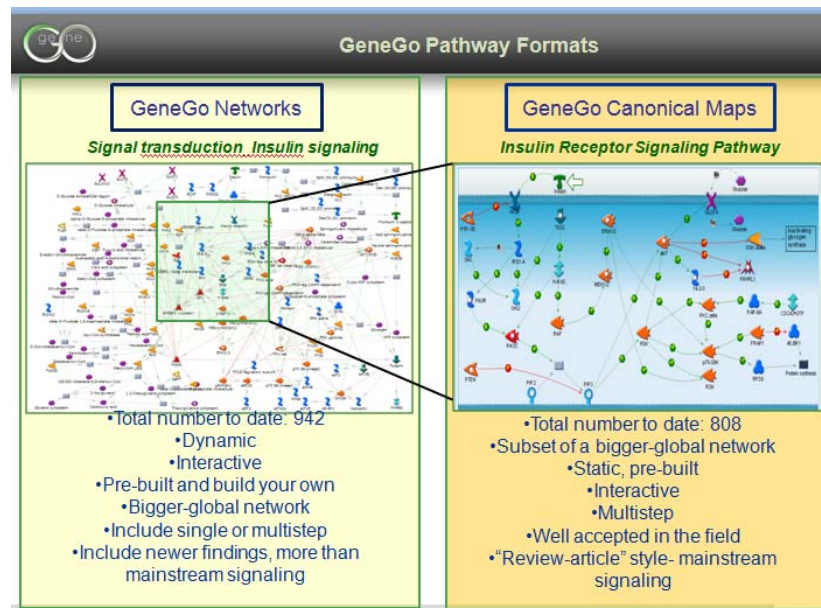


Figure 11

Every method of analysis provides 1) an interactive environment and 2) is accompanied by statistical measurements. Enrichment analyses are first displayed as histograms arranged with $-\log$ p-values organized by the most to least relevant. Each map or process network in every histogram links to an interactive visual representation of the enriched process in the form of a pre-built map or network, with the gene IDs from your original input list marked. Similar statistics are available for every network you build and are accessible/exportable from the "network statistics" button on a network page.

Helpful Hint: All enrichments and network-building statistics are calculated according to your gene ID and not the intensities. This implies that ***you do not need*** any values/intensities to do an analysis of a given list. Intensities are used to set thresholds (as described in the previous exercise) and to allow visualization of your gene IDs on maps and networks.

FUNCTIONAL ENRICHMENT ANALYSIS

Benefits

Ease of analysis – guided “3-click” tools
Ease of changing thresholds on the fly
Obtain analysis reports of enrichments AND networks

When to Use:

Determine most relevant biological functions represented by the data/gene list
Determine what diseases are represented by the data/gene list
Obtain quantitative evaluation of enrichments to facilitate prioritization
Visualize several datasets/types concurrently on canonical maps and networks

The MetaCore™ platform provides several mechanisms to report the relevance of a list of genes, proteins, metabolites or targets of uploaded compounds. The enrichment analysis option consists of mapping activated gene IDs (GX or MX files) of interest onto gene IDs grouped into ontologies. Enrichment analysis can be conducted for each ontology one at a time or collectively using a workflow. Enrichment analysis with MetaCore™ takes advantage of novel GeneGo ontologies or other publicly available ontologies (Gene Ontology www.geneontology.org). GeneGo ontologies include:

- canonical pathway maps
- cellular process networks
- disease biomarker networks
- drug target networks
- toxicity networks
- metabolic networks

GeneGo workflows are accessible from the analysis tab. They are designed to save you time in data analysis. Each is designed to combine several analysis steps under one umbrella. All workflows require data activation and a selection of parameters depending on the workflow. Each workflow is accompanied by a report summarizing the analysis. The workflow files are also saved under a workflow folder in the Data Manager. For a complete description of each workflow visit the help section of *portal.genego.com*. The different types of workflows include:

- Enrichment Analysis Workflow
- Analyze Single Experiment Workflow
- Compare Experiments Workflow
- Compare Compounds Workflow (only for MetaDrug™ software)
- Toxicity Workflow
- Biomarker Workflow

For this course we will be using the Enrichment Analysis workflow

Objectives: Using the Gamma secretase inhibition dataset, the goal is to determine the top most represented canonical maps, biological functions, and diseases.

- 1) In the Data Manager interface, ensure that you have active data files. In this example, you should have 2 GX files from the “**gamma secretase inhibition**” folder – specifically the normalized **untreated** and **GSI** files (a).
- 2) From the right, analysis guide side of the interface, select the “Analyze Data” (b) tab and choose “Enrichment Analysis” from Data Analysis Workflow option (c).

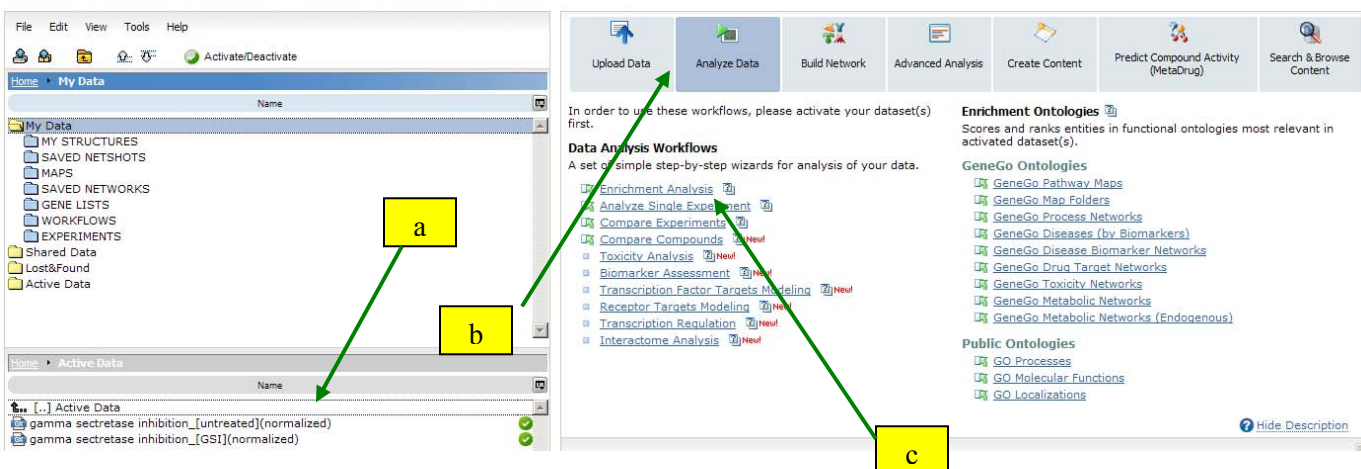


Figure 12: a. Active experiments. b. Go to the Analyze Data section. c. Click here to access the Analyze Single Experiment Workflow.

- 3) Next, select threshold settings (if you have not done so yet), and also choose what signals to use (up-regulated=”+”, down-regulated=”-“, or both), click Apply (**Figure 13a**).

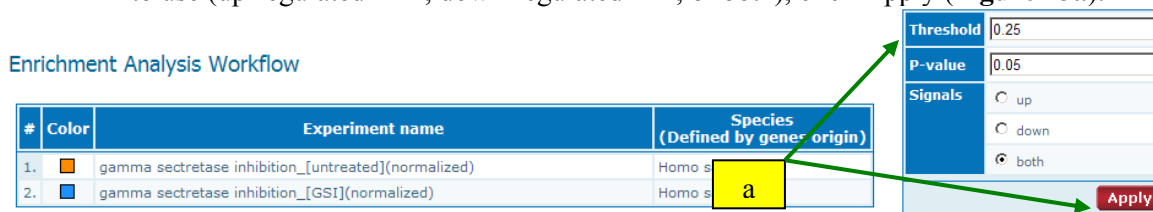


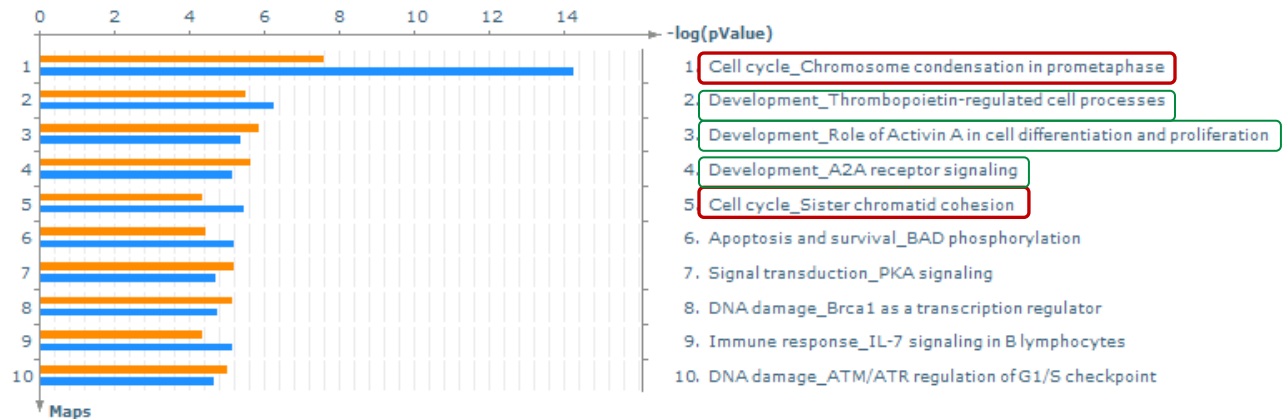
Figure 13: a. Set thresholds here and click Apply to start the comparison.

- 4) Wait for the results to calculate.
- 5) The histograms tabulate the distributions showing the top 10 most significant results according $-\log(p\text{-value})$ of a calculated hypergeometric statistic.

Below is a snapshot of the histograms for 2 GeneGo ontologies: 1) multi-step canonical GeneGo Maps and 2) GeneGo process networks that are most significantly enriched in our dataset.

GeneGo Pathway Maps

Distribution for: (1) gamma secretase inhibition_[untreated](normalized), (2) gamma secretase inhibition_[GSI](normalized)



GeneGo Process Networks

Distribution for: (1) gamma secretase inhibition_[untreated](normalized), (2) gamma secretase inhibition_[GSI](normalized)

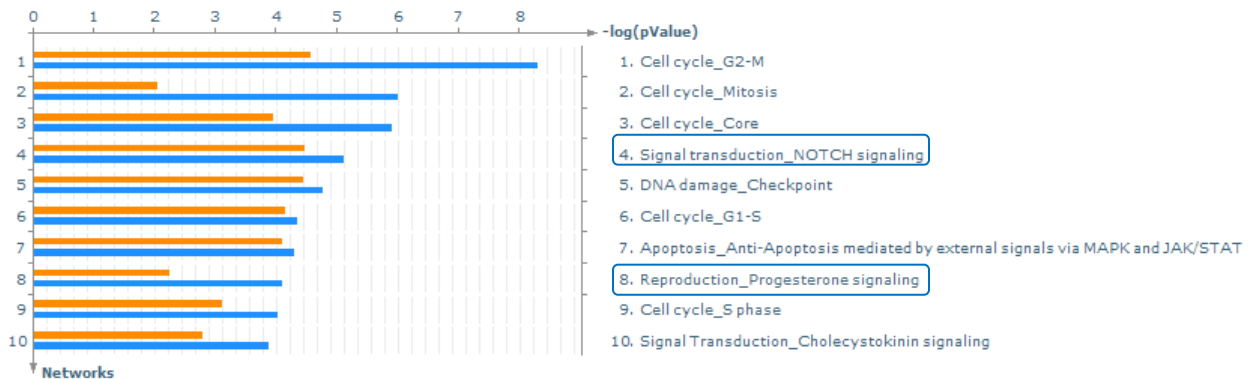


Figure 14

Analytical Observations: In this example, several functional insights can be made:

- Red Box:** The GeneGo Maps category lists cell cycle related maps in the top 3 results→ this suggests that inhibition of γ secretase caused changes in gene expression of several genes that collectively represent regulation cell cycle.
- Green Box:** GeneGo Maps enrichment also suggests the representation of developmental and cell cycle processes, consistent with the role of general notch signaling.
- Blue Box:** a closer look at the Biological Process further confirms a high representation of notch signaling along with other differentiation/development mediators. Cell-cycle processes are also appearing here, further strengthening our observations from the GeneGo Maps.
- You can also determine how representative these observations are for each data set. The longer the histogram bars, the more significant. Therefore in this case we can say that the first map or network is more represented by the T-cells treated with the GSI because the blue bar is longer than the orange one, which represents the untreated data file.

Helpful Hint: The histograms are interactive!

Mouse over the bar to determine the number of network objects or items from the gene list that are part of that particular map or network.

Clicking on the bar will allow you to open a pre-built map or network with the data overlaid.

- 6) Click on the first GeneGo Map: “Cell cycle_Chromosome condensation in prometaphase” to determine what genes are responsible for this enrichment (Figure 14).

Getting more from your enrichment results

At the top of the Map page:

- 1) Open the Visualization options panel and select your choice of filter to overlay. By selecting an option from one of the Marker categories, the objects associated with the selected category will be marked by a flashing icon.
- 2) Here we will mark the objects related to the lymphoma from the disease category (a) and all drug targets (b). ***These options are designed to facilitate driving your hypothesis further or confirming your current observations.***

This answers the questions:

What items on this map are biomarkers for general lymphoma (a)?

Which items have drugs that inhibit or activate them (b)?

Helpful Hint: The maps are interactive!

Click on the objects to: find all interactions (with details about up/downstream activation/inhibition, and mechanisms of interactions), diseases, drugs and GO processes

Click on the object to build a network of all interactions

Click on the interaction hexagon to see a window containing links to the PubMed articles where the interaction information was found and to export selected references to EndNote

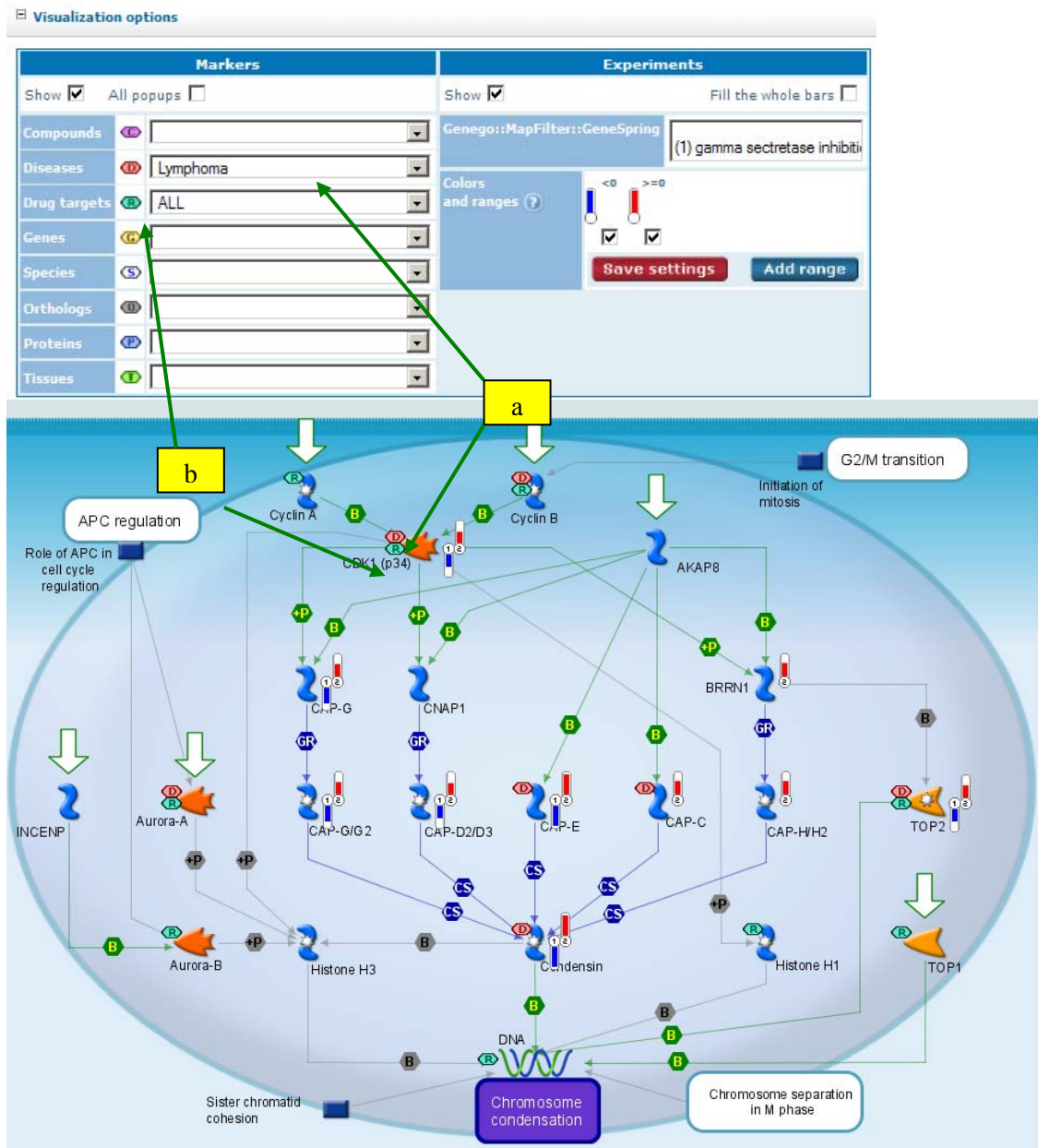


Figure 15: Data visualized as thermometer icons. a. Select visualization markers for disease of interest here. b. Select drug targetable objects here. Data values are represented by colored bars (thermometer-like icons) next to their associated genes. Blue = negative value and points downward, red = positive value and points up.

Analytical Observations: From this map the following observations can be made:

- The role of condensin and its upstream effects are dramatically altered by γ -secretase inhibition. For instance condensin, CAP G1/G2, CAP D1/D3 and CAPE all increase in expression (red bars labelled “2”) compared to the untreated sample (blue bars labelled “1”) that have a negative expression value.

- b) It is **possible** that these changes appeared to be mediated by CDK and TOP2 via BRNR1 not AKAP8 since AKAP8 expression remains unaltered.
- c) Interestingly, some genes from the data set are already markers of lymphoma (pink "D") such as CDK, CAP E, condensin and TOP2.
- d) If we consider that the cell line used was a T-cell lymphoma cell line, then in the untreated state, chromosome condensation is severely compromised, suggesting an oncogenic phenotype. By conducting the enrichment analysis and overlaying the data we can clearly suggest the treatment of these cells with a γ -secretase inhibitor reverses this oncogenic phenotype.

Helpful Hint 1: It is possible to export a list of genes on the map under the "Export options" header. Genes from the map can be exported as a list to be stored in the Data Manager, to Excel or to Array Studio, Rosetta Resolver, GeneSpring, Spotfire, and DecisionSite by selecting the desired tool in the "Export to" row. Exporting the map image itself is also possible in 300 dpi publication-quality png format.

- 7) Go back to your Enrichment Histograms and select the Signal transduction_NOTCH signaling network from the GeneGo Biological Process Networks. The pre-built network is another way GeneGo presents pathways.

Remember: *Networks differ from Canonical Pathway Maps in that they are more expanded pathways that describe several arms of a mediator and include interactions that may be experimentally proven once or twice. Therefore, they contain more information outside of the usual spectrum of a focused hypothesis.*

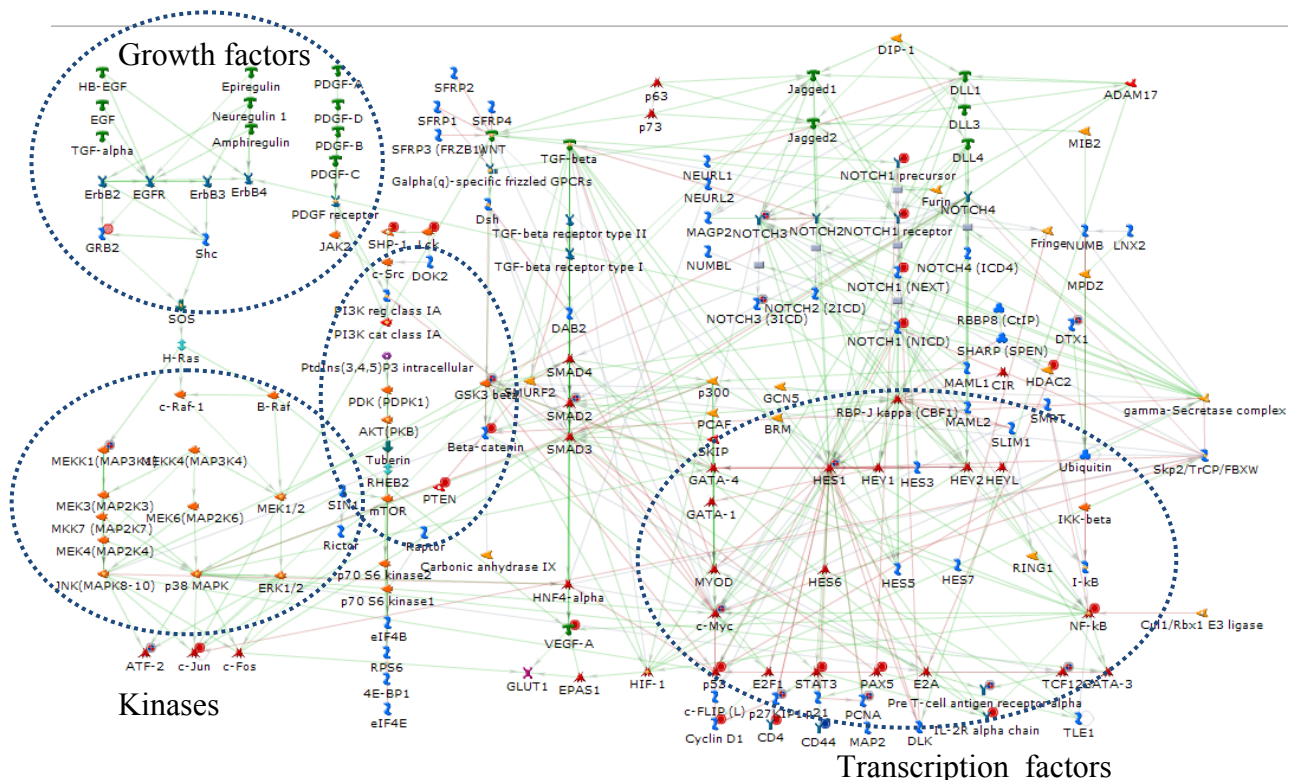


Figure 16

- 8) To determine the part of the Notch Pathway that is represented by the data only, focus on the objects with red/blue circles or open the EDIT tab → choose Selection → choose Show user data only.

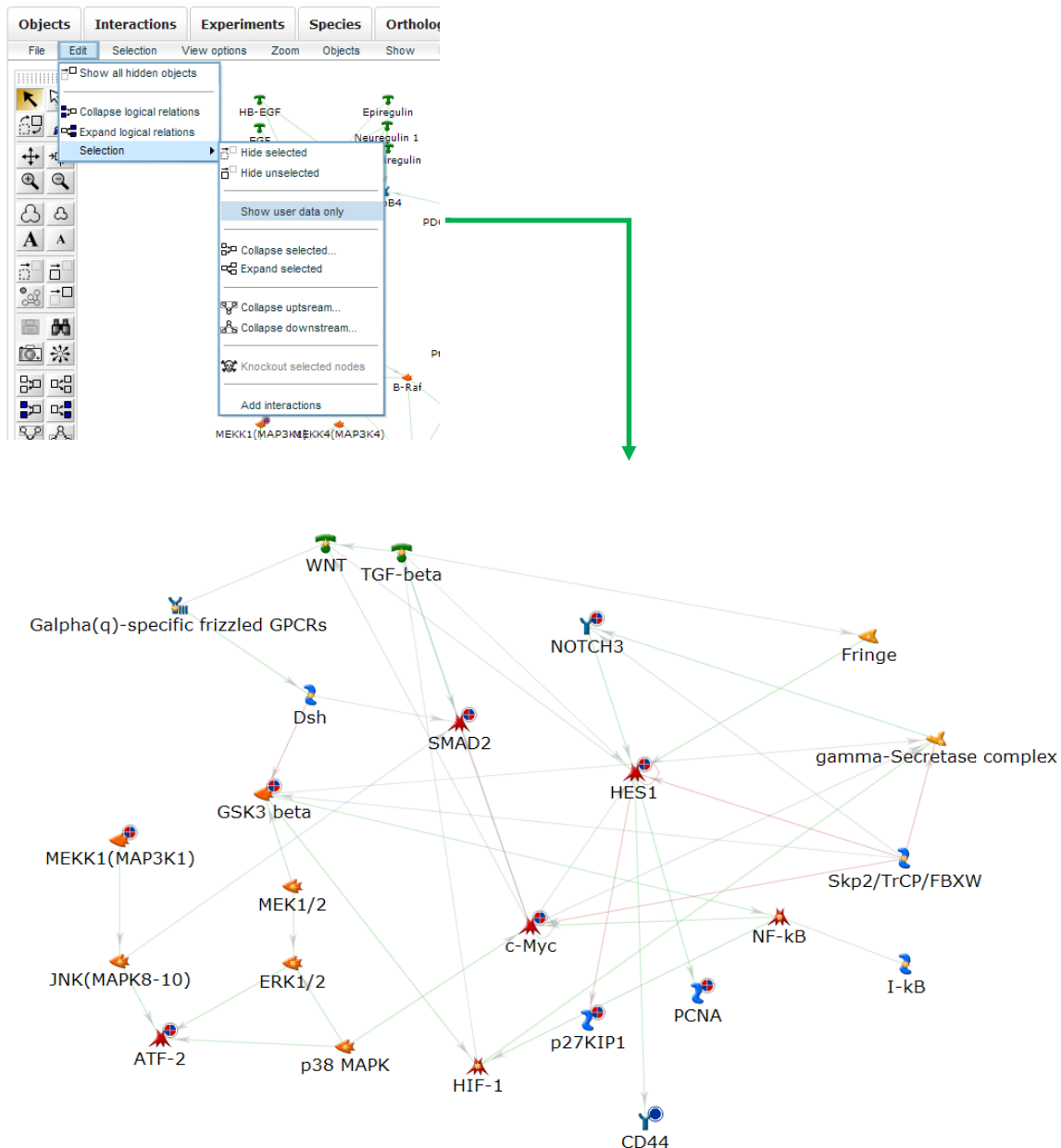


Figure 17

- 9) The objects that are expressed in the data are marked by red/blue circles in the upper right-hand corner of each object.
- 10) Unlike the maps, the data here is not represented side by side from each file. To determine which experiments the objects are expressed from, select the Experiments tab.
- 11) A window summarizing the data will appear. Select the data you wish to view or animate to cycle between all data sets.

Analytical Observations: From the pre-built network, we determined that Notch signaling incorporates several growth hormone (TGF, PDGF, HGF), kinase (MEKK, PI3K, Akt) and transcription factor (Nfk β and Hes) hubs of regulation (dotted circles Figure 16). By opting to show only the sections of the Notch network represented by the data we find that the most affected center of regulation is through the transcription factors HES1, SMAD2 and c-myc.

Breakout Session II: Determining the biological relevance of your gene list

Exercise 1: Getting Familiar with quantitative analysis using enrichment histograms

Next to each histogram:

- list your observation pertaining to each experiment (bar);
- list your observation pertaining to the 2 GX files collectively;
- list your overall impression of the biology represented.

Ontology	Observations per Experiment	Observations of all Experiments Collectively
GeneGo Pathway Maps		
GeneGo Biological Processes		
GeneGo Diseases (by Biomarkers)		
GeneGo Toxicity Networks		
GO Processes		

Exercise 2: Using threshold options to drive your hypothesis

Changing threshold settings allows you to narrow or expand the gene list that is analyzed or can highlight functions specific to a particular dataset. Repeat Exercise 1 at different thresholds. Record your observations and be prepared to discuss.

ASSESSING CONNECTIVITY USING NETWORK ALGORITHMS

Applications of Building Networks

- Do items in my list directly interact with each other?
- Are there any known interactions that connect genes from my list?
- How do I determine the biological relevance of a network?
- What processes and diseases does my network represent?
- What are the central hubs of regulation?

MetaCore™ is your ultimate tool for building networks using state-of-the-art algorithms and knowledge bases. Networks are built around network objects, including compounds, RNA or DNA, various protein classes and reactions. There are 10 different algorithms designed to build specific types of interaction networks. For more definitions about network objects and how each algorithm works, refer to the help section on *portal.genego.com*.

When you choose “Build network” you will be lead to the Network Options page where you can:

- select an algorithm
- edit the object list
- filter your list according to species, tissue, type of interaction, cellular localization and more

Helpful Hint: You can read about network-building algorithms in detail in the Network-Building Algorithms white paper or download recorded online WebEx seminars on how to use network algorithms, provided in the help section.

The screenshot shows the 'Network Options' page in MetaCore. At the top, there are tabs for 'Location' (new window, current window, parent window), 'Network type' (Metabolic, Regulatory, Merged), and a checkbox for 'Use canonical pathways' (checked). Below this is a section for 'Pre-filters' and 'Standard options'. The 'Standard options' section contains several algorithm options: 'Analyze network' (selected), 'Transcription regulation', 'Analyze network (transcription factors)', 'Analyze network (receptors)', 'Shortest paths', 'Self regulation', 'Direct interactions', 'Expand by one interaction', and 'Auto expand'. At the bottom, there are checkboxes for 'Advanced options' and 'Network objects'. Three green arrows point from text boxes to specific parts of the interface: one to the 'Pre-filters' section, one to the 'Network objects' checkbox, and one to the 'Advanced options' checkbox.

Open this tab to pre-filter for tissue type, species, and cellular localization, ortholog, object and interaction types.

Use Network objects section to select direction of signaling and add more objects.

Use Advanced options to further filter for types of interactions, reactions and nodes as well as number of signaling steps.

Figure 18

To appreciate the use of having several network-building options and how to increase the biological space around selected interactions, the following exercise will start with algorithms with more direct interactions and transit through more expansive algorithms.

Objectives:

- To build direct networks with a selected list and compare to more expanded networks using different algorithms (i.e. answer: do any of my gene list items directly interact with each other?)
- To become familiar with network operations (saving, exporting, aligning and post filters)
- To compare results of a direct network to those of another algorithm such as the analyze network algorithm.

- 1) Start with the Build Network tab from the analysis portion of the interface (a), and select “Build network with experimental data” (b).

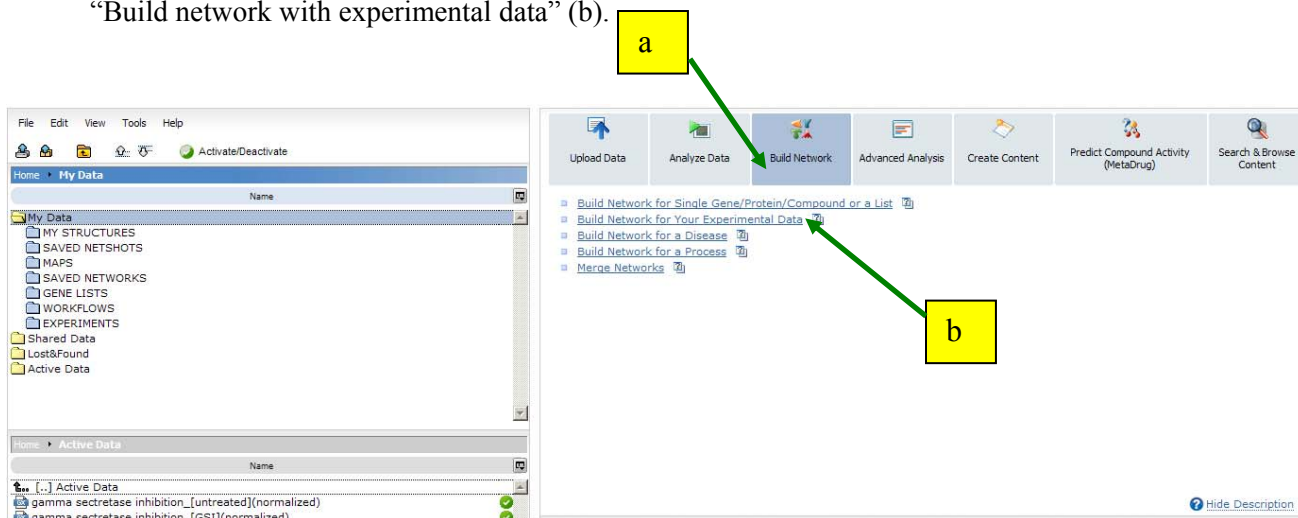


Figure 19: a. Select Build Network. b. Click here to start building from a list.

- 2) To determine if **any of the items on the data file connect directly to each other**, select the “direct interaction” from the Network Options page to obtain the following image:

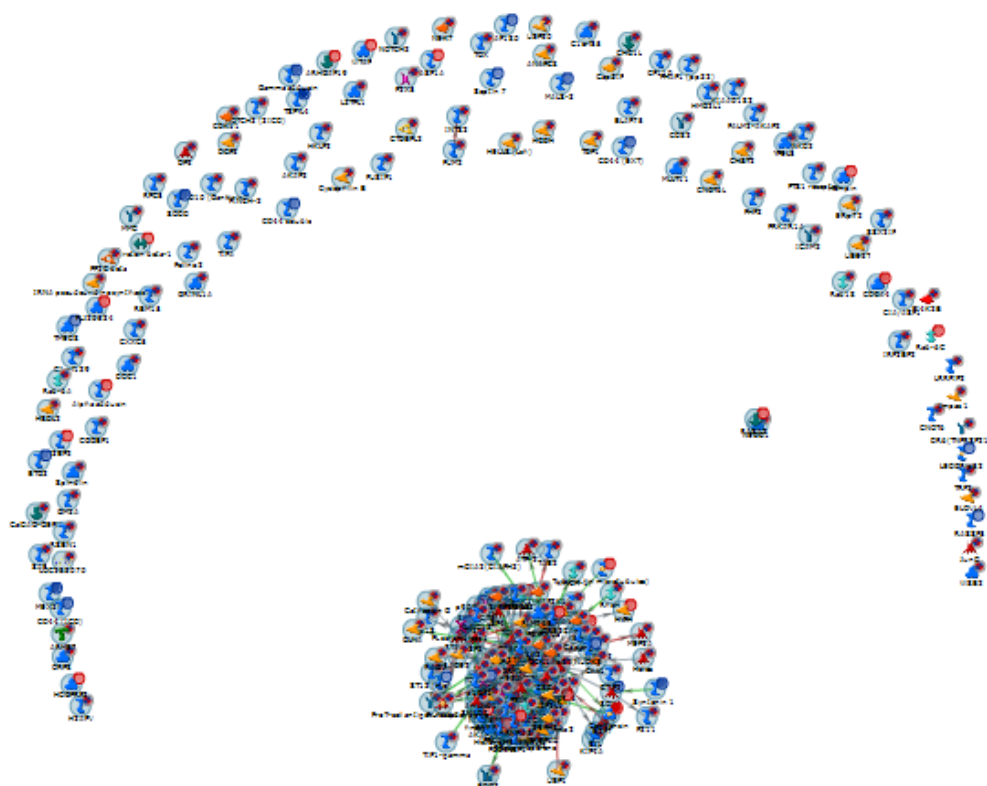


Figure 20

- 3) You will notice a cluster of interactions and several items that are not connected. To remove any unconnected items and take a closer look at the cluster, from the right hand

tool bar, select the connection icon  and hide all small clusters.

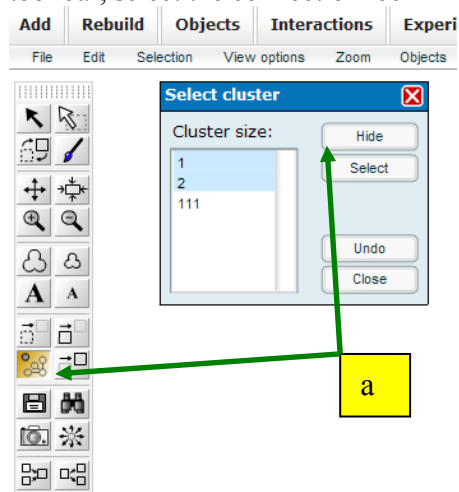


Figure 21

- 4) Next, to disperse the connected cluster pick one item (a) by drawing a box over it and drag it to one corner of the screen. Then from the top tool bar, choose “selection”→ select all (b). Find the align button on the right-hand tool bar (c), and align vertically and horizontally (d).

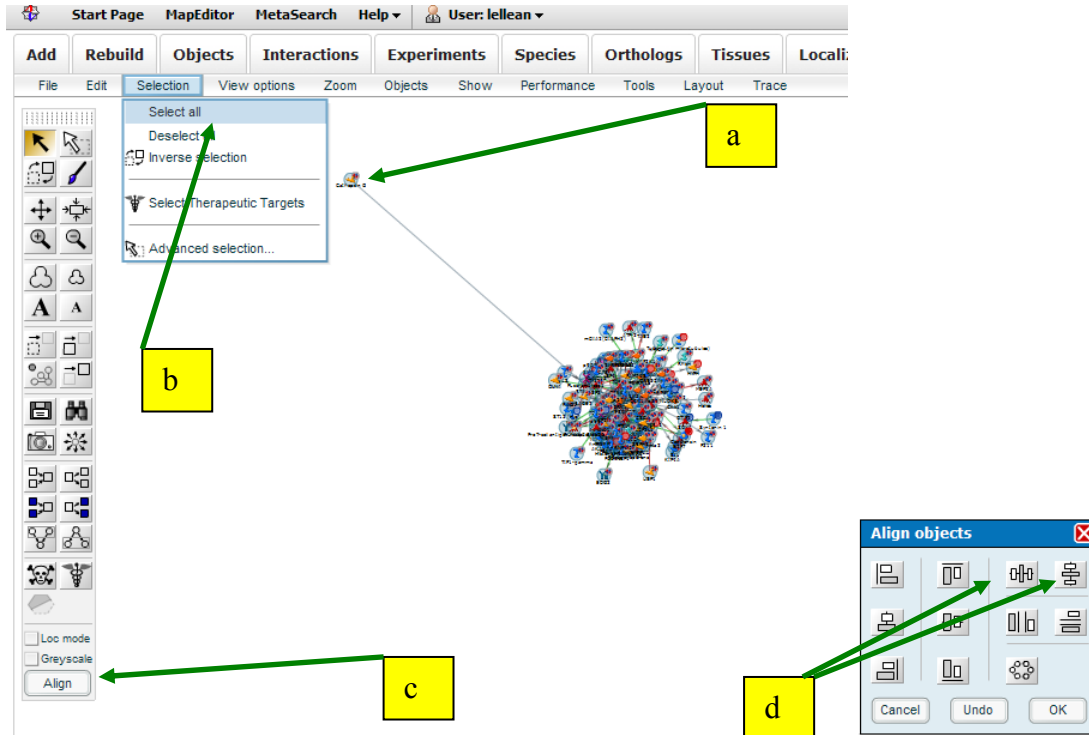


Figure 22: Working with a clustered network. To disperse the connected cluster pick one item by drawing a box around it and drag it to one corner of the screen (a). Then from the top tool bar, choose “selection”→ select all (b). Find the align button on the right hand tool bar (c), and align vertically and horizontally (d).

- 5) With the final image you can use the post filters (top tabs) to ask:
“What does the network mean biologically?” by selecting the GO Process tab.

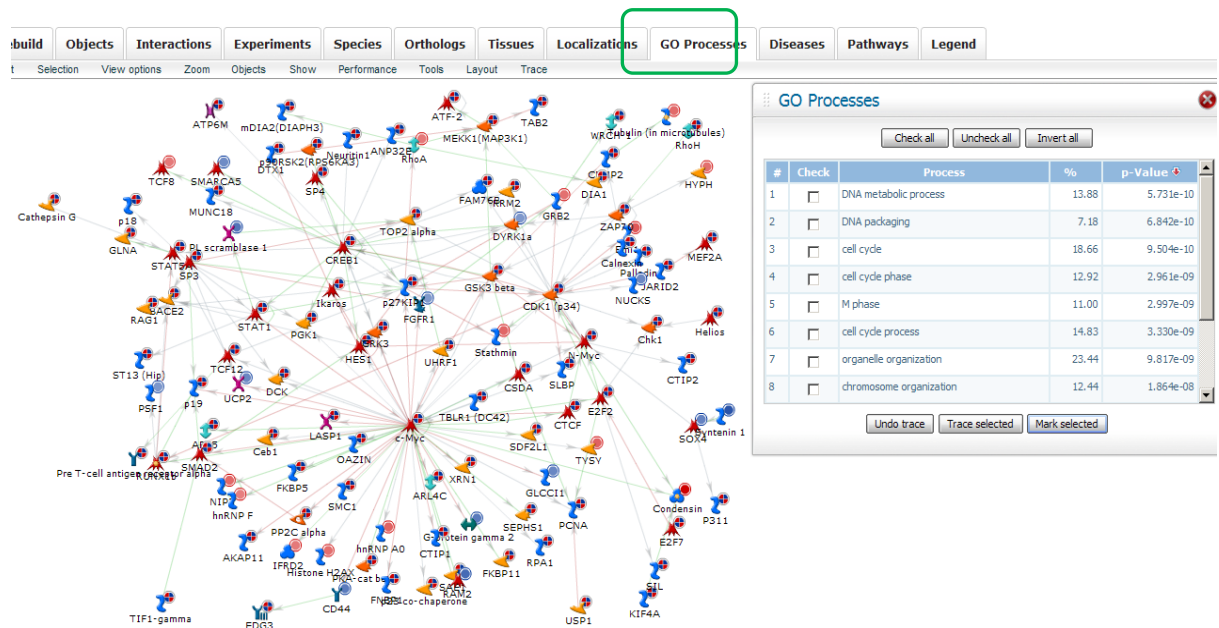


Figure 23

Analytical Observations:

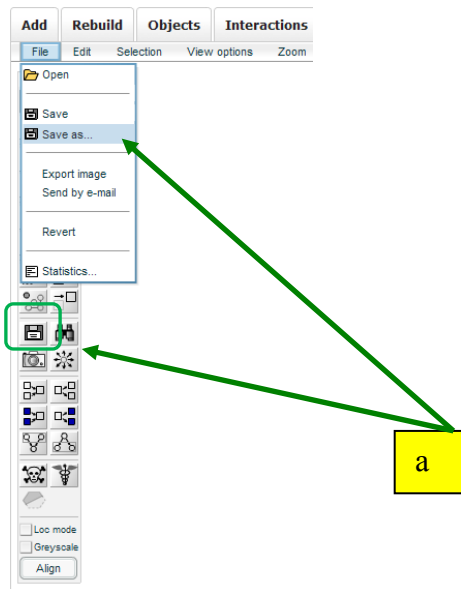
- In this exercise we specially wanted to know if any of the items from the data set are directly related/connected to each other. This option only uses the data in the data manager and does not bring in anything else from the database that could be an intermediate step. By doing so we identified that a subset of the data is connected to c-myc suggesting this transcription factor as a central hub.
- Using the GO Process filter we also determined that the items connected on this network collectively represent DNA metabolic process (highest score), DNA packaging other cell cycle functions.

Challenge question:

Is c-myc the most significant because it is well-connected within the dataset or because it is well-studied and an abundance of data is available in the public domain—GeneGo database? Hint to Answer: the interactome tool.

- Using the post filters to also ask:
 - What is the most represented disease?
 - What canonical pathways are represented?

- 7) To save this network use the options under “File” or the disk symbol from the left-hand tool bar.



The network will be saved in
“Saved Networks” folder in the
data manager

Figure 24

Next we will determine how all of the data is connecting, using an algorithm that will draw additional players from the database to facilitate processes.

- 8) Go to the “Network Options” window and select the Analyze network algorithm. This algorithm works by:
- Identifying the most connected items on the list;
 - Trying to connect to other items on the list using shortest path, and pulling in database objects according to most logical associations;
 - Creating sub-networks according to defined network space (default 50 objects);
 - Prioritizing according to hypergeometric distribution statistics and according to any interactions that are part of a known canonical pathway.

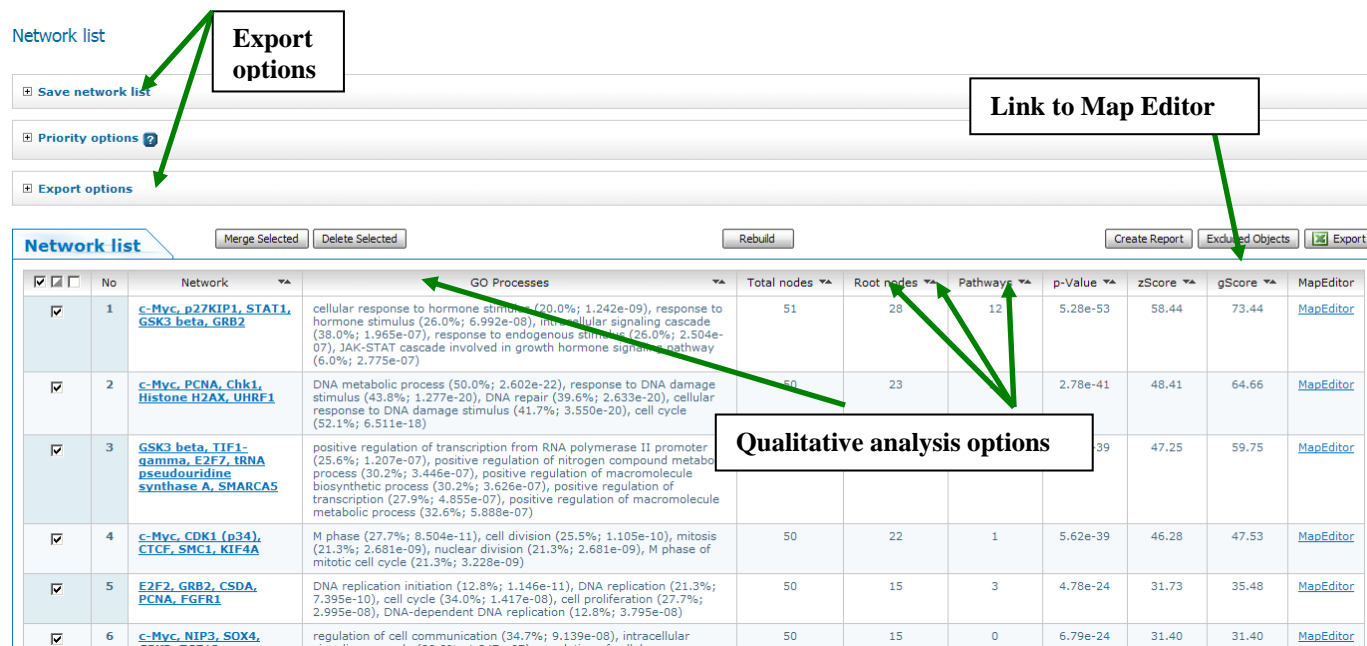


Figure 25

Analytic Observations: This algorithm also places c-myc as a central hub for several networks. By scrolling through the GO process column, the c-myc oriented sub-networks are consistent with our previous observations: the representation of cell cycle.

Helpful Hint: For more details on p-value, g-score and z-score calculations refer to the pdf document in the help section of MetaCore™. Use the scores to help prioritize your analysis steps.

- 9) For this exercise choose a sub-network that describes a process other than cell cycle. For example the MEKK1/TOP2 sub-network is centered on the role of phosphorylation. This network is also interesting because both MEKK and TOP2 were key players in our enrichment analysis for Notch and chromosome condensation signaling respectively. This sub-network may provide insight to a link between the 2 observations.

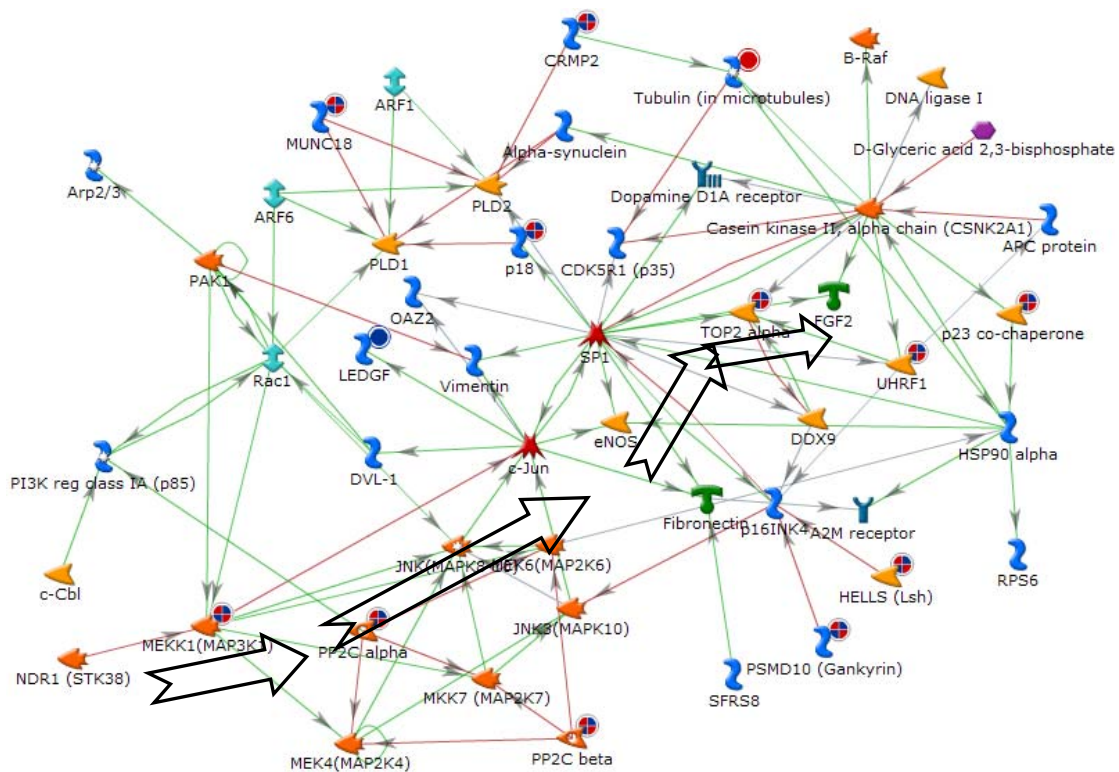


Figure 26

Analytic Observations: Indeed this network does provide a link between MEKK and TOP2 though c-Jun and sp1, downstream of NDR1 (highlighted above).

10) The image generated is a visual representation of connectivity. If you prefer a tabular view, scroll to above the network, and select “Network Statistics”. This is exportable to Excel and can help you determine the number of incoming and outgoing interactions for each object, the presence of different protein classes on the network and cellular localization as well as a summary of representative processes, toxic pathologies and diseases.

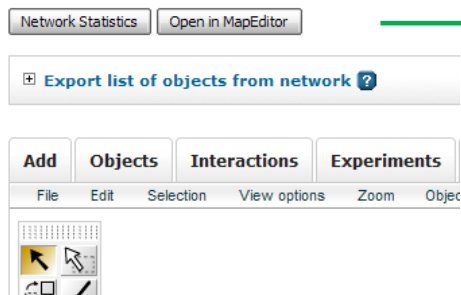
11) Repeat the saving steps above



Build new network

- [Select network objects / Build network from](#)

Title: MEKK1(MAP3K1), TOP2 alpha, UHRF1



Network statistics:

MetaCore™ version	6.0 build 21893
Name	MEKK1(MAP3K1), TOP2 alpha, UHRF1, PP2C alpha, Tubulin (in microtubules)
Description	
Total Nodes	
Total Edges	104
Low Trust Edges	1

- [Options](#)
 - [Advanced options](#)
- [Experiments](#)
 - [General parser](#)
 - [\(1\) gamma secretase inhibition \[untreated\]\(normalized\)](#)
 - [\(2\) gamma secretase inhibition \[GSI\]\(normalized\)](#)
 - [\(3\) T-cell lymphoma GeneGo biomarkers 12-11-2009 13:55:28](#)
- [Interactions](#)
- [Hubs](#)
- [Divergence hubs](#)
- [Convergence hubs](#)
- [Nodes](#)
- [Transcription factors](#)
- [Membrane receptors](#)
- [Secreted proteins & peptides](#)
- [GO processes](#)
- [Diseases](#)
- [Toxic Pathologies](#)
- [Tissues](#)

Figure 27

Break Out Group Session III

Exercise 1: Working with different network-building algorithms

Go back to the Network Option page and select a different algorithm.

- 1) Use the help section to summarize the role of the network algorithm you chose
- 2) Build a network or set of networks using your designated algorithm
- 3) Why did you choose this algorithm?
- 4) Discuss your results

Exercise 2: Working in pairs or groups

Use the network-building option Build Network for Single Gene/Protein/Compound or List and type in a list of up to 10 items when prompted.

- 1) Designate an algorithm to each member of the group, using direct interactions, auto expand, expand by one or shortest path (several members of the group can use shortest path but with a different number of steps).
- 2) Once the network is built use the select and align options and the left-hand toolbar to optimize visual effects.
- 3) Use the GO processes, disease and pathway filters to learn more about each others' networks.
- 4) Prepare to discuss.

ASSESSING CONNECTIVITY USING INTERACTOME ANALYSIS

When to Use

- What are the most connected items on my gene list within the database?
- What are the most connected items on my gene list within the dataset?
- What items are regulating (downstream interactions) and are being regulated (upstream interactions)?
- What is the percent of incoming and outgoing interactions within one dataset or between datasets?

The interactome tools allow the ability to analyze the categorical connectivity (density of interactions) of experimental data within the database. There are seven interactome tools and a description of each is found in the help section of the interface. The results are tabulated according to the statistically significant number of interactions in the set. It also allows analyzing interaction densities within and between several datasets, and in particular, estimating how one dataset affects another, such as the role of transcriptional regulation from one set on the other.

Objectives:

- To identify key regulatory protein classes and hubs within the dataset
 - To identify key regulatory protein classes and hubs not expressed in the dataset
 - To determine if the connectivity of a hub is true or a result of over annotation due to what is available in the literature.
- 1) Activate “**gamma secretase inhibition**” folder – specifically the normalized **untreated** and **GSI** files and access the interactome tool from the Advanced Analysis tab (a). Select the “Interactions by Protein function” (b) options.

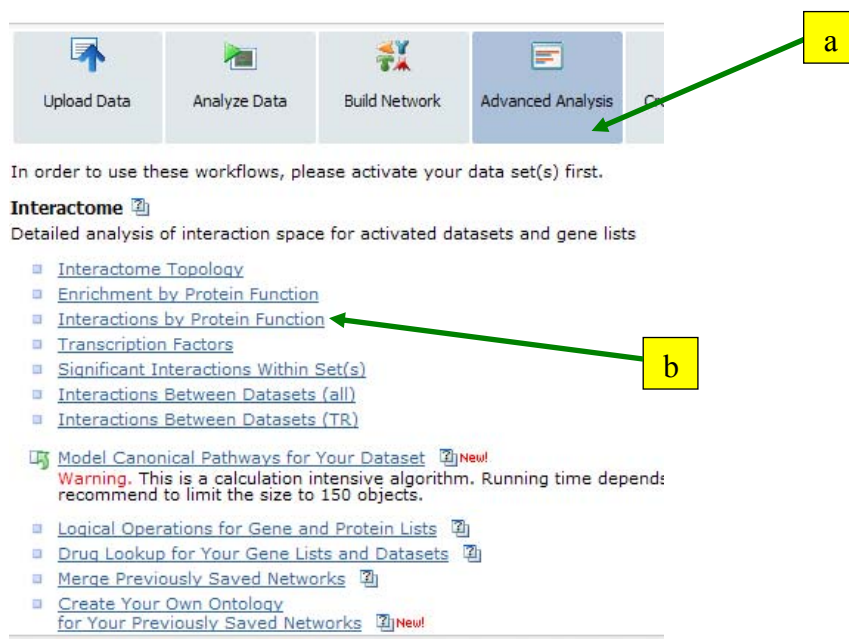


Figure 28: a. Select Advanced Analysis. b. Click here to open the Interactome tool.

- 2) The output of all interactome analyses are tables that are easy to read and exportable. The following is an image of the legend for the tables accessible from each results page.

Legend	
Columns have the following meaning:	
IDs in active data set	original probe/gene IDs in the activated dataset(s)
Object name	network object name in MetaBase
r	number of network objects in the activated dataset(s) which interact with the chosen object
n	number of network objects in the activated dataset(s)
R	number of network objects in the complete database which interact with the chosen object
N	total number of gene-based objects in the complete database
mean	mean value for hypergeometric distribution ($n \cdot R / N$)
z-score	z-score $((r - \text{mean}) / \sqrt{\text{variance}})$
p-value	probability to have the given value of r or higher (or lower for negative z-score)

Figure 29

- 3) The results table has two tabs: Over-connected and Under-connected (see image below for a visual representation of connectivity) with a score:
- If the number of interactions for the object is more than (or equal to) the mean (mathematical expectation), such objects are considered over-connected (their z-score ≥ 0).
 - If the number of interactions is less than the mean, the objects will be considered under-connected (their z-score < 0).


Overconnected	Underconnected	<input type="checkbox"/> Active data set only	<input checked="" type="checkbox"/> Show all columns	 Export					
IDs in active data set		Object name	r	n	R	N	mean	z-score	p-value
Transcription factors									
-1413709642; 202431_s_at	E2F1	58	405	558	21684	10.42	15.07	1.045e-26	
	c-Myc	108	405	1925	21684	35.95	12.71	3.568e-26	
596206521	E2F4	45	405	328	21684	6.126	15.98	7.26e-26	
	STAT3	51	405	476	21684	8.89	14.42	3.838e-24	
	c-Myb	35	405	249	21684	4.651	14.29	1.127e-20	
	c-Fos	39	405	342	21684	6.388	13.13	1.304e-19	
458018884	SP1	101	405	2100	21684	39.22	10.48	2.069e-19	
	c-Jun	51	405	645	21684	12.05	11.5	2.676e-18	
695458262	RelA (p65 NF-kB subunit)	50	405	633	21684	11.82	11.38	6.175e-18	
	NF-kB1 (p50)	31	405	242	21684	4.52	12.64	2.945e-17	
482975076	p53	74	405	1349	21684	25.2	10.14	3.877e-17	
1976466315; 210365_at	AML1 (RUNX1)	31	405	248	21684	4.632	12.44	5.979e-17	
200887_s_at	STAT1	40	405	462	21684	8.629	10.9	7.242e-16	
	NUR77	21	405	109	21684	2.036	13.45	1.007e-15	
	RBP-J kappa (CBF1)	19	405	85	21684	1.588	13.98	1.293e-15	
	Bcl-3	17	405	63	21684	1.177	14.75	1.356e-15	
	RUNX3	17	405	68	21684	1.27	14.11	5.524e-15	

Figure 30

- 4) Reading the table:
One way to evaluate the connectivity of each item within the dataset vs. the entire database is to calculate the r/n ratios. For instance:

For Myc:

Data: $r/n = 108/405 = 0.267$ and Database: $R/N = 1925/21684 = 0.09$

For E2F1

Data: $r/n = 58/405 = 0.14$ and Database: $R/N = 558/21684 = 0.025$

These calculations are taken into account when the p-value is determined.

Analytic Observations In this exercise, myc was also identified as a significant over-connected hub, since the number of interactions with the data is more than those with the database. This is consistent with our previous observations.

Using this method we are also able to say that E2F1, which is not expressed in the data, is a key regulator transcription factor.

- 5) Scroll through the rest of the list to determine key regulatory receptors, enzymes etc.

Make a list of the top most-connected items from each category that are also expressed in the data, and the top most-connected that are not expressed in the data. This will be needed in the next exercise.

Helpful Hint: The interactome by protein function results can also tell you about the types of pathways represented by your data. After scrolling through the table, can you relate any functions to some of the key hubs? Are there any reoccurring functions?

CREATING A CUSTOM MAP WITH MAPEDITOR™

Applications of MapEditor

Custom editing of pre-built GeneGo canonical pathway maps
Conversion of saved networks into map visualizations
Creation of new maps from maps from scratch
Addition of custom objects and links
Ability to “publish” to MetaCore™ to use for data overlay and enrichments

MapEditor™ is a Java module, which enables custom editing of canonical pathway maps available in MetaCore™, conversion of networks into map visualizations, and drawing of regulatory and metabolic maps from scratch. This tool is linked to the underlying database and enables a user to search from the MapEditor™ interface and simplifies adding objects to the map. A user may choose the objects for his/her maps from the MetaCore™ content of genes, proteins, compounds and interactions, *or introduce new objects*.

Computer requirement: Java 1.4.2 (except Java 1.4.2_06 and 1.4.2_07) is required on the client computer. If Java is not present it is automatically installed on the client computer when Map Editor™ is first accessed.

NOTE: a full Map Editor manual is available from the help section on *portal.genego.com*.

Objectives:

For this exercise we will use the interactome results to create a custom map and determine how the top-connected nodes interact with each other.

- 1) To access MapEditor™ click on the link from the data manger interface from the top tool bar or from the Create Content tab.

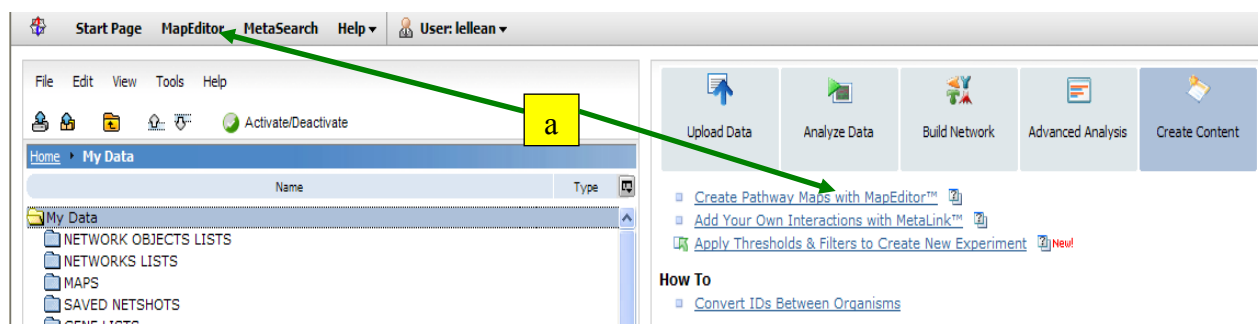


Figure 31

- 2) You will be prompted to connect to the MetaCore database. To connect, enter the same username and password used for MetaCore login.

Helpful Hint: You can also access MapEditor™ from the network interface

- 3) To open a new map, select File → new → Map

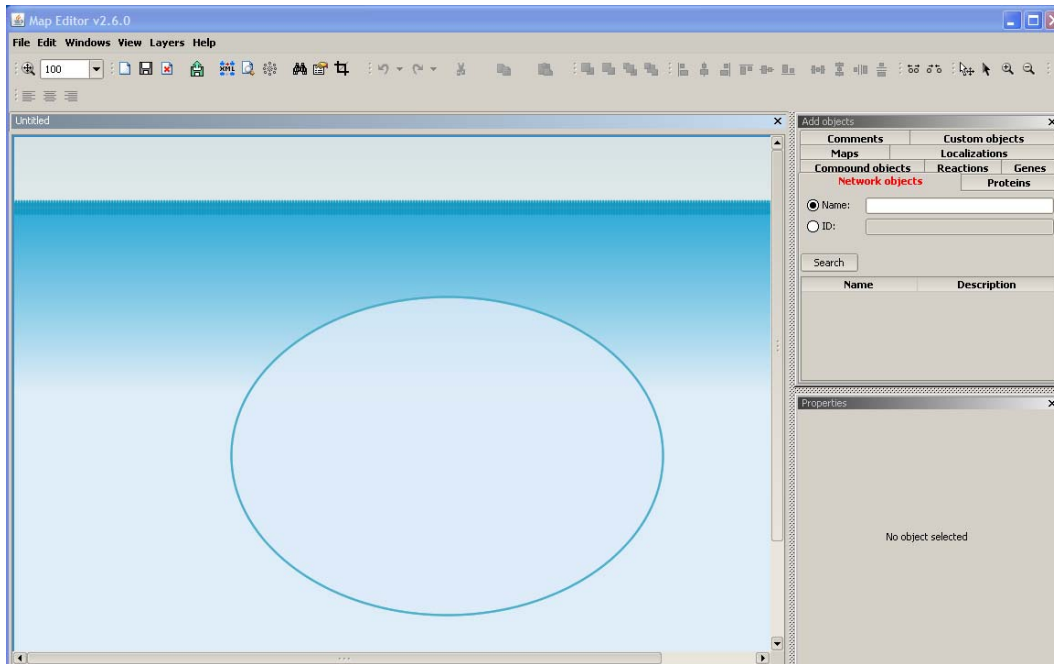


Figure 32

- 4) The new map instantly sets a plasma membrane and an additional intracellular compartment. Select the intracellular circle and make the space smaller
- 5) MapEditor is linked to the database and items can be searched and used for building a pathway here. Using the Network object tab search for the first item on the list from the interactome results:
- E2F4 – not expressed but most connected. From the results drag and drop E2F4 into the circle designating an intracellular space.
 - Next, search c-myc

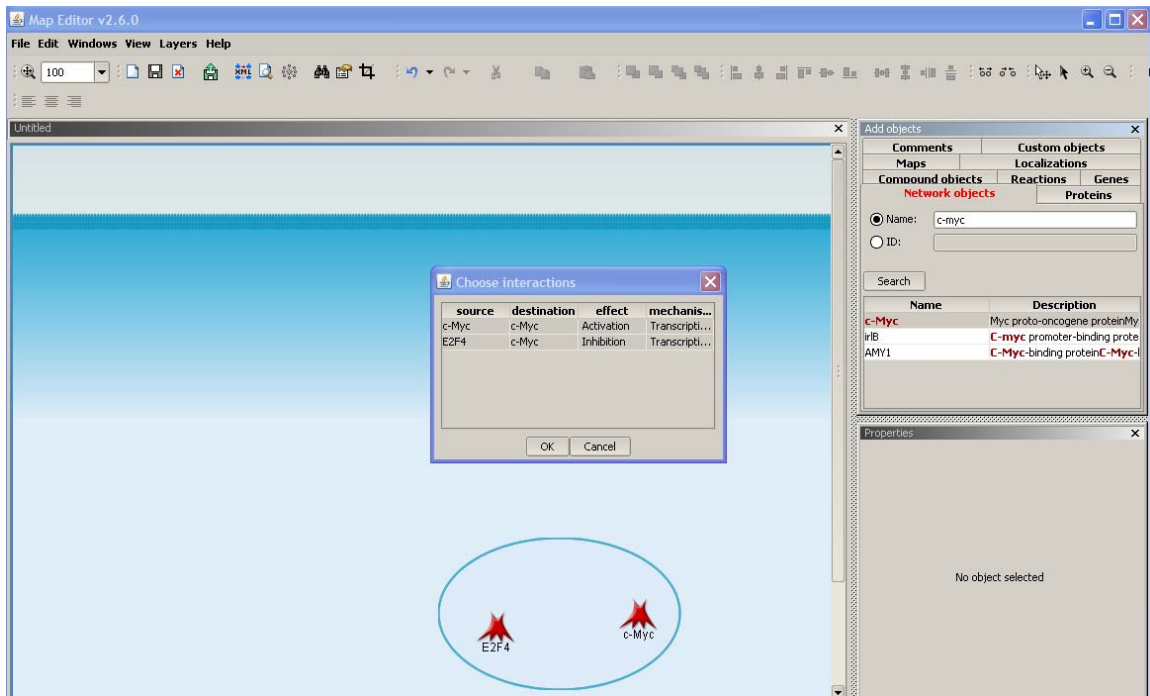


Figure 33

- 6) If any interactions in the database exist between c-myc and E2F4, an interaction box will appear for you to select desired interactions. In this case, accept both interactions.
- 7) Next search for the top-connected receptors from the interactome results (FCGRT and FGFR). Since these are receptors, you can drag and drop into the plasma membrane.
- 8) Continue searching and adding network objects and interactions. Note: you can also double-click on the arrows to change their size.

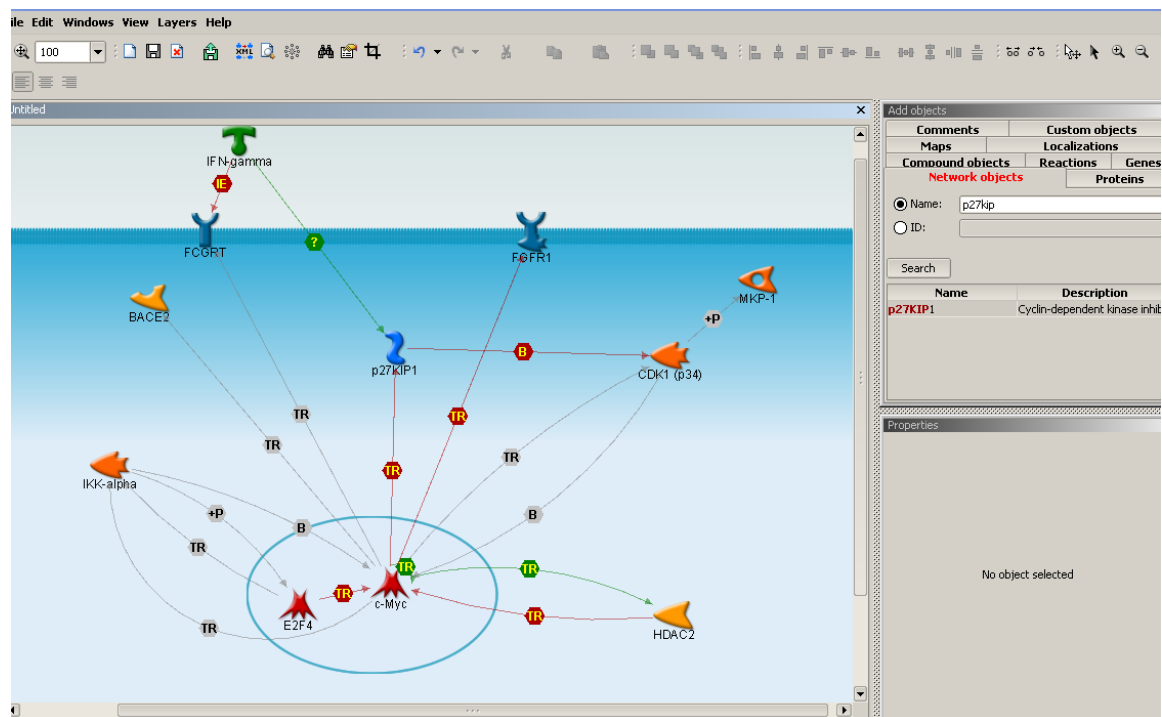


Figure 34

- 9) To add a custom item, select the Custom objects tab and drag the general protein network object symbol.
- 10) To change this symbol, double-click and use the properties window in the lower right corner. Use the Object information tab to change the icon by clicking on the icon to open the legend for more options.
- 11) In the same properties window you can also change the name of the object. In this example, an compound symbol was chosen and renamed to FGFR inhibitor.

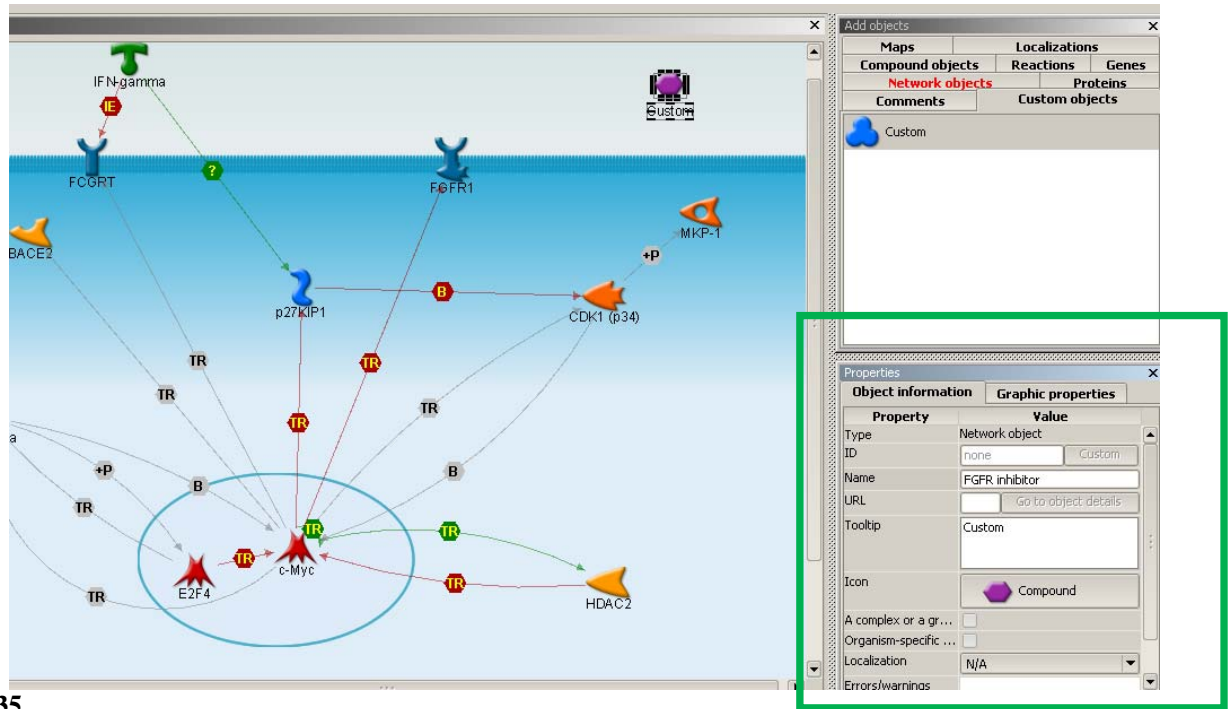



Figure 35

- 12) To create a link between the new object and the FGFR, select the link tool in the upper tool bar , then click on the inhibitor symbol followed by the FGFR symbol.

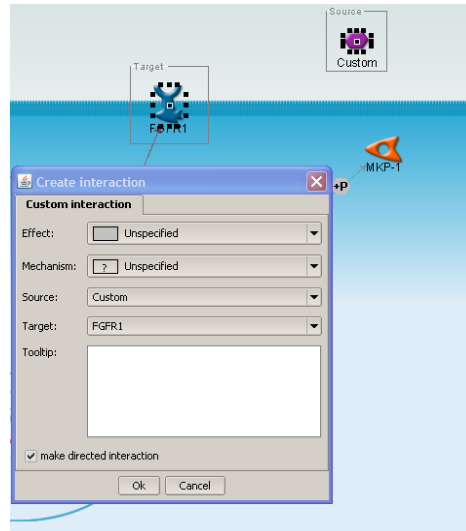



Figure 36

- 13) A custom interaction box will appear where you can select the color or the arrow (i.e. red for inhibition) and mechanism (i.e. binding).
- 14) Be sure to select the selection arrow before continuing. 
- 15) To save the custom map, you can use the menu from the File option. You can save as an xml file or publish it to the Map folder ontology. For this exercise, select the "Publish" option.

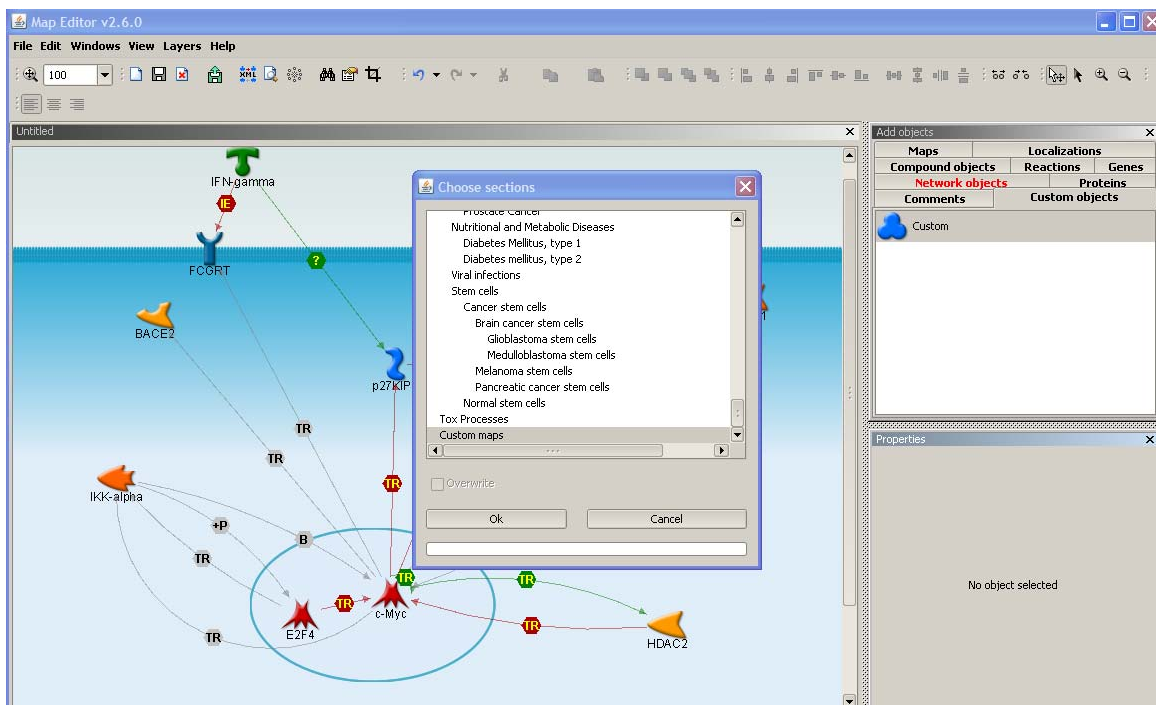


Figure 37

- 16) Once it is published it will be saved to your “Maps” folder in the data manager. Refresh the data manager then right-click on it and call it “GSI-interactome Map”.

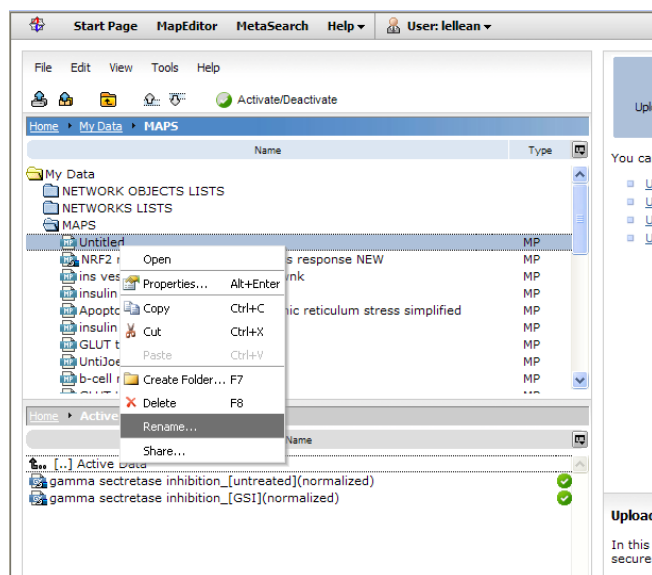


Figure 38

- 17) Make sure you have data in the active window. If you like, activate the biomarker list that was exported from the search exercise along with the GSI data. Double-click on the new map to overlay the data on the custom map.

References

1. Goal, G. (2006) *J. Biomolecular Techniques* 17, 252-260
2. Wolfe, M.S. (2006) *Biochemistry* 45, 7931-7939
3. Fortini, M. E. (2002) *Nat. Rev. Mol. Cell Biol* 3, 673-684
4. Zhao, B., Yu, M., Neitzel, M., Marugg, J., Jagodzinski, J., Lee, M., Hu, K., Schenk, D., Yednock, T., and Basi, G. (2008) *J. Bio Chem.* 283 2927-2937
5. Shih I, M., and Wang, T. L. Notch Signaling, γ -Secretase Inhibitors, and Cancer Therapy (2007) *Cancer Res.* 67, 1879-1882
6. Miele, L. Notch Signaling (2006) *Clin Cancer Res* 12(4) 1074-1078
7. Lanz, T.A., Himes, C.S., Pallante, G., et al. The γ -secretase inhibitor, DAPT, reduces A β levels in vivo in plasma and CSF in young (plaque-free) and aged (plaque-bearing) Tg2576 mice. (2003) *J Pharm. & Exp. Therap.* 10 1124