

# IPA Training: Maximizing the Biological Interpretation of Gene, Transcript & Protein Expression Data with IPA

# Log into IPA: www.ingenuity.com

The screenshot shows the Ingenuity website at <http://www.ingenuity.com/proc>. A red circle highlights the URL bar. Another red circle highlights the 'LOGIN' button in the top navigation bar. A blue arrow points from the 'NEW: INSTALL IPA CLIENT' button on the left to the 'Welcome! Please login' form on the right.

http://www.ingenuity.com/proc Ingenuity Products Login - ...

Share Browser WebEx

INGENUITY® PRODUCTS SCIENCE BLOG LOGIN

Choose a product to login

INGENUITY PATHWAY ANALYSIS

Comprehensive pathway and network analysis of complex 'omics data

NEW: INSTALL IPA CLIENT

Or use legacy Web Start login

INGENUITY iREPORT

For current customers

INGENUITY VARIANT ANALYSIS

Rapidly find causal variants using a knowledge-driven approach

Welcome! Please login

Email

Password

Remember my password

**LOG IN**

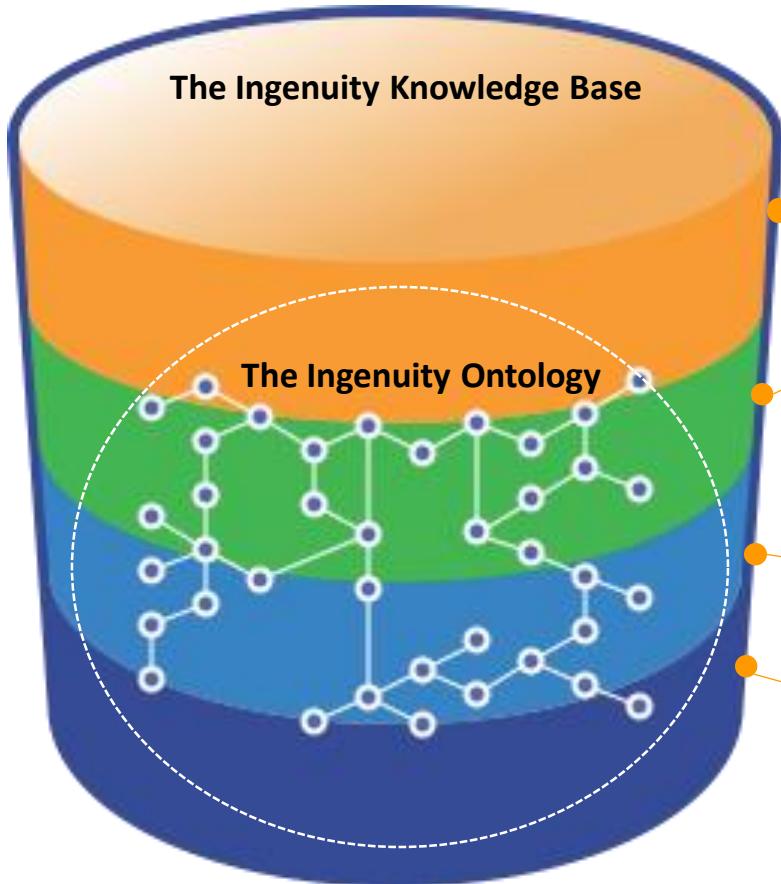
[Sign Up](#) | [Forgot Password](#)

# How can IPA help you?

## IPA

- Deep pathway understanding of a single gene/protein
  - Drug/therapeutic target discovery
- Biological understanding of large data sets, including
  - Differential gene expression, array and RNAseq (transcriptomics)
    - **Isoform annotation (New)**
  - Differential protein expression (proteomics)
  - **Genes with loss/gain-of-function variants (New)**
  - Metabolomics
  - miRNA expression
  - Gene List
    - Chip-seq
    - siRNA screening
    - Sequence Variants (see also Ingenuity Variant Analysis)
  - Methylation
  - Protein phosphorylation

## Ingenuity Knowledge Base



### Ingenuity Findings

**Ingenuity® Expert Findings** – Manually curated Findings that are reviewed, from the full-text, rich with contextual details, and are derived from top journals.

**Ingenuity® ExpertAssist Findings** – Automated text Findings that are reviewed, from abstracts, timely, and cover a broad range of publications.

### Ingenuity Modeled Knowledge

**Ingenuity® Expert Knowledge** – Content we model such as pathways, toxicity lists, etc.

**Ingenuity® Supported Third Party Information** – Content areas include Protein-Protein, miRNA, biomarker, clinical trial information, and others





## Core Analysis Steps

---

- File > New > Core Analysis
  - Or File > Data Set Upload
- Upload Data (gene expression, protein expression, metabolomics, etc.)
- Set Core Analysis Settings
- Run Analysis
- Interpret Results

# Data Upload

# Data Upload Format Examples

Typical value-types that are uploaded to IPA

## Identifier List

A
1 ID
2 NM_130786
3 NR_015380
4 NM_138932
5 NM_014576
6 NM_138933
7 NM_000014
8 NR_026971
9 NM_144670
10 NM_001080438
11 NM_017436
12 NM_016161
13 NM_015665

+differential expression

A	B
_130786	0.14
015380	-0.99
138932	-0.02
014576	-0.02
138933	0.02
000014	-4.79
026971	-0.67
144670	-5.96
001080438	-1.97
017436	-1.09
NM_016161	2.02
NM_015665	-0.27

Log2Ratio

+significance stat

A	B	C
	Log2Ratio	p-value
	0.14	8.68E-01
	-0.99	2.24E-01
	-0.02	9.83E-01
	-0.02	9.85E-01
	0.02	9.79E-01
	-4.79	1.02E-01
	-0.67	6.17E-01
	-5.96	1.30E-01
	-1.97	3.47E-01
	-1.09	5.02E-01
	2.02	5.97E-02
	-0.27	5.68E-01

Log2Ratio p-value

+RPKM

(maximum RPKM between experimental condition and control recommended for RNAseq)

A	B	C	D
	Log2Ratio	p-value	Intensity/ RPKM/FPKM
	0.14	8.68E-01	2931.69
	-0.99	2.24E-01	1649.26
	-0.02	9.83E-01	1.67
	-0.02	9.85E-01	1.77
	0.02	9.79E-01	1.83
	-4.79	1.02E-01	239.75
	-0.67	6.17E-01	213.79
	-5.96	1.30E-01	610.64
	-1.97	3.47E-01	3.91
	-1.09	5.02E-01	6186.83
	2.02	5.97E-02	149.85
	-0.27	5.68E-01	13330.34

# Uploading Multiple Observations

## Format for multi-observation upload

- Multiple experimental differential expressions can be grouped into a single spreadsheet and upload
  - Nice-to-have if you are comparing a series of expression values such as a time-course
  - Be sure and name your observations at the time of upload in IPA



	A	B	C	D	E	F	G
1	ID	12 Hour Log2Ratio	12 Hour p-value	12 Hour Intensity/	24 Hour Log2Ratio	24 Hour p-value	24 Hour Intensity/
2	NM_130786	0.14	8.68E-01	2931.69	-0.83	4.65E-01	4791.17
3	NR_015380	-0.99	2.24E-01	1649.26	0.72	5.32E-01	198.72
4	NM_138932	-0.02	9.83E-01	1.67	1.58	8.31E-03	7879.80
5	NM_014576	-0.02	9.85E-01	1.77	-0.77	1.26E-02	46757.06
6	NM_138933	0.02	9.79E-01	1.83	0.90	2.03E-02	26426.36
7	NM_000014	-4.79	1.02E-01	239.75	-0.01	9.82E-01	2117.73
8	NR_026971	-0.67	6.17E-01	213.79	0.12	8.64E-01	14076.24
9	NM_144670	-5.96	1.30E-01	610.64	-1.62	1.46E-01	31.85
10	NM_001080438	-1.97	3.47E-01	3.91	0.12	8.25E-01	10491.96
11	NM_017436	-1.09	5.02E-01	6186.83	2.02	4.44E-01	14788.50
12	NM_016161	2.02	5.97E-02	149.85	-0.57	1.09E-01	273101.00
13	NM_015665	-0.27	5.68E-01	13330.34	0.36	4.87E-01	11876.00
14	NM_023928	-1.12	1.03E-02	22828.45	-0.17	7.18E-01	3339.36

## Verify the differential expression calculation

- Recommend  $\text{Log}_2(\text{ratio})$  differential expression

$$\text{Log}_2\left(\frac{\text{Experimental Condition Exp.}}{\text{Control Exp.}}\right)$$

- Ratio differential expression

$$\left(\frac{\text{Experimental Condition Exp.}}{\text{Control Exp.}}\right)$$

- Fold Change

- If increased differential expression

$$\left(\frac{\text{Experimental Condition Exp.}}{\text{Control Exp.}}\right)$$

- If decreased differential expression

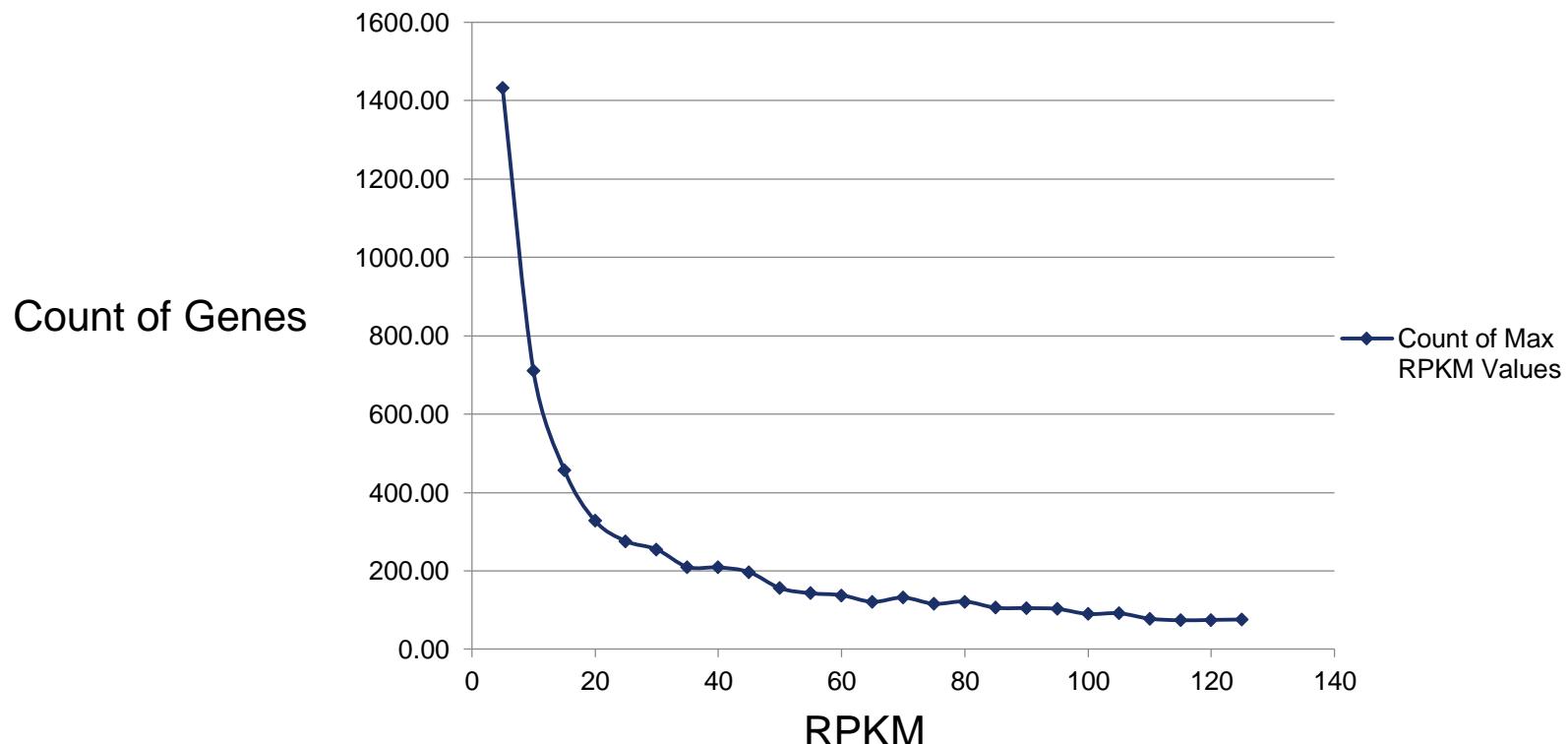
$$-1 \left(\frac{\text{Control Exp.}}{\text{Experimental Condition Exp.}}\right)$$

Fold change will never have values between 1 and -1

# Typical Distribution of RPKM Values in RNAseq Data

## Filtering on absolute expression

- RNAseq measurements often result in many significant differential fold changes at low absolute transcript expression levels
- Including the maximum RPKM value of your experimental condition and control allows for later filtering on absolute expression value in addition to fold change and p-value



## Best practices

- Calculate metrics outside of IPA (e.g. fold-change, p-value)
- Create an Excel spreadsheet or tab delimited file
  - Only 1 header row allowed
  - One column must have identifiers, preferably the left-most column
  - Can have up to 20 observations
  - IPA will only look at the top worksheet in an Excel workbook
- Group related observations into a single spreadsheet if possible
  - Time course, drug concentration, cell lines, etc.
- Specify array platform (chip) if possible
  - It is OK to use “Not specified/applicable”
- Pre-filter data at the lowest threshold that you have confidence in
  - For example, probe measurement p-value of .05 or other criteria
  - Use the Recalculate button to refresh the screen

- Examples of data set types
  - Differential gene expression, array and RNAseq (transcriptomics)
    - **Isoform annotation (New)**
  - Differential protein expression (proteomics)
  - **Genes with loss/gain-of-function variants (New)**
  - Metabolomics
  - miRNA expression
  - Gene List
    - Chip-seq
    - siRNA screening
    - Sequence Variants (see also Ingenuity Variant Analysis)
  - Methylation
  - Protein phosphorylation

## Why don't all of the molecules map?

- The gene ID might not correspond to a known gene product. For example, most ESTs are not found in the knowledge base (exception: ESTs that have a corresponding Entrez Gene identifier are found in the knowledge base).
- A gene/protein ID might correspond to several loci or more than one gene. Such identifiers are left unmapped in the application due to the ambiguity of the identity.
- Identifiers for species other than human, mouse or rat must map to human, mouse or rat orthologues in order to map in IPA.
- SNPs must map to a single gene. SNPs that fall greater than 2 KB upstream or 0.5 KB downstream of a gene coding region will not be mapped in IPA during data upload, since they cannot be unambiguously mapped to a single gene.
- There may be insufficient findings in the literature regarding some molecules.

## How Do I Choose The Reference Set?

If you are using a standard vendor platform supported by IPA, then that platform should be selected as your reference set.

If you do not know the platform or the data were taken from different platforms, select a reference set that best estimates the entire population you evaluated.

- For gene expression data, select the “Ingenuity Knowledge Base (genes only)”
- For metabolomics, select the “Ingenuity Knowledge Base (endogenous chemicals only)”
- You have the option to having your uploaded data set used as the reference set (User Data Set)

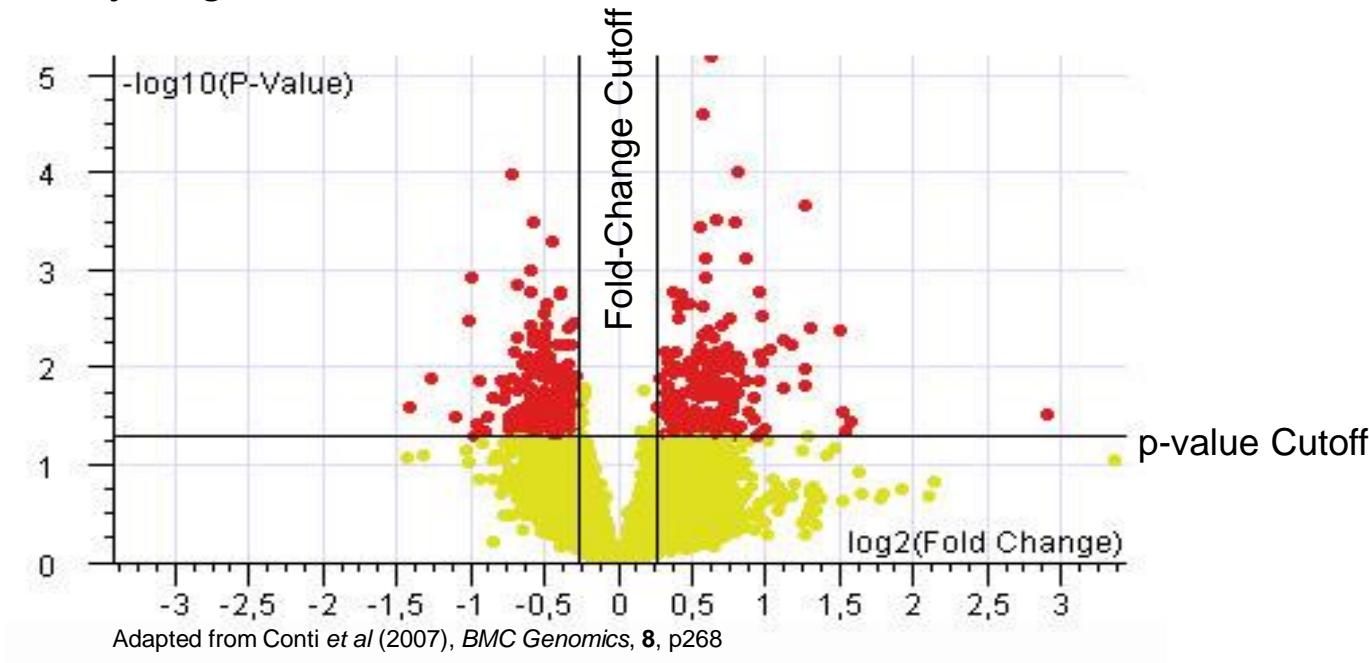


## Core Analysis Set-up

## Use cutoffs to refine set size per observation

'Ideal' set size for IPA core analysis from gene expression data is typically 200-3000

- Small sets will not have many directional effect z-scores (downstream functions, upstream regulators)
- Very large data sets will tend to have more 'noise'



# Creating an IPA Core Analysis

**Create Core Analysis - [analysis : GSE26129\_MCF-7\_A2780\_IPA.xls]**

**General Settings**

**Population of genes to consider for p-value calculations:**  
Reference Set: Whole Human Genome Microarray 4x44K v2

**Relationships to consider:**  
Affects networks and upstream regulator analysis  
 Direct and Indirect Relationships  
 Direct Relationships

**Optional Analyses:**  
 My Project

**Analysis Filter Summary**  
Consider only relationships where confidence = Experimentally Observed

**Set Cutoffs**

**Expression Value Type**: Cutoff Range Focus On

- Log Ratio: -8.8771 to 8.2299 Both Up/Downregulated RECALCULATE
- Fold Change: 3.0 -470.206 to 300.24 Both Up/Downregulated
- p-value: 0.01 0.0 to 0.9999

Click here to apply filter cutoffs and see number that are network and function eligible

19495 analysis-ready molecules across observations

**Preview Dataset GSE26129\_MCF-7\_A2780\_IPA.xls Observation: MCF-7TxtD10 vs. MCF-7cc (19495)**

**Analysis-Ready (19495) Mapped IDs (31984) Unmapped IDs (41000)**

**View other observations if a multi-observation data set**

Symbol	Entrez Gene Name	Location	Type(s)	Drug(s)
A1BG*	alpha-1-B glycoprotein	Extracellular Space	other	
A1BG-AS1	A1BG antisense RNA	unknown	other	
A1CF*	ANPBP1	Nucleus		

**RUN ANALYSIS** **CANCEL**

Make sure reference set matches source of molecules

Assembles networks and identifies transcriptional regulators with only direct relationships. Results in networks in which members are nearer neighbors of one another and biases for binding relationships.

Network and function eligible molecules should be 100-2000 for best results, but other values can work

# Creating an IPA Core Analysis- Network Generation

Option to exclude endogenous chemicals from networks

Option to turn off molecular networks for a faster analysis

Fine-tune format of networks

Turn On Causal Network (Advanced Analytics)

**GENERAL SETTINGS**

Select the following Networks (increases analysis time)

Interaction networks

Include endogenous chemicals  
Genes are always included

Molecules per network: 35

Networks per analysis: 25

Causal networks  
Store functions, diseases and genes for each network (max 50)

**ADVANCED** **SAVE AS DEFAULTS**

**Set Cutoffs**

Expression Value Type	Cutoff	Range	Focus On
Log Ratio		-8.8771 to 8.2299	Both Up/Downregulated
Fold Change	3.0	-470.206 to 300.24	Both Up/Downregulated
p-value	0.01	0.0 to 0.9999	

**RECALCULATE** 2026 analysis-ready molecules across observations

Preview Dataset GSE26129\_MCF-7\_A2780\_IPA.xls Observation: MCF-7TxtD10 vs. MCF-7cc (840)

Analysis-Ready (840) Mapped IDs (31984) Unmapped IDs (9016) All IDs (41000)

**ADD TO MY PATHWAY** **ADD TO MY LIST** **CREATE DATASET** **CUSTOMIZE TABLE**

	Log Ratio	Fold Change	p-value	ID	Notes	Symbol	Entrez Gene Name	Location	Type(s)	Drug(s)
<input type="checkbox"/>	↑3.668	↑12.712	1.69E-04	A_24_P235429	D	ABCA1*	ATP-binding cas	Plasma Membra	transporter	probucol
<input type="checkbox"/>	↑1.697	↑3.242	7.76E-04	A_24_P67096	D	ABCA5*	ATP-binding cas	Plasma Membra	transporter	
<input type="checkbox"/>	↑4.280	↑30.815	1.10E-04	A_22_082827	D	ABC1	ATP binding cas	Plasma Membra	transporter	

Rows: 1 - 100

**RUN ANALYSIS** **CANCEL**

# Creating an IPA Core Analysis: Using Filters

**Create Core Analysis - [analysis : Time course. Treated vs untreated]**

Several filters available. Set criteria to filter out findings of less interest.

**General Settings**

- Network Generation Options**
- Data Sources** All
- Confidence** Experimentally Observed
- Species** All
- Tissues & Cell Lines** All
- Mutation** All

**ADVANCED** **SAVE AS DEFAULTS**

**Population of analysis** **Entrez Gene** **Protein** **Small Molecule** **Pathway** **Cell Type** **Organism** **Regulation**

Relationships: **Indirect Relationships** **Direct Relationships**

Optional Analyses:

- My Project**
  - My Pathways
  - My Lists
- Ingenuity CWS**
  - My Pathways
  - My Lists
- Alcon**
  - My Pathways
  - My Lists

**Analysis Filter Summary**

Consider only relationships where confidence = Experimentally Observed

**Set Cutoffs**

Expression Value Type	Cutoff	Range	Focus On
Fold Change	2	-17.2747 to 46.8718	Both Up/Downregulated
p-value	.05	0.0 to 0.9994	

**RECALCULATE** 363 analysis-ready molecules across observations

**Preview Dataset** Time course. Treated vs untreated Observation: 120 hours (304)

Analysis-Ready (304) | Mapped IDs (461) | Unmapped IDs (20) | All IDs (481)

**ADD TO MY PATHWAY**
**ADD TO MY LIST**
**CREATE DATASET**
**CUSTOMIZE TABLE**

	Fold Change	p-value	ID	Notes	Symbol	Entrez Gene Name	Location	Type(s)	Drug(s)
<input type="checkbox"/>	↑2.974	2.00E-06	209459_s_at		<b>ABAT</b>	4-aminobutyrate a	Cytoplasm	enzyme	valproic acid, vig...
<input type="checkbox"/>	↑14.652	1.59E-04	207692_s_at	D	<b>ACAN*</b>	aggrekan	Extracellular Space	other	
<input type="checkbox"/>	↑41.570	0.00E00	205132_at	D	<b>ACTC1</b>	actin, alpha, cardia	Cytoplasm	enzyme	
<input type="checkbox"/>	↑19.161	0.00E00	226814_at	D	<b>ADAMTS9*</b>	ADAM metallopep	Extracellular Space	peptidase	
<input type="checkbox"/>	↑4.054	0.00E00	213974_at	D	<b>ADAMTSL3</b>	ADAMTS-like 3	unknown	other	
<input type="checkbox"/>	↑2.267	3.00E-06	205771_s_at		<b>AKAP7</b>	A kinase (PRKA) an	Plasma Membrane	other	
<input type="checkbox"/>	↑4.429	1.00F-06	215783_s_at		<b>ALPL</b>	alkaline phosphata	Plasma Membrane	phosphatase	

Rows: 1 - 50

**RUN ANALYSIS** **CANCEL**

# Creating an IPA Core Analysis- Advanced Settings

**Create Core Analysis - [analysis : Time course. Treated vs untreated]**

**General Settings**

**Population of genes to consider for p-value calculations:**  
Reference Set: Human Genome U133 Plus 2.0 Array

**Relationships to consider:**  
Affects networks and transcription factor analysis  
 Direct and Indirect Relationships  
 Direct Relationships

**Optional Analyses:**

**Analysis Filter Summary**  
Consider only relationships where confidence = Experimentally Observed

**Advanced Settings**

Select expression value for node coloring: Fold Change

This expression value type will be used to calculate the directionality of functions and will be displayed in color on pathways and networks.

**Duplicate Resolution**

When IDs map to the same gene, protein, or other molecule:

Apply cutoffs before consolidating IDs: Yes (recommended)

Resolve duplicates using Exp Value: Fold Change

Consolidate IDs using the expression value: maximum

**Observation to Include**

Include	Observation Name	Analysis Ready
#1 <input type="checkbox"/>	2 hours	9
#2 <input checked="" type="checkbox"/>	24 hours	169
#3 <input checked="" type="checkbox"/>	120 hours	304

OK CANCEL ANALYSIS CANCEL

**Annotations**

Make sure molecule coloring is set for a metric such as fold change, log ratio, etc.

Confirm how you would like to resolve duplicates

Deselect any observations that you would like to exclude from the analysis

Set criteria to filter out findings of less interest.

- Species
- Tissue

Filter stringency

- A “Stringent” setting requires that each of a pair of molecules and the relationship that connects them meet the filter criteria
- A “Relaxed” filter requires that the gene or protein expression of the molecules connected by a relationship meet the filter criteria

## Using Core Analysis Pre-filters, Cont.

Unspecified refers to findings or molecules where cell/tissue/organ is not specified or classified

### protein-protein interactions [1]

+ Binding of **MATRILYSIN [MMP7]** protein and human **TIMP2** protein occurs in a cell-free system.

#### Pre-filter Advantages

- Focuses IPA analysis on networks, biological functions, and canonical pathways on molecules and relationships closely related to the experiment.

#### Pre filter Disadvantages

- Loss of information
- Loss of relationships that may be applicable to your species or tissue but were described in a different speices or tissue.

Create Core Analysis - [analysis : Prostate Disease.txt]

Filters and General Settings

General Settings ?

Species All ?

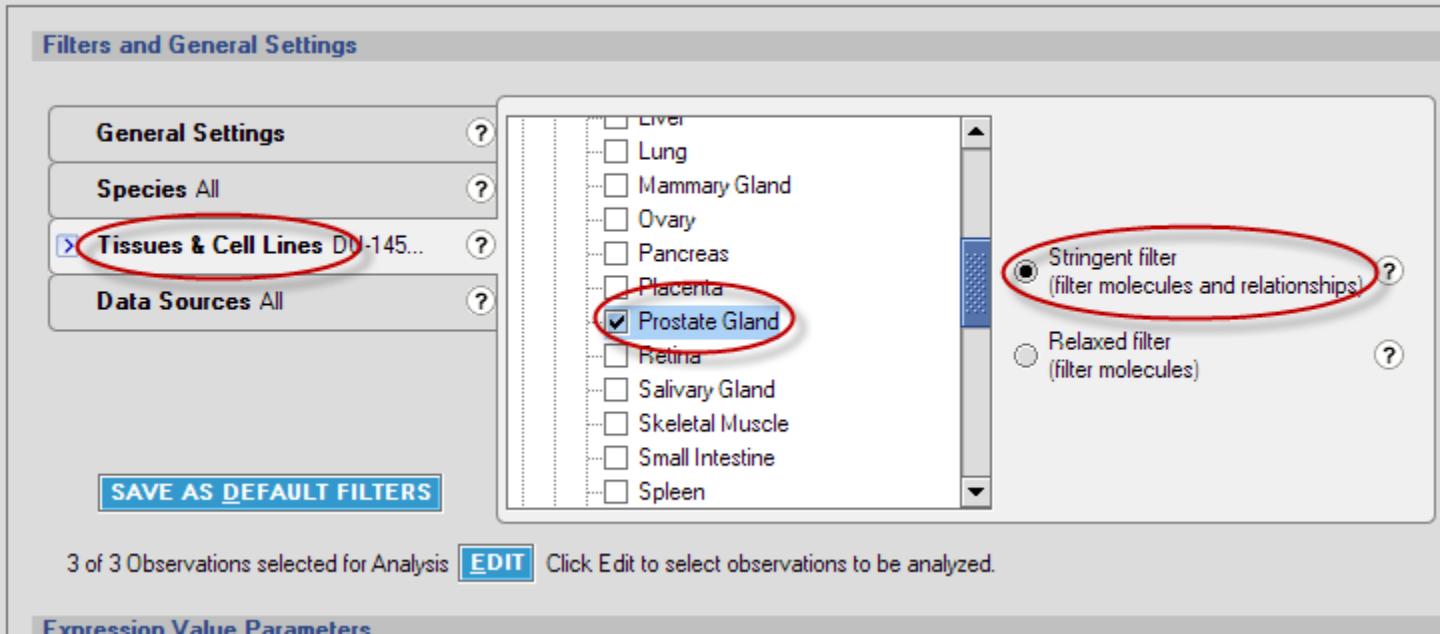
Tissues & Cell Lines DN-145... ?

Data Sources All ?

SAVE AS DEFAULT FILTERS

3 of 3 Observations selected for Analysis **EDIT** Click Edit to select observations to be analyzed.

Expression Value Parameters



Liver  
Lung  
Mammary Gland  
Ovary  
Pancreas  
Placenta  
 Prostate Gland  
Retina  
Salivary Gland  
Skeletal Muscle  
Small Intestine  
Spleen

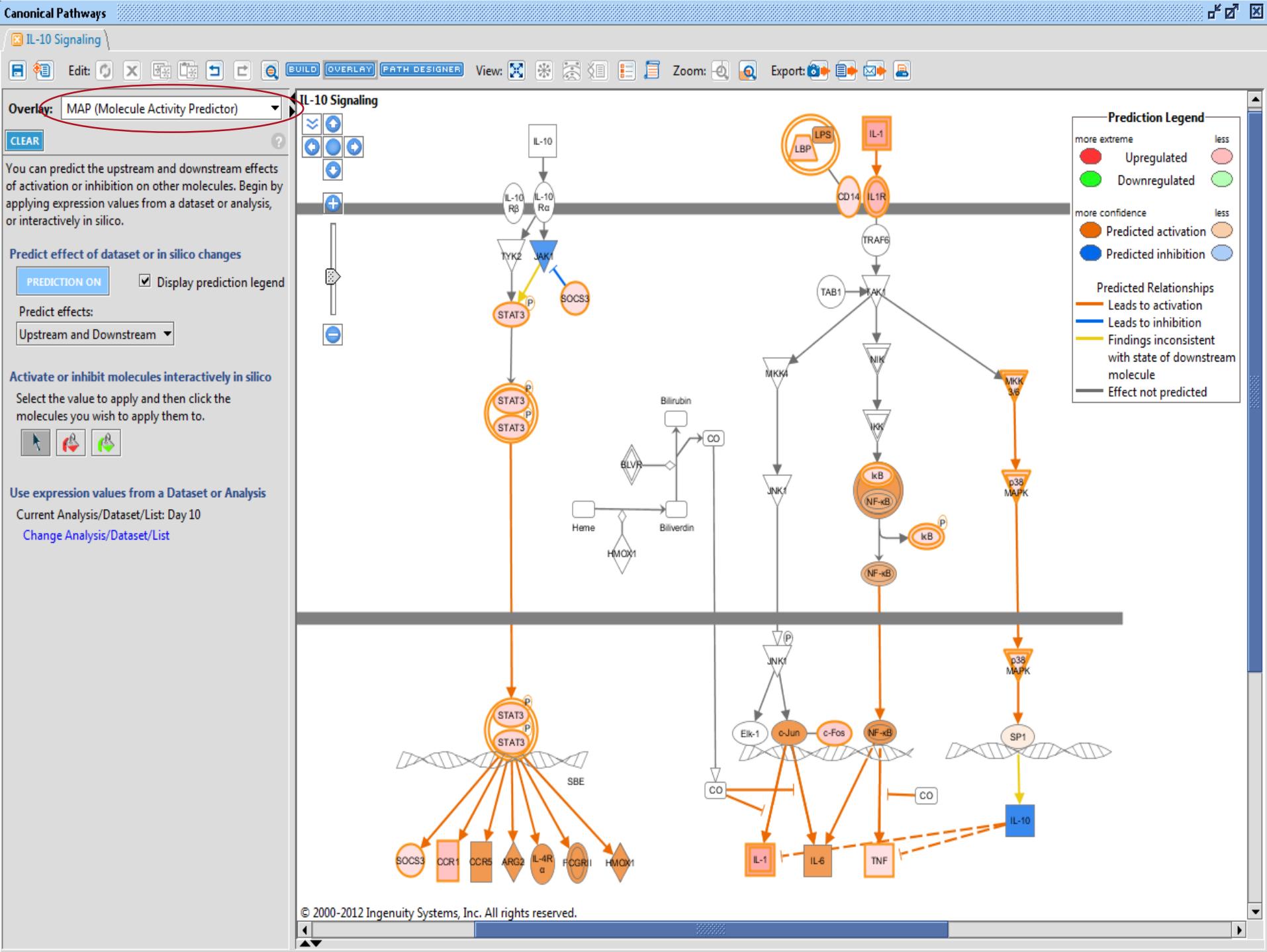
Stringent filter (filter molecules and relationships) ?

Relaxed filter (filter molecules) ?

# Large Scale Data Analysis

## IPA Core Analysis

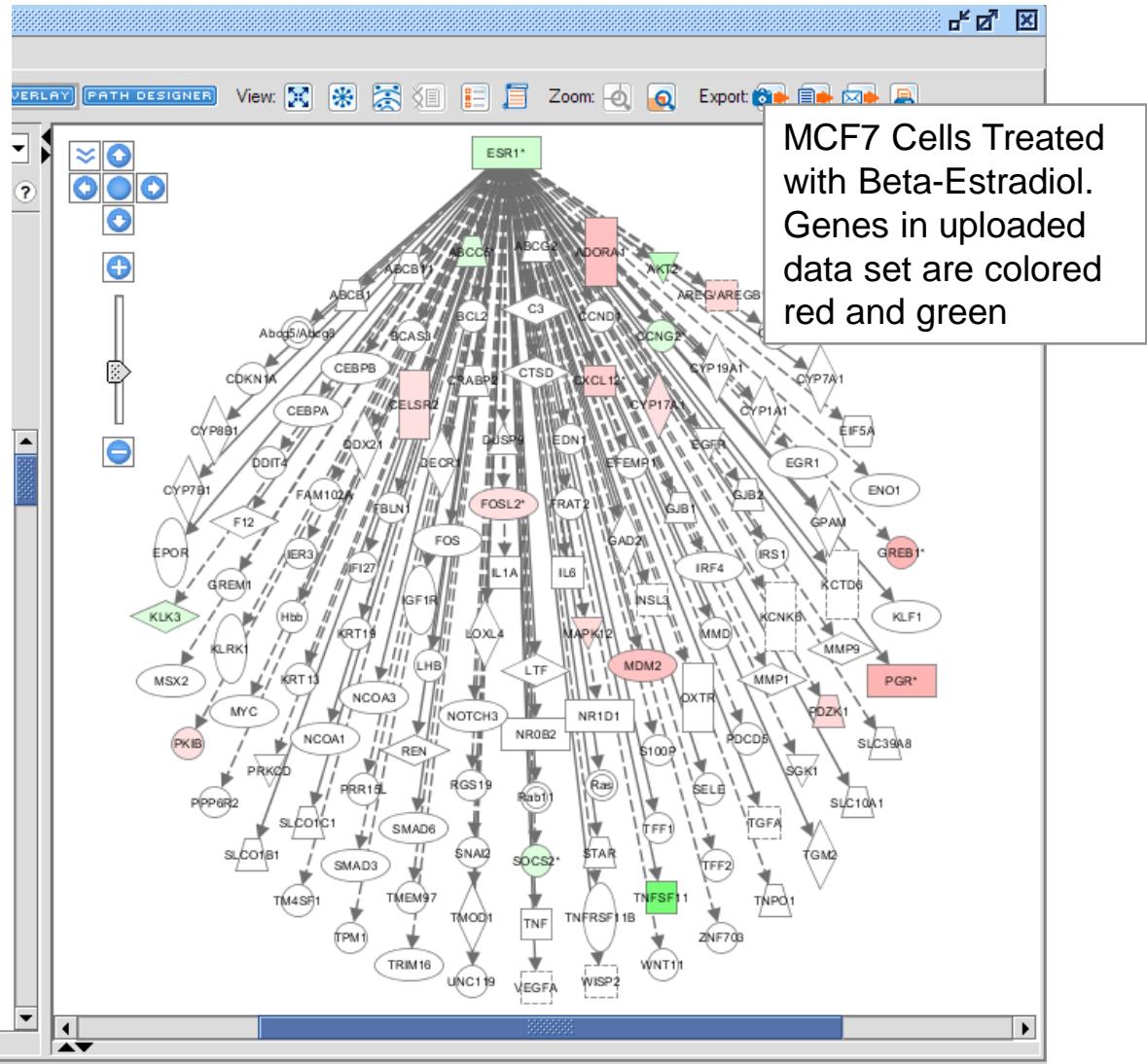
- Pathway Analysis
  - Predicts pathways that are changing based on gene expression
  - New tools to predict directional effects on the pathway (MAP overlay tool)
- Upstream Regulator Analysis
  - Predicts what regulators caused changes in gene expression
  - Predicts directional state of regulator
  - Creates de novo pathways based on upstream regulators (Mechanistic Networks)
- Diseases and Functions Analysis
  - Predicts effected biology (cellular processes, biological functions) based on gene expression and predicts directional change on that effect
    - “Increase in cell cycle”
    - “Decrease in apoptosis”
- Regulator Effects
  - Models pathway interactions from predicted upstream regulators, through differentially expressed genes, to biological processes
- Networks
  - Predicts non-directional gene interaction map



# Identify likely upstream regulators and their activity state

## IPA Upstream Regulator Analysis

- Use published experimental molecular interactions to identify upstream regulators
- Identify upstream regulators by determining gene enrichment in downstream genes
- Predict the activity state of regulators by correlating literature reported effects with observed gene expression

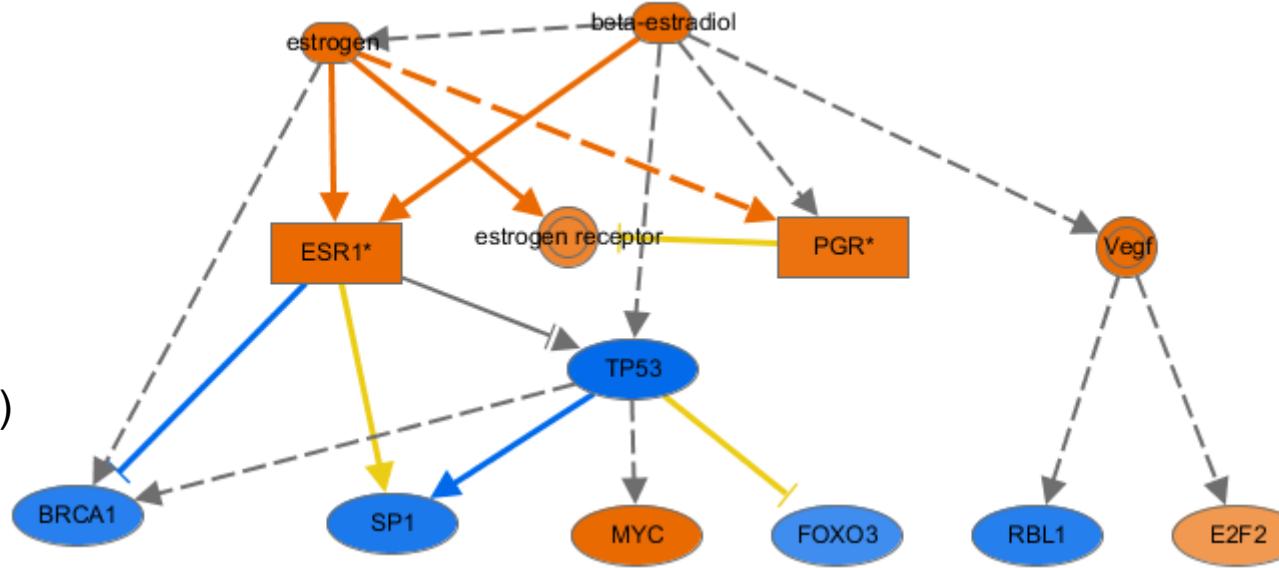


# Create *de novo* pathways of regulators and genes

## IPA Mechanistic Networks

Upstream Analysis								
Upstream Regulators		Causal Networks						
ADD TO MY PATHWAY		ADD TO MY LIST		CUSTOMIZE TABLE		DISPLAY AS NETWORK		MECHANISTIC NETWORKS
Upstream Regulator	Fold Change	Molecule Type	Predicted Activation	Activation z-score	p-value of overlap	Target molecules in	Mechanistic Net...	
beta-estradiol		chemical - endogenous group	Activated	6.097	1.24E-26	↓ABCA1, ↓... all 122	186 (13)	
Mek		chemical drug	Activated	3.683	7.37E-07	↑ABCE1, ↑... all 16		
estrogen		ligand-dependent nuclear receptor	Activated	3.661	2.00E-04	↓ABCA1, ↑... all 19	129 (13)	
ESR1	↓-1.708	cytokine	Activated	3.504	2.84E-13	↓ABCC5, ↑... all 37	183 (13)	
IL3				3.190	1.74E-02	↑ADA, ↑AR, ... all 16		

- Identify potential upstream regulator signal transduction
- Using shared downstream gene effects and gene-gene interactions, pathways (mechanistic networks) are created.



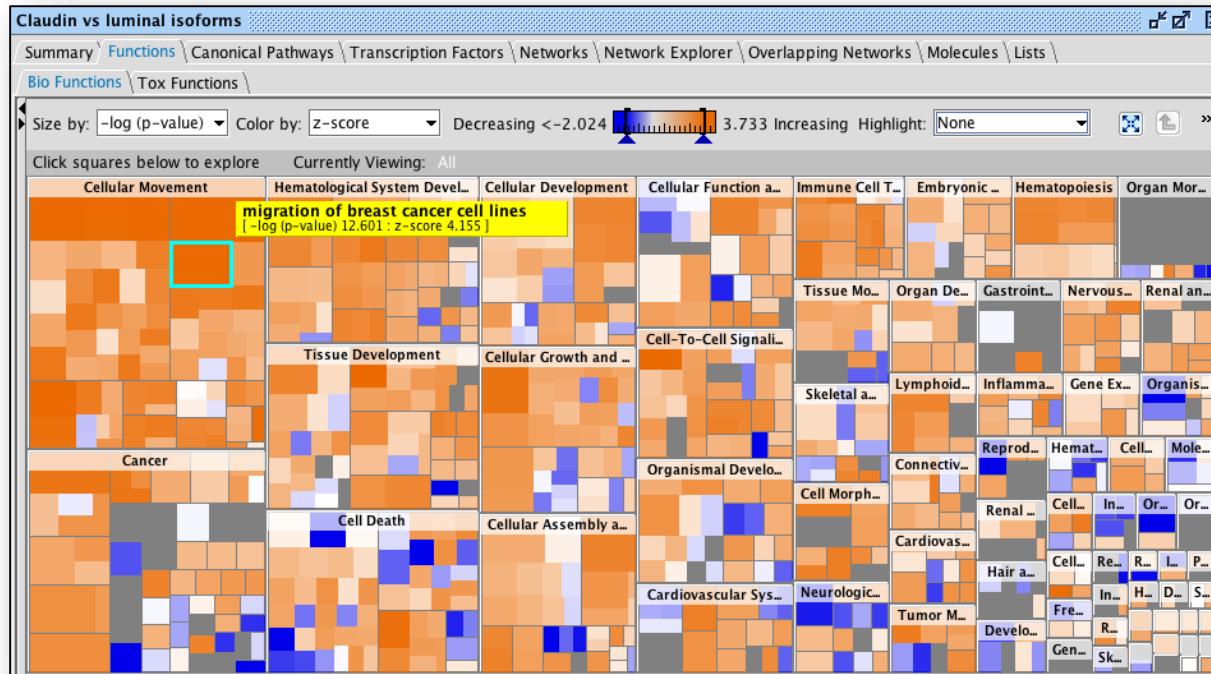
IPA Winter Release 2012

# A novel approach to visualize and predict biological impact of gene expression changes

## Downstream Effects Analysis

Identify key biological processes influenced by differentially expressed genes

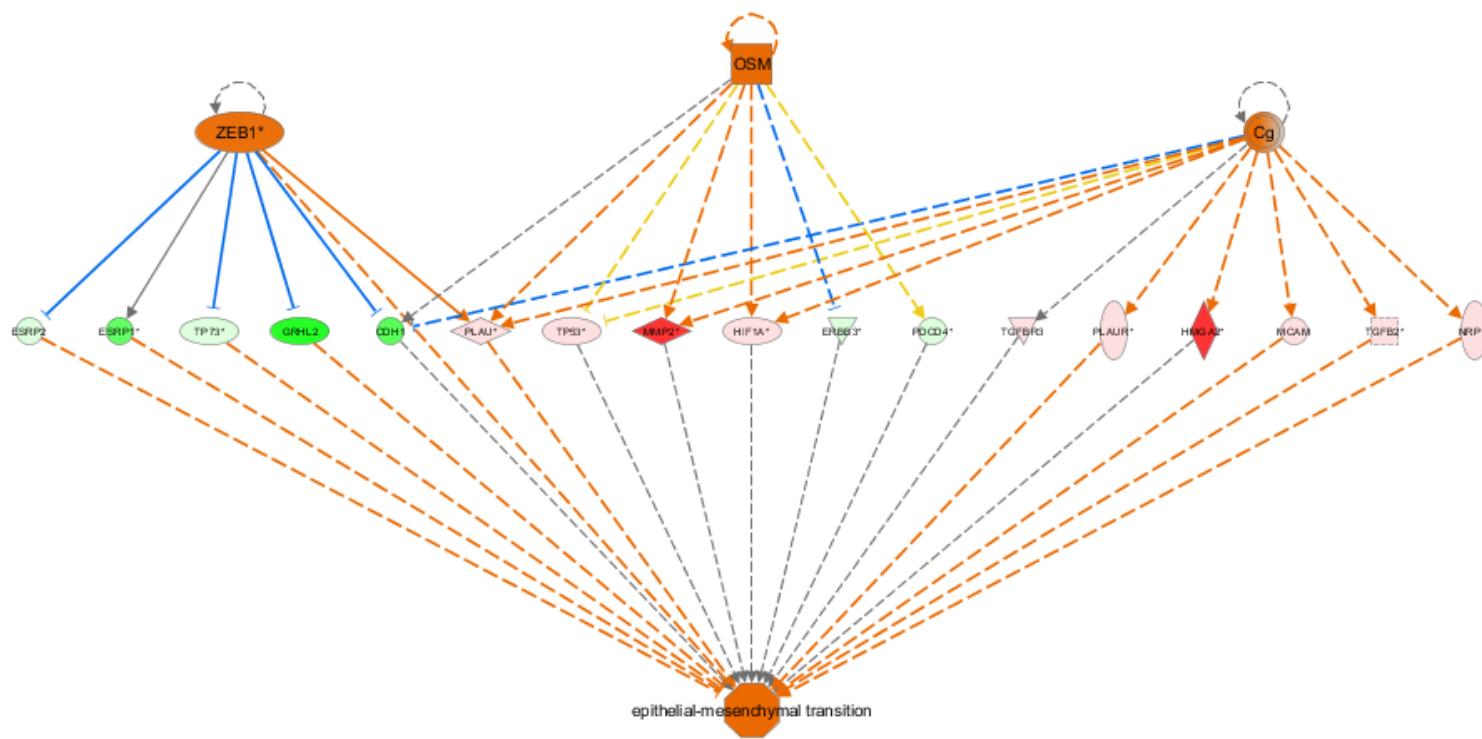
Understand whether cellular processes are being driven up or down by correlating observed expression with reported experimental gene effects



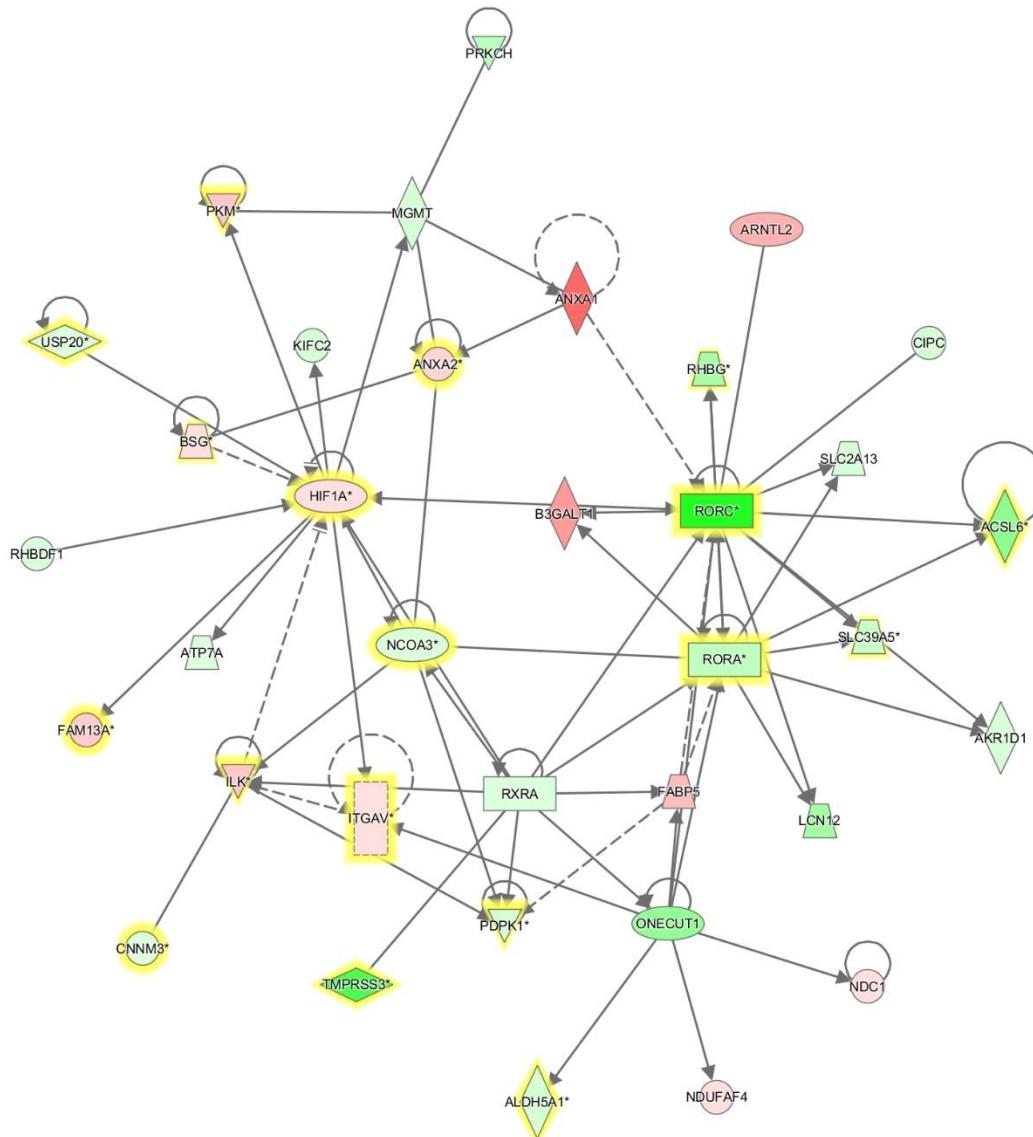
- Each box represents a biological process or disease
- The size of the box represents gene enrichment
- The color of the box indicates the predicted increase or decrease

## Regulator Effects

- Hypothesis for how a phenotype, function or disease is regulated in the dataset by activated or inhibited upstream regulators
- Explain impact of upstream molecules on downstream biology
- Explain potential mechanism for a phenotype or drug
- Define drug targets
- 



# Network Analysis



# Analyzing and Interpreting Results



## Approaches to Viewing Results

---

- IPA will subdivide your data into slices based on molecule connectivity (networks), cellular functions, and involvement in canonical pathways
- Spend time surveying the information. Not everything is of scientific interest, look for slices of your data that address your scientific question, are consistent with known biological processes, are consistent with pathology, etc.
- Typically the goal will be to find a set of genes/molecules that can be looked at in greater detail by building a custom pathway
- If you are comparing observations, run comparison analysis.

IPA calculates two distinct statistics as part of a core analysis

- P-value:
  - Calculated using a Right-Tailed Fisher's Exact Test
  - Reflects the likelihood that the association or overlap between a set of significant molecules from your experiment and a given process/pathway/transcription neighborhood is due to random chance. The smaller the p-value the less likely that the association is random.
  - The p-value does not consider the directional effect of one molecule on another, or the direction of change of molecules in the dataset.
  
- Z-score:
  - Applied in some analysis types and provides predictions about upstream or downstream processes.
  - Takes into account the directional effect of one molecule on another molecule or on a process, and the direction of change of molecules in the dataset.

# Analyzing Results

## Canonical Pathway Analysis



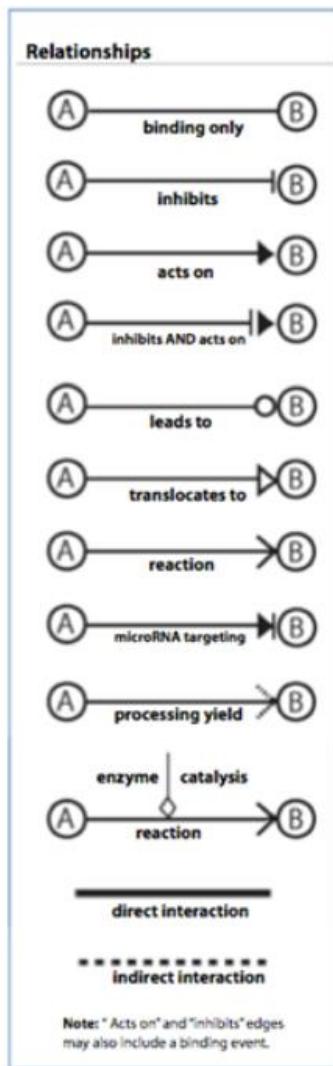
# Pathway Analysis

---

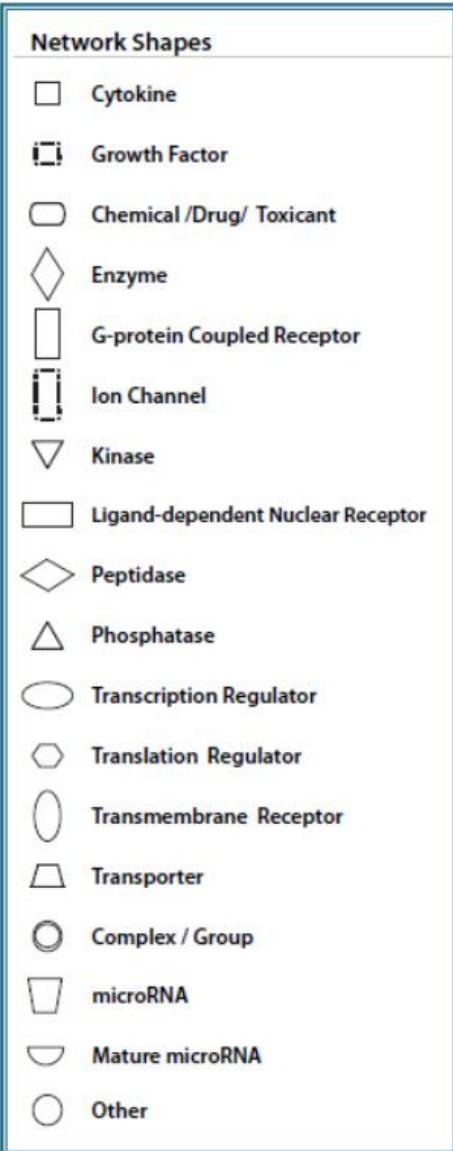
- What known biological pathways appear most significantly affected by the genes in my data set?
  
- What genes within a pathway are changing in expression and what effect might that change have on the pathway?

- Bar-chart represents significance of gene enrichment for any given pathway
  - Significance is most important metric
- Ignore bumpy yellow line: ratio/percent coverage of a pathway subject to pathway size bias
- Bar-chart color indicates predicted directionality
  - When considering pathway directionality, focus on  $2 < z\text{-score} < -2$
  - Just because a pathway does not have a good z-score does not make it uninteresting
- To open pathway, look for open pathway button on far right after bar-chart selection

# Go to Help > Legend...Print it out



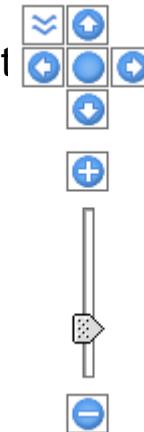
Relationship Labels	
A	Activation
B	Binding
C	Causes/Leads to
CC	Chemical-Chemical interaction
CP	Chemical-Protein interaction
E	Expression (includes metabolism/ synthesis)
EC	Enzyme Catalysis
I	Inhibition
L	Proteolysis (includes degradation for Chemicals)
LO	Localization
M	Biochemical Modification
miT	microRNA Targeting
MB	Group/complex Membership
nTRR	Non-Targeting RNA-RNA Interaction
P	Phosphorylation/Dephosphorylation
PD	Protein-DNA binding
PP	Protein-Protein binding
PR	Protein-RNA binding
PY	Processing Yields
RB	Regulation of Binding
RE	Reaction
RR	RNA-RNA Binding
T	Transcription
TR	Translocation
UB	Ubiquitination





## Interpretation Tips

- Look for pathway biological themes
  - Use Overlapping Pathway tab to filter and view pathways with shared genes
  - Often lesser scoring pathways of a theme are simply subsets of genes found in a better scoring pathway
- Scan CP names for pathways of particular interest
  - Statistical significance does not equal biological significance and visa-versa
  - Pathways may have many second messengers which can be regulated post-transcriptionally
- View pathways by clicking the bar-chart and the OPEN PATHWAY button on right
- Use MAP tool (OVERLAY tool) to help interpretation
- Overlay other analyses as applicable
- Toggle overlay options

- Scroll-wheel on mouse controls zoom, or use toolbar zoom button
- Left-click selects (turns blue)
- Left-click-drag on nodes moves the node
- Right-click hold-and-drag moves your view
- Right-click brings up menu for controlling
  - tool tip (mouse-over node pop-up)
  - copy/paste
  - Highlighting (colored outline)
  - selection
- Node shapes indicate a protein's primary function, see Help>Legend
- Relationship lines indicate the type of relationship and the mouse-over letter the type of relationship, see Help>Legend

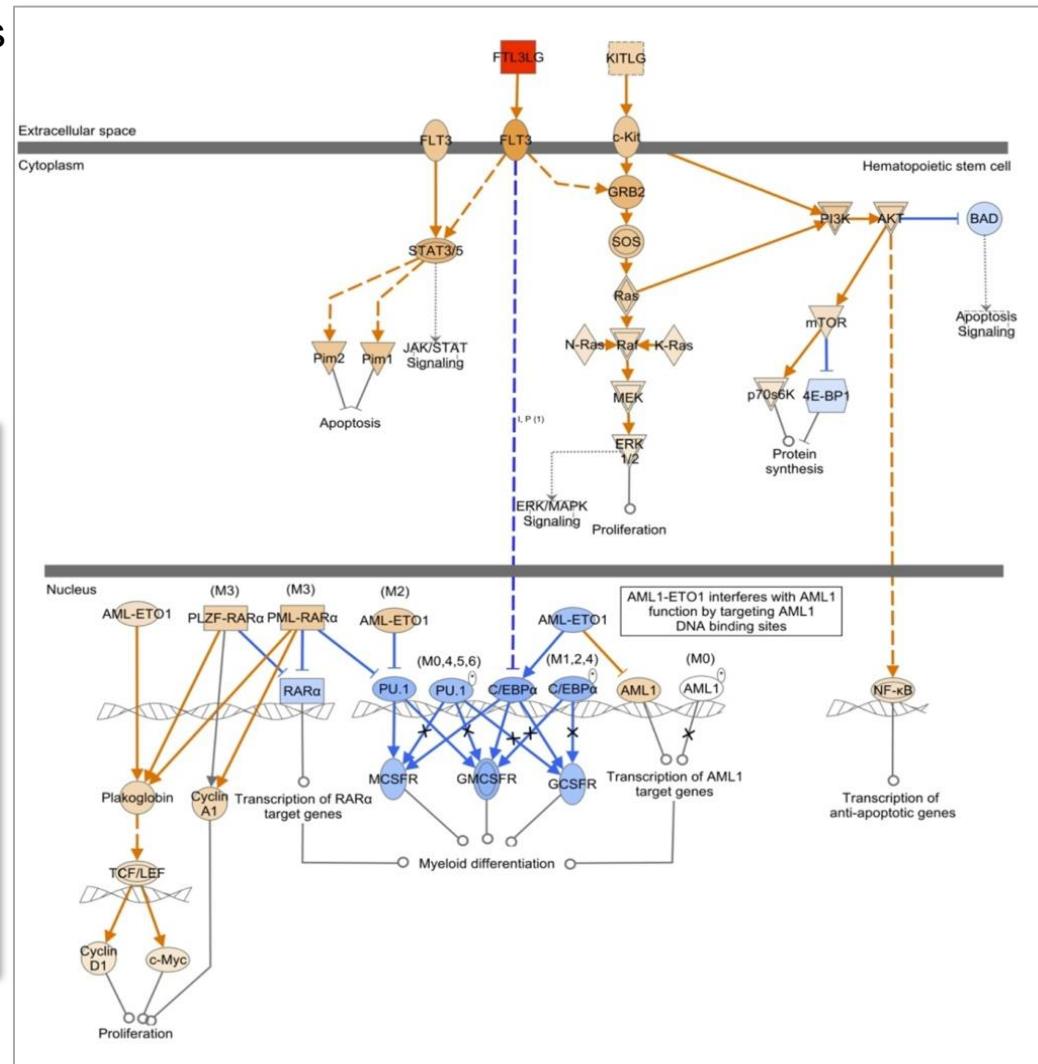
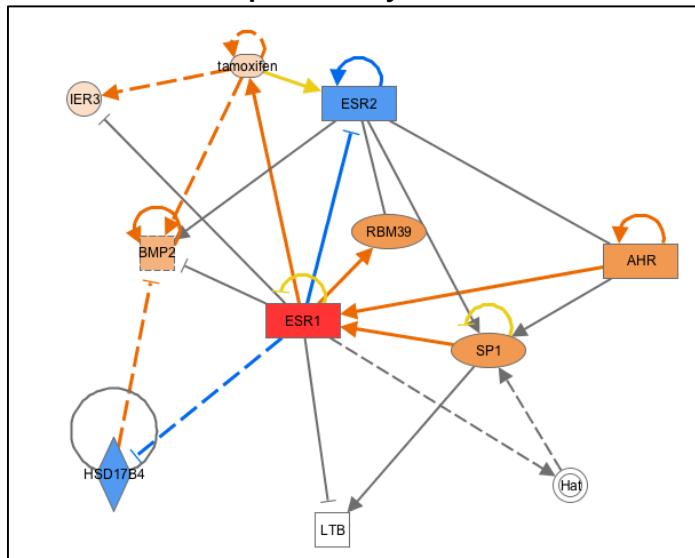
Navigation Control

- Double-clicking a node brings up the node summary
  - You can navigate to the Gene/Chem View page by clicking the protein name at the top of the summary window pane.
- Double-clicking a relationship line brings up the relationship summary
  - You can to the literature evidence findings by clicking the "View relationships between:..." link at the top of the summary window pane.
- Groups
  - Groups are represented by a double outline applicable to any molecule shape. These represent cases where findings use a general gene name to describe a gene class or group of isoforms
  - Complexes of different proteins are also given a double outline
  - **View members by left-click selecting, then right-click>Show Membership**

# Molecule Activity Predictor (MAP)

OVERLAY button -> MAP (Molecule Activity Predictor)

- Use observed expression changes to suggest functional effects on neighboring molecules
- Manually set activation states to observe predicted effects on canonical pathways



## Overlay other uploaded data sets, analyses

OVERLAY button -> Analyses, Data sets, and Lists

- Select other analyses from projects
- Useful for comparisons

## Analyzing Results Upstream Regulators



- What transcription factors likely led to observed gene expression changes?
- What *de novo* pathways can be created based on predicted upstream regulator interactions?

# IPA Upstream Regulator Analysis

Identify important signaling molecules for a more complete regulatory picture

The screenshot shows the IPA Upstream Regulator Analysis interface. The top navigation bar includes links for Summary, Functions, Canonical Pathways, Upstream Analysis (which is selected), Networks, Molecules, Lists, and My Pathways. Below this is a secondary navigation bar with Upstream Regulators and Causal Networks. The main content area is a table titled "Upstream Regulators". The columns are: Upstream Regulator, Fold Change, Molecule Type, Predicted Activation, Activation z-score, p-value of over..., Target molecules in ..., and Mechanistic Net... . The table lists various regulators like beta-estradiol, raloxifene, ESR1, etc., with their respective fold changes, activation status, and mechanistic networks.

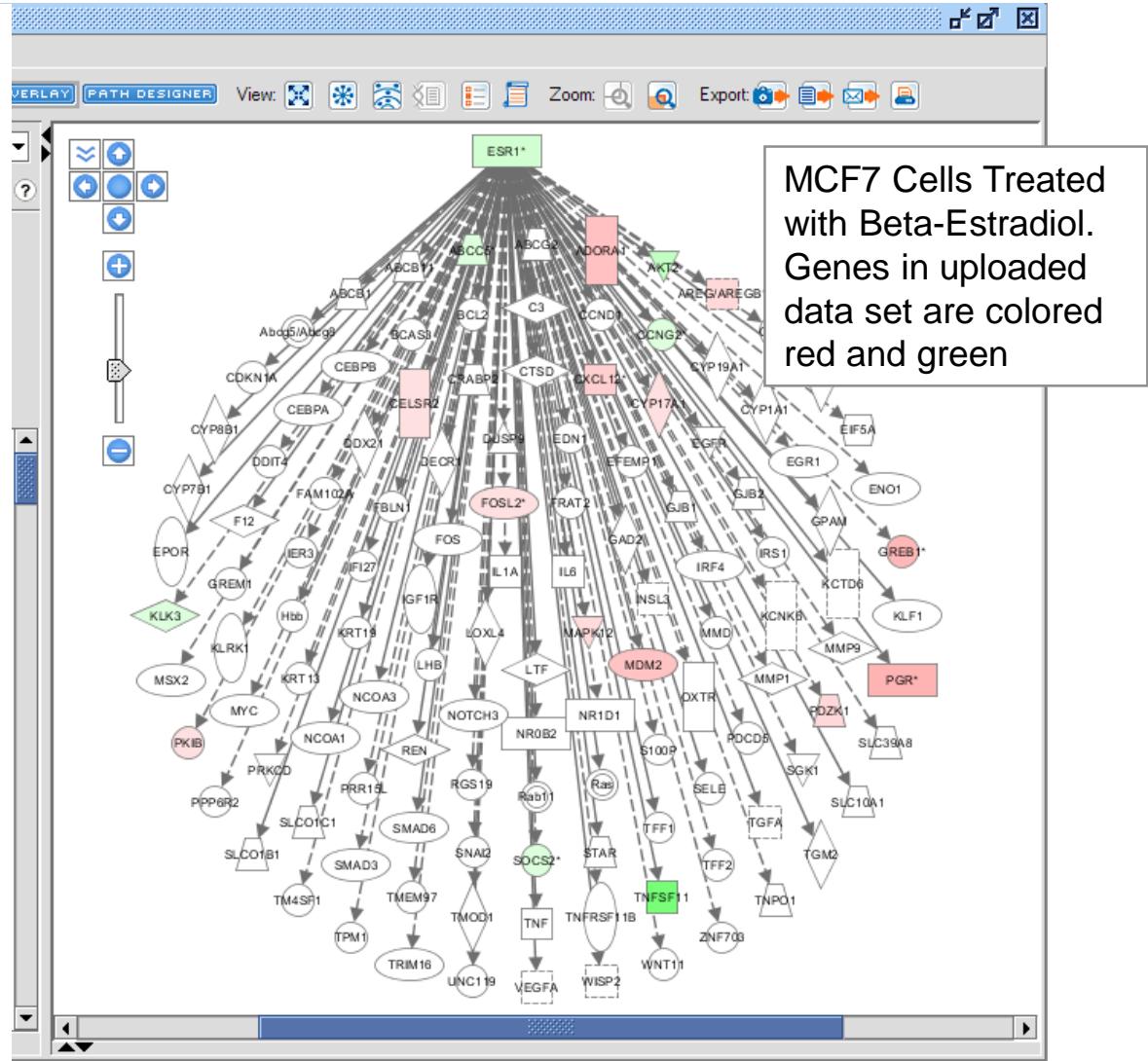
Upstream Regulator	Fold Change	Molecule Type	Predicted Activation	Activation z-score	p-value of over...	Target molecules in ...	Mechanistic Net...
beta-estradiol		chemical - endogenous	Activated	6.097	1.24E-26	↓ABCA1, ↓... all 122	186 (17)
raloxifene		chemical drug		-0.751	1.83E-14	↑AREG/AREGB...all 28	125 (15)
ESR1	↓-1.708	ligand-dependent nu	Activated	3.504	2.84E-13	↓ABCC5, ↑... all 37	186 (20)
trichostatin A		chemical drug		-0.620	1.02E-11	↓ABCA1, ↑... all 45	189 (20)
valproic acid		chemical drug		-1.593	5.88E-11	↓ABCG1, ↓... all 33	193 (18)
fulvestrant		chemical drug	Inhibited	-2.826	2.72E-10	↑ADRBI, ↑... all 27	181 (22)
TGFB1		growth factor		-0.325	4.02E-10	↓ABCA1, ↑... all 86	187 (20)
RAF1		kinase		-0.321	6.96E-10	↑AREG/AREGB...all 25	145 (19)
ESR2		ligand-dependent nu		0.095	9.98E-09	↑ADORA1, ↑... all 18	184 (19)
MYC	↑1.855	transcription regulat	Activated	2.599	1.02E-08	↑ABC1, ↓... all 52	157 (15)
CCND1	↑1.371	other		0.777	1.28E-08	↑BCL2, ↑B... all 24	154 (18)
ERBB2	↓-1.822	kinase		0.591	1.68E-08	↑AREG/AREGB...all 43	144 (20)
TNF		cytokine		-0.134	2.00E-08	↓ABCA1, ↓... all 77	227 (22)
dexamethasone		chemical drug		-0.930	2.64E-08	↑ABHD2, ↓... all 79	203 (18)
Salmonella enterica s		chemical toxicant		1.149	5.11E-08	↑AREG/AREGB...all 20	
ZNF217	↓-1.315	transcription regulat		0.555	7.38E-08	↓ADM, ↓A... all 13	
PGR	↑5.528	ligand-dependent nu		1.879	8.41E-08	↑AREG/AREGB...all 18	168 (20)
tretinoin		chemical - endogenous	Inhibited	-2.611	1.02E-07	↓ABCA1, ↑... all 70	159 (20)
methylselenic acid		chemical reagent		0.152	1.16E-07	↓ACSL3, ↑... all 24	180 (12)

Selected/Total upstream regulators : 0/709

- Quickly filter by molecule type
- Filter by biological context
- Generate regulators-targets Network to identify key relationships

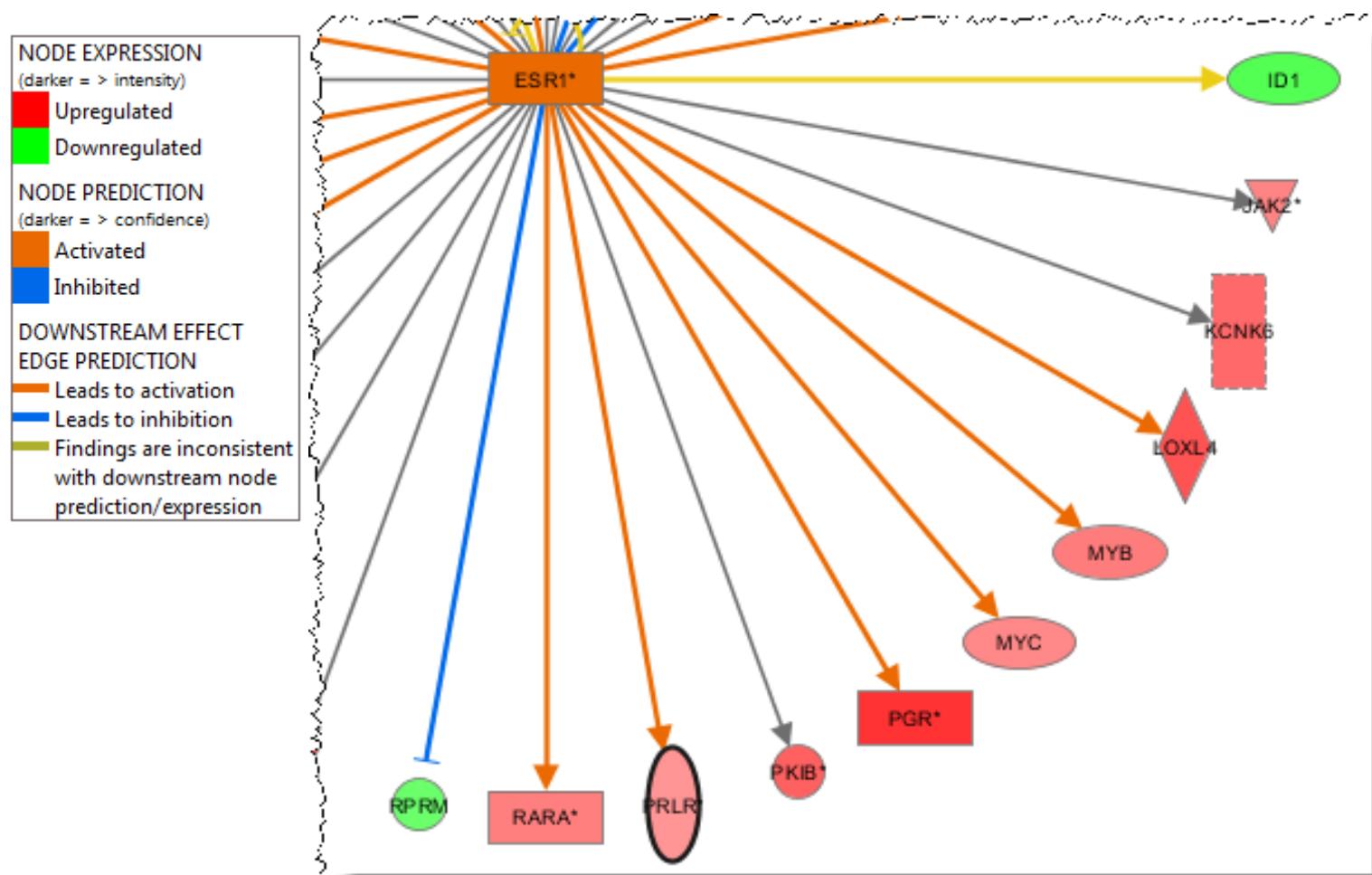
# IPA Upstream Regulator Analysis

- Use experimentally observed relationships (not predicted binding) between regulators and dataset genes to predict upstream transcriptional regulators.
- Calculate z-score to predict activation or inhibition of regulators based on relationships with dataset genes and direction of change of dataset genes.



# IPA Upstream Regulator Analysis

Directional Effects: Molecule Activity Predictor  
Examine Expression Relationship Consistency



# IPA Upstream Regulator Analysis

Hr12FC

Summary | Functions | Canonical Pathways | **Upstream Analysis** | Networks | Molecules | Lists | My Pathways |

Upstream Regulators | Causal Networks |

**CUSTOMIZE TABLE**

Upstream Regulator	Fold Change	Molecule Type	Predicted Activation	Activation z-score	p-value of over...	Target molecules in ...	Mechanistic Net...
beta-estradiol		chemical - endogenous	Activated	6.097	1.24E-26	↓ABCA1, ↓...all 122	186 (17)
raloxifene		chemical drug		-0.751	1.83E-14	↑AREG/AREGB...all 28	125 (15)
ESR1	↓-1.708	ligand-dependent nu...	Activated	3.504	2.84E-13	↓ABC5, ↑...all 37	186 (20)
trichostatin A		chemical drug		-0.620	1.02E-11	↓ABCA1, ↑...all 45	189 (20)
valproic acid		chemical drug		-1.593	5.88E-11	↓ABCG1, ↓...all 33	193 (18)
fulvestrant		chemical drug	Inhibited	-2.826	2.72E-10	↑ADRB1, ↑...all 27	181 (22)
TGFB1		growth factor		-0.325	4.02E-10	↓ABCA1, ↑...all 86	187 (20)
RAF1		kinase		-0.321	6.96E-10	↑AREG/AREGB...all 25	145 (19)
ESR2		ligand-dependent nu...		0.095	9.98E-09		
MYC	↑1.855	transcription regulator	Activated	2.599	1.02E-08		
CCND1	↑1.371	other		0.777	1.28E-08		
ERBB2	↓-1.822	kinase		0.591	1.68E-08		
TNF		cytokine		-0.134	2.00E-08		
dexamethasone		chemical drug		-0.930	2.64E-08		
Salmonella enterica s		chemical toxicant		1.149	5.11E-08		
ZNF217	↓-1.315	transcription regulator		0.555	7.38E-08		
PGR	↑5.528	ligand-dependent nu...		1.879	8.41E-08		
tretinoin		chemical - endogenous	Inhibited	-2.611	1.02E-07		
methylselenic acid		chemical reagent		0.152	1.16E-07		

Selected/Total upstream regulators : 0/709

**Molecule Types**

- Unfiltered
- Transcription Factors
- miRNA
- Drugs and Chemicals
- Select from list below

Select all

- biologic drug
- chemical - endogenous mammalian
- chemical - endogenous non-mammalian
- chemical - kinase inhibitor
- chemical - other
- chemical - protease inhibitor
- chemical drug

**Apply** **Cancel**

- Quickly filter by molecule type
- Filter by biological context
- Generate regulators-targets Network to identify key relationships

- Entities with positive z-scores are known to elicit the same gene expression changes as seen in your data
  - Entities you might want to knock-down to inhibit effects of experiment
- Entities with negative z-scores are known to elicit the opposite gene expression when active
  - Entities you could add to an experiment to counter effects of experiment
- Contradictions between z-score direction prediction and measured gene expression could be the result of
  - A discrepancy between protein activity and expression level
  - Lag time between change in gene expression and effect of that expression
- A regulator with significant z-score but poor p-value could represent a situation where only a few downstream genes in your experimental condition correlate in expression, but many other genes may be expressed in other conditions (or is junk).
- A regulator with insignificant z-score and significant p-value could represent a situation where the genes in your data are downstream of the regulator, but their expression pattern is unique to your experimental condition (or is junk).

# IPA Upstream Regulator Analysis

Summary | Functions | Canonical Pathways | **Upstream Analysis** | Networks | Molecules | Lists | My Pathways |

Upstream Regulators | Causal Networks |

**ADD TO MY PATHWAY** **ADD TO MY LIST** **CUSTOMIZE TABLE** **DISPLAY AS NETWORK** **MECHANISTIC NETWORKS**

How might these upstream regulators interact?

	Upstream Regulator	Fold Change	Molecule Type	Predicted Activation	Activation z-score	p-value of overlap	Target molecules in...	Mechanistic Net...
	beta-estradiol		chemical - endogenous group	Activated	6.097	1.24E-26	↓ABCA1, ↓... all 122	186 (17)
	Mek		chemical drug	Activated	3.683	7.37E-07	↑ABCE1, ↑... all 16	116 (13)
	estrogen		ligand-dependent nuclear receptor	Activated	3.661	2.00E-04	↓ABCA1, ↑... all 19	160 (18)
	ESR1	↓-1.708	cytokine	Activated	3.504	2.84E-13	↓ABCC5, ↑... all 37	186 (20)
	IL3		transcription regulator	Activated	3.190	1.74E-02	↑ADA, ↑AR... all 16	
	MYC	↑1.855					↑ABCE1, ↓... all 52	157 (15)
	LEP						↑ADM, ↓ESR1... all 13	
	Cg						↑ADM, ↑A... all 21	215 (19)
	Vegf						↑ADM, ↑C... all 27	141 (19)
	PBK (complex)						↑ADM, ↑... all 14	
	FSH		complex	Activated	2.291	8.10E-07	↑ADM, ↑A... all 29	202 (22)
	F2		peptidase	Activated	2.287	1.50E-03	↑B4GALT1, ↑... all 16	163 (16)
	NFkB (complex)		complex	Activated	2.258	2.02E-02	↓ABCG1, ↑... all 24	
	Immunoglobulin		complex	Activated	2.236	7.32E-02	↓ADM, ↑B... all 10	
	lithium chloride		chemical drug	Activated	2.213	4.22E-04	↑BCL2, ↑CDC6... all 9	147 (17)
	ERK		group	Activated	2.200	1.05E-02	↑AREG/AREGB... all 12	
	MYB	↑2.039	transcription regulator	Activated	2.199	6.99E-02	↑BCL2, ↓CD... all 5	
	NFKBIA	↓-1.204	transcription regulator	Activated	2.183	7.24E-02	↑ATP11A, ↑... all 17	
	CSF1	↓-1.195	cytokine	Activated	2.154	4.52E-01	↑BCL2, ↑EGR3 ... all 5	

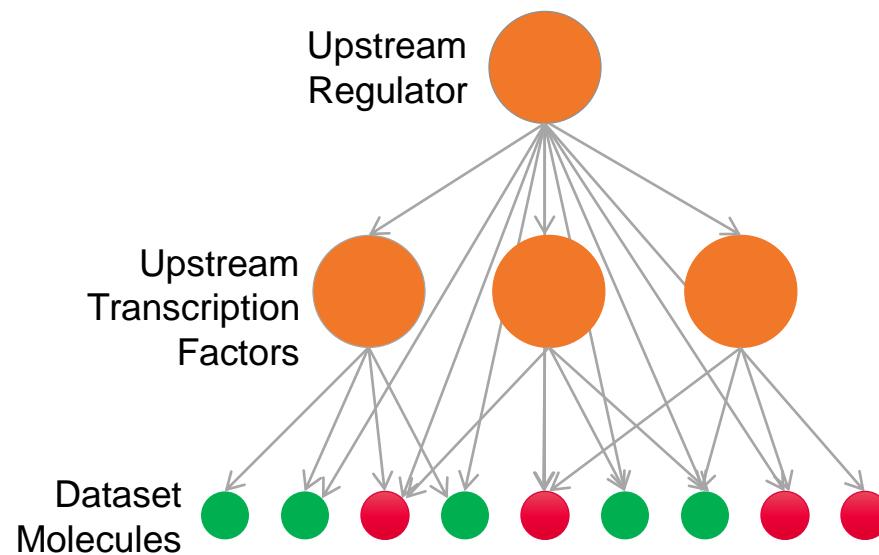
Selected/Total upstream regulators : 0/709

- Quickly filter by molecule type
- Filter by biological context
- Generate regulators-targets Network to identify key relationships

## IPA Mechanistic Networks

Goal: To discover plausible sets of connected upstream regulators that can work together to elicit the gene expression changes observed in a dataset

How: Take IPA Upstream Regulator results and computationally seek pairs of regulators predicted to affect the expression of a similar set of genes. Repeat to build a network:

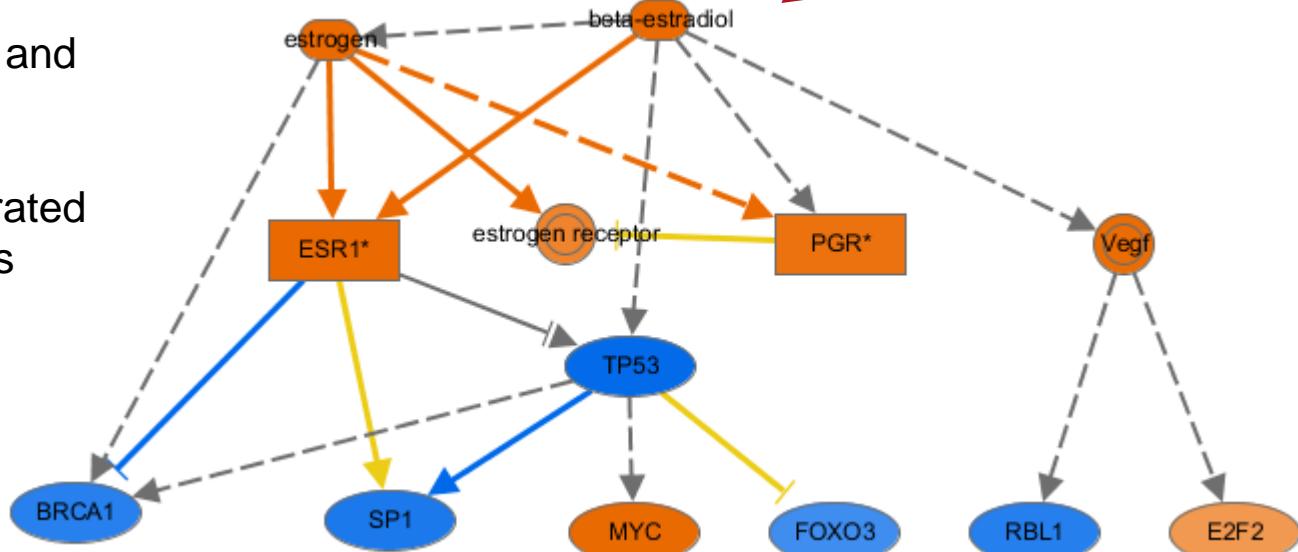


How might the upstream molecule drive the observed expression changes?

Upstream Analysis								
Upstream Regulators		Causal Networks						
<input type="checkbox"/> Upstream Regulator	Fold Change	Molecule Type	Predicted Activation	Activation z-score	p-value of overlap	Target molecules in	Mechanistic Net...	
<input type="checkbox"/> beta-estradiol		chemical - endogenous group	Activated	6.097	1.24E-26	↓ABCA1, ↓... all 122	186 (13)	
<input type="checkbox"/> Mek		chemical drug	Activated	3.683	7.37E-07	↑ABCE1, ↑... all 16		
<input type="checkbox"/> estrogen		ligand-dependent nuclear receptor	Activated	3.661	2.00E-04	↓ABCA1, ↑... all 19	129 (13)	
<input type="checkbox"/> ESR1	↓-1.708	cytokine	Activated	3.504	2.84E-13	↓ABCC5, ↑... all 37	183 (13)	
<input type="checkbox"/> IL3			Activated	3.190	1.74E-02	↑ADA, ↑AR... all 16		
<input type="checkbox"/> ARMC			Activated	3.090	1.02E-01	↑ABCA1, ↓... all 19	186 (13)	

Hypothesis generation and visualization

Each hypothesis generated indicates the molecules predicted to be in the signaling cascade



IPA Winter Release 2012

# Mechanistic Networks

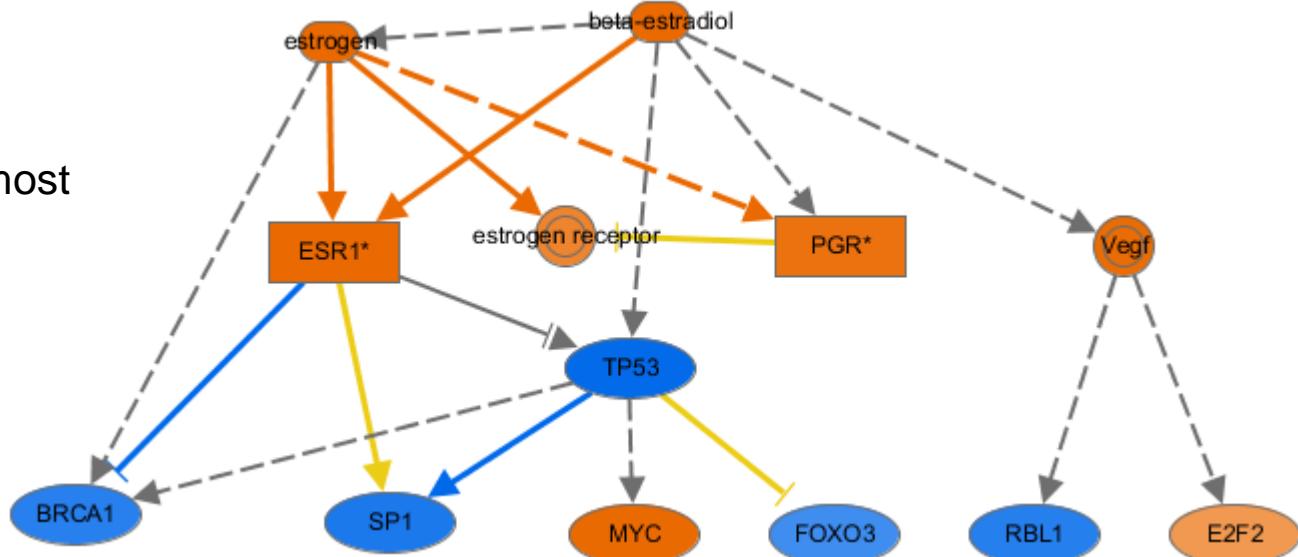
Summary | Functions | Canonical Pathways | Upstream Analysis | Networks | Molecules | Lists | My Pathways |

Upstream Regulators | Causal Networks |

**ADD TO MY PATHWAY** **ADD TO MY LIST** **CUSTOMIZE TABLE** **DISPLAY AS NETWORK** **MECHANISTIC NETWORKS**

Upstream Regulator	Fold Change	Molecule Type	Predicted Activation	Activation z-score	p-value of overlap	Target molecules in network	Mechanistic Net...
beta-estradiol		chemical - endogenous group	Activated	6.097	1.24E-26	↓ABCA1, ↓... all 122	186 (13)
Mek		chemical drug	Activated	3.683	7.37E-07	↑ABCE1, ↑... all 16	
estrogen		ligand-dependent nuclear receptor	Activated	3.661	2.00E-04	↓ABCA1, ↑... all 19	129 (13)
ESR1	↓-1.708	cytokine	Activated	3.504	2.84E-13	↓ABCC5, ↑... all 37	183 (13)
IL3			Activated	3.190	1.74E-02	↑ADA, ↑AR... all 16	

Recommend increasing stringency of criteria in most cases



# Advanced Analytics

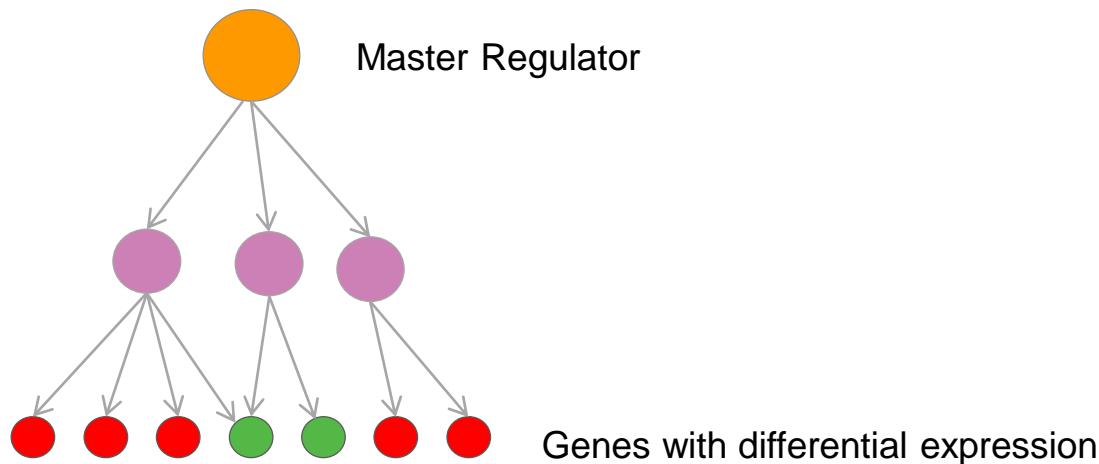
## Causal Networks

Advanced Analytics requires an additional subscription fee

## Advanced Analytics

- Alternate method of predicting upstream regulators based on causal relationships and allowing multiple interaction steps to gene expression changes
- Identify potential novel master-regulators of your gene expression by creating pathways of literature-based relationships
- Expands predictions to include indirect upstream regulators not in mechanistic networks

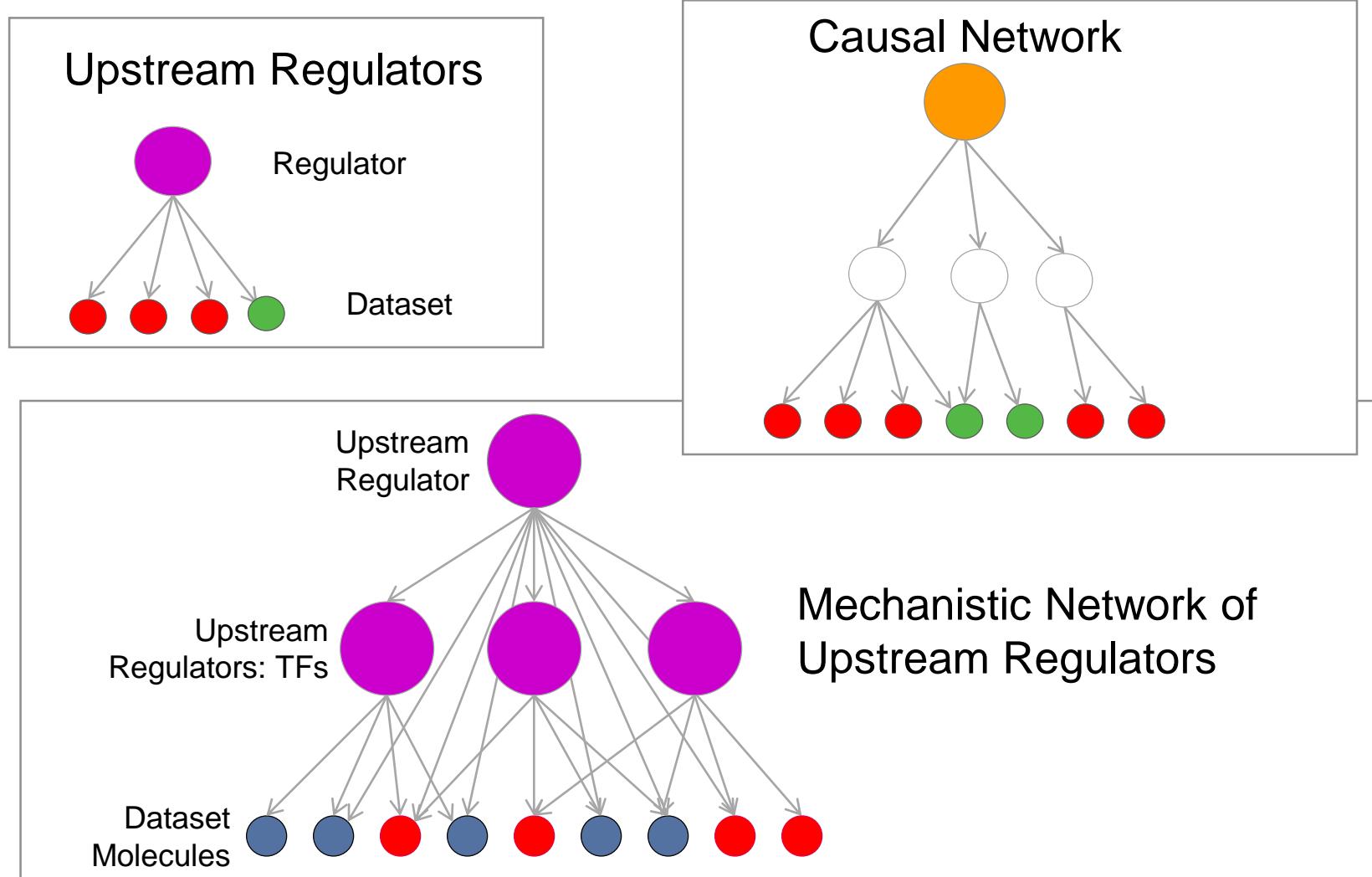
## Causal Networks



# Single- vs. Mechanistic- vs. Causal Networks

## Leveraging the network to create more upstream regulators

### Advanced Analytics: Causal Network Analysis



# Turning on Causal Networks (with Advanced Analytics)

Create Core Analysis - [analysis : E2 of MCF7 P05.xls]

**General Settings**

**Networks Interaction & Caus...**

**Node Types**

**Data Sources All**

**Confidence Experimentally Ob...**

**Species All**

**Tissues & Cell Lines All**

**Mutation All**

**ADVANCED** **SAVE AS DEFAULTS**

**Generate the following Networks (increases analysis time)**

**Interaction networks**

**Include endogenous chemicals** Molecules per network 35 Networks per analysis 25

*Genes are always included*

**Causal networks**

Score master regulators for relationships to diseases, functions, genes, or chemicals (max 50)

Score using causal paths only

**ADD...** **REMOVE**

**Set Cutoffs**

Expression Value Type	Cutoff	Range	Focus On
Exp Fold Change	<input type="text"/>	-22.7434 to 25.1208	Both Up/Downregulated ▾
Exp p-value	<input type="text"/>	0.0 to 0.05	

**RECALCULATE** 9574 analysis-ready molecules across observations

**Preview Dataset E2 of MCF7 P05.xls** Observation: Hr12FC (4532) ▾

Analysis-Ready (4532) ▾ Mapped IDs (13871) ▾ Unmapped IDs (1496) ▾ All IDs (15367) ▾

# Advanced Analytics: Causal Network Analysis

Hr12FC

Summary \ Functions \ Canonical Pathways \ Upstream Analysis \ Networks \ Molecules \ Lists \ My Pathways \

Upstream Regulators Causal Networks

Add to My Pathway Add to My List Print More Info

p-value of over... 2.02E-21 - 2.42E-11 (p1 of 10) <> More Info

Add column(s) Rela... Add column In... D...

Master Regulator	Fold Cha...	Molecule Type	Participati...	Depth	Predicted ...	Activation z-score	p-value...	Net...	Target molec...	Causal network	Targe...	In... D...
SULT1E1		enzyme	beta-estr... all 3	2	Inhibited	-4.951	2.02E-21	1.00E-04	▲ABCAL, ▲AB... all 104	102 (3)	2	an... all 8
beta-estradiol		chemical - endogenous ...	beta-estr... all 1	1	Activated	5.528	2.52E-21	1.00E-04	↓ABCAL, ↓ABCCS ... all 99	99 (1)	1	8-bro... all 8 astres
trans-hydroxytamoxif...		chemical drug	trans-hy... all 1	1		0.928	4.24E-20	1.00E-04	▲ABC5, ↑ARE... all 29	29 (1)	1	
fulvestrant		chemical drug	Akt, AKT1 ... all 21	2		-1.555	2.12E-16	1.00E-04	▲ABCAL, ↑ADA, ... all 93	93 (21)	21	
bisindolylmaleimide I		chemical - kinase inhibitor	ATF2, b... all 44	2	Inhibited	-2.502	3.98E-16	2.00E-04	↑ABC1, ▲AB... all 108	108 (44)	44	
SORBS3		other	AKT1, a... all 49	3		1.362	1.69E-15	1.00E-04	↑ABC1, ↑AB... all 138	138 (49)	47	EGF ... all 1
raloxifene		chemical drug	Akt, AKT1 ... all 15	2		0.000	5.66E-15	1.00E-04	▲ABCAL, ↑ADA, ... all 82	82 (15)	15	
NR1B		ligand-dependent nucleic acid binding protein	acetami... all 23	3		-0.913	6.47E-15	6.00E-04	↓ABCAL, ▲AB... all 120	120 (23)	20	.. 8... all 10 andr.
1,4-bis[2-(3,5-dichlorophenoxy)ethoxy]benzene		chemical toxicant	.. Ahr-a... all 18	3		-0.368	6.78E-15	7.00E-04	▲ABCAL, ▲AB... all 118	118 (18)	18	
CSF1	↓-1.195	cytokine	Akt, AKT1 ... all 37	2		1.616	1.18E-14	5.00E-04	↑ABC1, ▲AB... all 124	124 (37)	37	C3, ... all 12 forsk.
BAG1	↑1.110	other	AR, ↑BAG1... all 9	2		1.890	1.31E-14	1.00E-04	↑ABHD2, ↑ADA, ... all 63	63 (9)	9	IL2, ↑... all 2
RAC3		enzyme	Akt, AKT1 ... all 41	3	Activated	2.514	1.85E-14	3.00E-04	↓ABCAL, ↑AB... all 124	124 (41)	39	
UBE2L3	↓-1.071	enzyme	AR, EGFR, ... all 8	2		1.722	1.95E-14	1.00E-04	↑ABHD2, ↑ARE, ... all 57	57 (8)	7	FSH
MMP11		peptidase	AGT, Akt ... all 63	3	Activated	2.214	2.46E-14	2.00E-04	↓ABCAL, ▲AB... all 160	160 (63)	61	↓FURIN ... all 3 progr.
Ap1		complex	Akt, AKT1 ... all 49	3		0.709	3.00E-14	3.00E-04	↓ABCAL, ▲AB... all 161	161 (49)	49	beta... all 38 AR, ↑
NCOA4		transcription regulator	AHR, AR, ... all 5	2		1.980	3.04E-14	1.00E-04	↑ADA, ↓ADM, ↑... all 50	50 (5)	4	
BAD		other	Akt, AKT1 ... all 45	3	Inhibited	-2.227	3.57E-14	1.00E-04	↓ABCAL, ↑AB... all 147	147 (45)	44	8-(4... all 32 Calc.
dihydrotestosterone		chemical - endogenous ...	Akt, AR, ... all 26	2	Activated	2.636	4.20E-14	5.00E-04	↓ABCAL, ↑AB... all 121	121 (26)	26	finast
androstenedione		chemical - endogenous ...	Akt, AKT1 ... all 33	3		1.281	5.48E-14	7.00E-04	↓ABCAL, ↑AB... all 103	103 (33)	31	8-bro... all 6 ESR
CMA1		peptidase	APP, C... all 52	3		-0.077	6.06E-14	1.50E-03	↓ABCAL, ↑AB... all 167	167 (52)	49	
SMARCE1	↓-1.177	transcription regulator	Akt, AKT1 ... all 29	3		1.400	6.65E-14	5.00E-04	↓ABCAL, ↑AB... all 100	100 (29)	29	
FKBP4	↑2.700	enzyme	AKT1, a... all 36	3		-1.754	9.04E-14	1.70E-03	↑ABC1, ▲AB... all 130	130 (36)	35	Ca2+ ... all 1
CSF2		cytokine	Akt, AKT1 ... all 34	2	Activated	2.307	9.28E-14	2.00E-03	↓ABCAL, ↑AB... all 137	137 (34)	34	CCL21 ... all 9 beta-
Cxcl11		cytokine	Akt, +C... all 16	3	Activated	3.305	1.13E-13	4.00E-04	↓ABCAL, ↑ABC1 ... all 77	77 (16)	13	MAP... all 2 CSF2,
raloxifene		chemical drug	raloxifene ... all 1	1		-0.784	1.19E-13	1.00E-04	↑AREG/AREGB, ↑... all 26	26 (1)	1	
FURIN	↓-1.256	peptidase	ADAM12 ... all 47	3		1.698	1.76E-13	6.00E-04	↓ABCAL, ↑AB... all 153	153 (47)	44	MAPK3 ... all 2 PPP2
SMAD2	↓-1.153	transcription regulator	androgen ... all 53	3		0.811	1.83E-13	9.00E-04	↓ABCAL, ↑AB... all 152	152 (53)	50	ACV... all 20 beta..
Hdac		group	Akt, AKT1 ... all 52	3	Inhibited	-2.280	1.83E-13	6.00E-04	↑ABC1, ▲AB... all 130	130 (52)	51	7S NGF ... all 4 romi.
VEGFA		growth factor	ADAM17 ... all 111	3	Activated	2.154	2.22E-13	6.00E-04	↓ABC5, ↑AB... all 194	194 (111)	94	12-h... all 50 ACV..
TCF7L2	↓-1.576	transcription regulator	Akt, AKT1 ... all 33	3		-1.106	2.42E-13	8.00E-04	↓ABCAL, ↑ABC1 ... all 99	99 (33)	32	LNX2 ... all 1 triam
cardiolipin		chemical - endogenous ...	↓ABL1, ... all 36	3	Activated	2.108	2.75E-13	1.00E-04	↑ABC1, ↓ABC1 ... all 90	90 (36)	30	

Selected/Total rows : 0 / 968

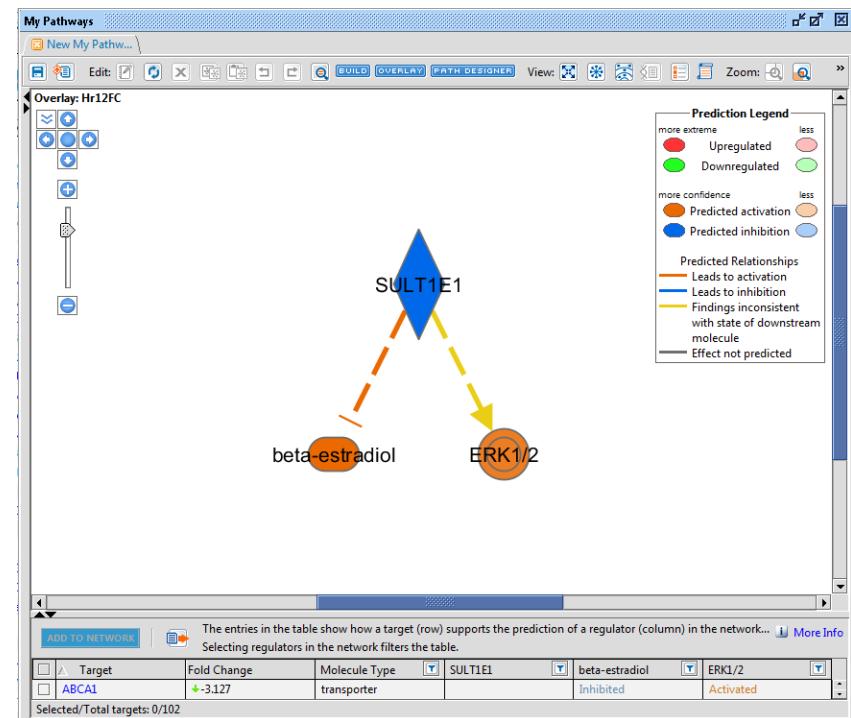
Beta-estradiol of MCF7 cells at 12 hr.  
 SULT1E1 is top master regulator, but does not appear in upstream regulator table



# SULT1E1 Causal Network

## Advanced Analytics: Causal Network Analysis

- SULT1E1 is an enzyme that converts estrone and estradiol to an inactive form
- Causal network predicts the absence, inhibition, or saturation of this enzyme in this experiment where estradiol was added exogenously
- SULT1E1 does not have downstream gene expression relationships and, thus, does not appear in the Upstream Regulator table or Mechanistic Networks
- Hypothesis: increasing SULT1E1 activity can have an anti-estrogen effect



## Parameters

---

### Advanced Analytics: Causal Network Analysis

Only considers edges of unambiguous direction of regulation to downstream genes

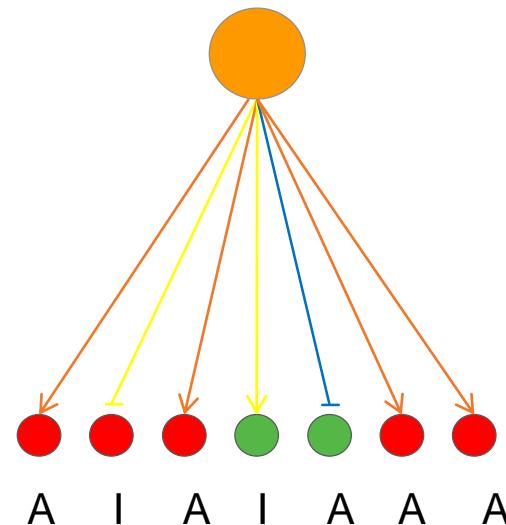
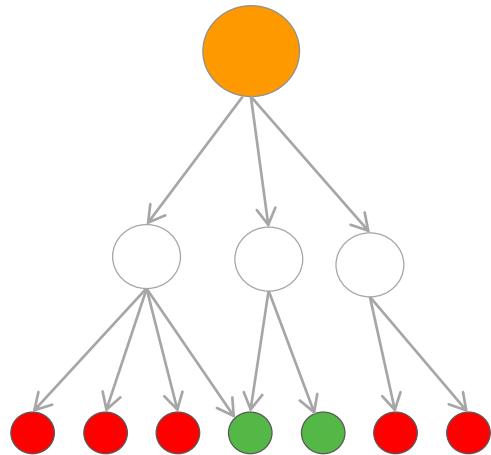
Edges that cannot be assigned a direction of regulation, including all types of binding edges are excluded

- Included relationship types
  - activation (A)
  - inhibition (I)
  - expression (E)
  - transcription (T)
  - group/complex
  - membership edges (MB, considered activating)
  - phosphorylation (P)

Up to 3 interactions edges from root are considered

- Expression/Transcription must be last edge type

## Advanced Analytics: Causal Network Analysis



2 inhibitory edges  
5 activating edges

## Advanced Analytics: Causal Network Analysis

Two p-values are calculated

- Fishers Exact Test of whether there is a greater than expected proportion of downstream data set genes than expected by chance
- Network bias corrected p-value is a measure of how often a more significant result was seen in 10K iterations of selecting random data sets of genes with similar relationship number.

z-score

- Activation z-score is calculated and represents the bias in gene regulation that predicts whether the upstream regulator exists in an activated or inactivated state

$$z = \frac{N^+ - N^-}{\sqrt{N}},$$

z-score represents the number of standard deviations from the mean of a normal distribution of activity edges.

# Analyzing Results

## Diseases & Functions (Downstream Effects)



## Downstream Function/Process Analysis

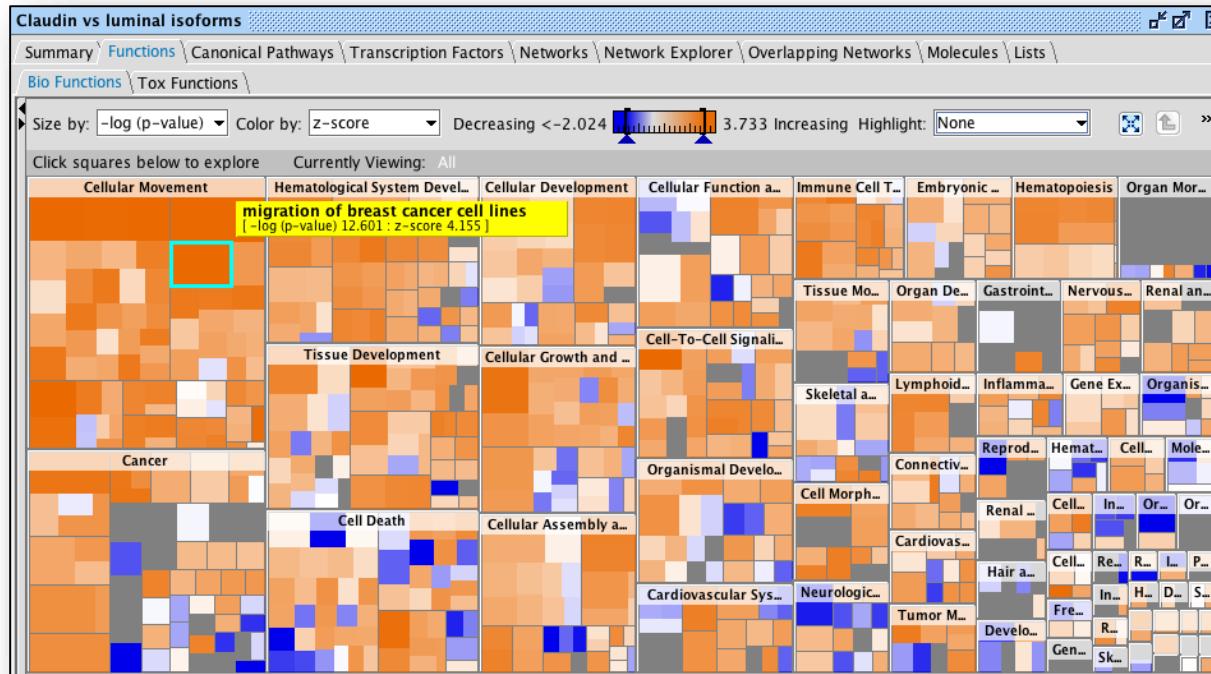
---

- How are cellular processes predicted to be changing based on my gene expression data?
- What genes are driving these directional changes?

## Downstream Effects Analysis

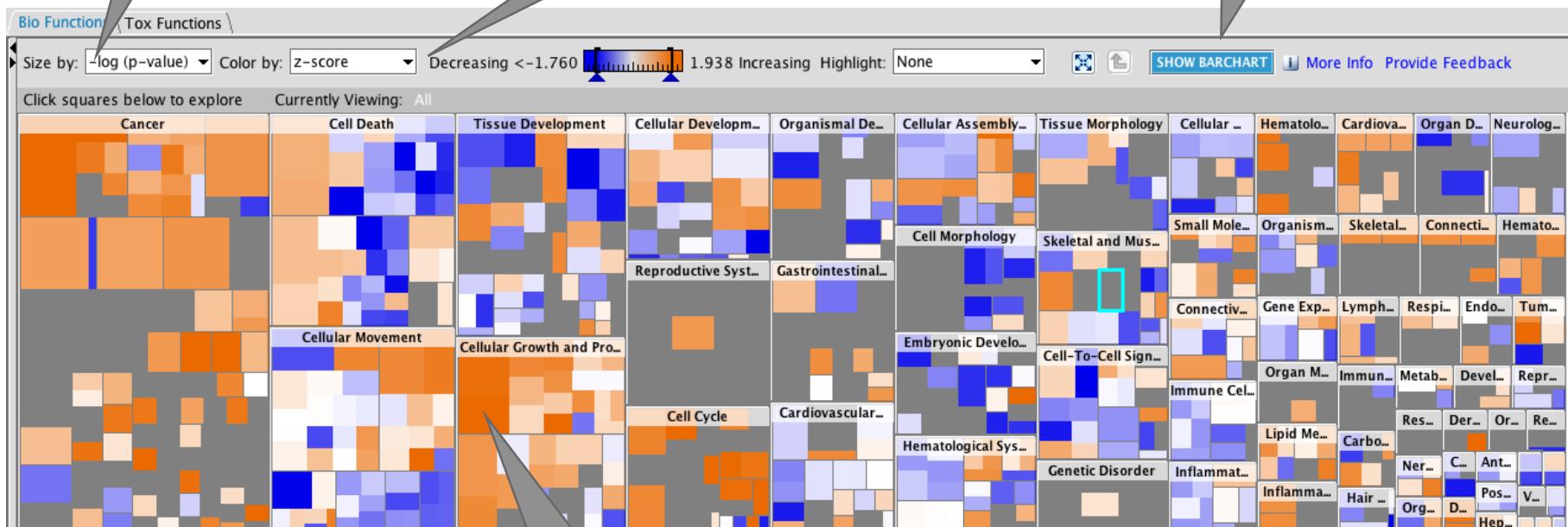
Identify key biological processes influenced by differentially expressed genes

Understand whether cellular processes are being driven up or down by correlating observed expression with reported experimental gene effects

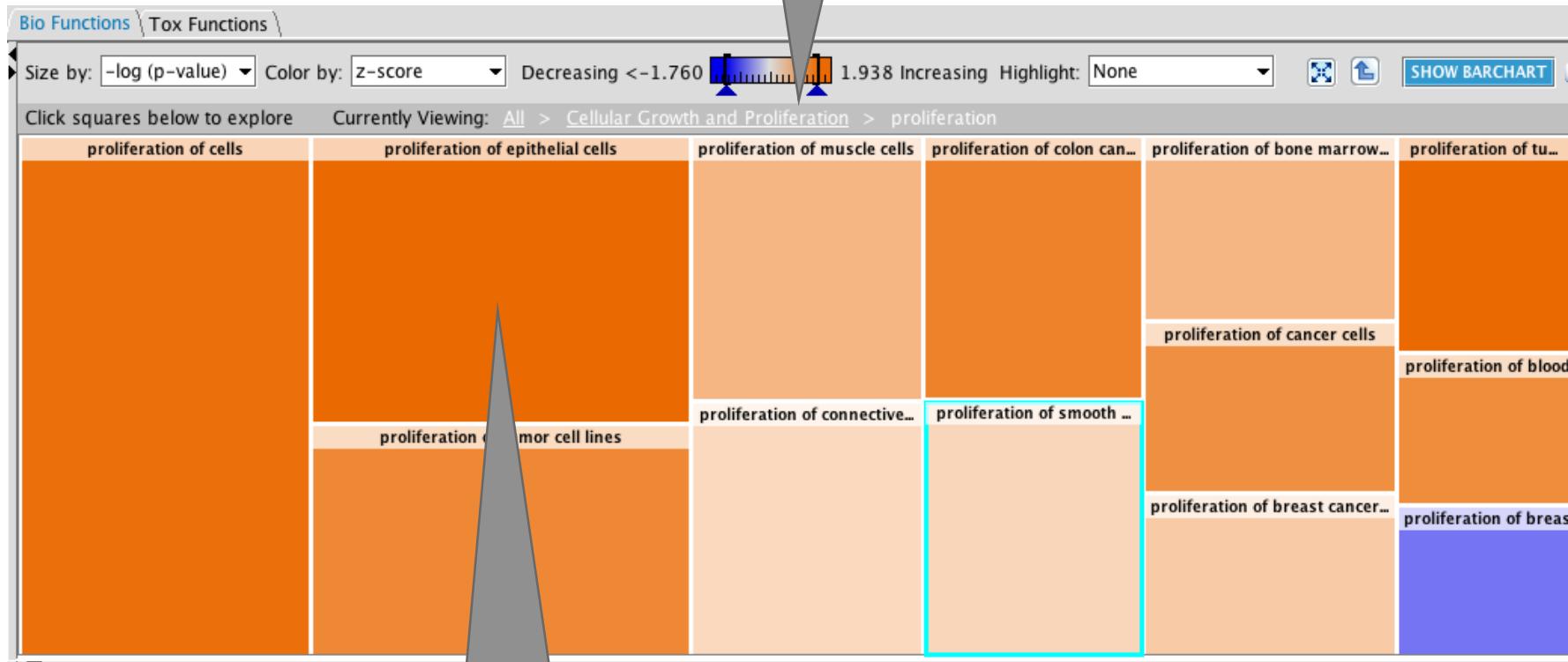


- Each box represents a biological process or disease
- The size of the box represents gene enrichment
- The color of the box indicates the predicted increase or decrease

# Downstream Effect on Bio Function



## Ontology Levels



Click to See the Specific  
Genes and Findings

Functional Category and Statistical Result

Access Findings

Downstream Effects Analysis: Evidence for Effects  
proliferation of epithelial cells(z-score 1.991). Overlap p-value 8.80E-10

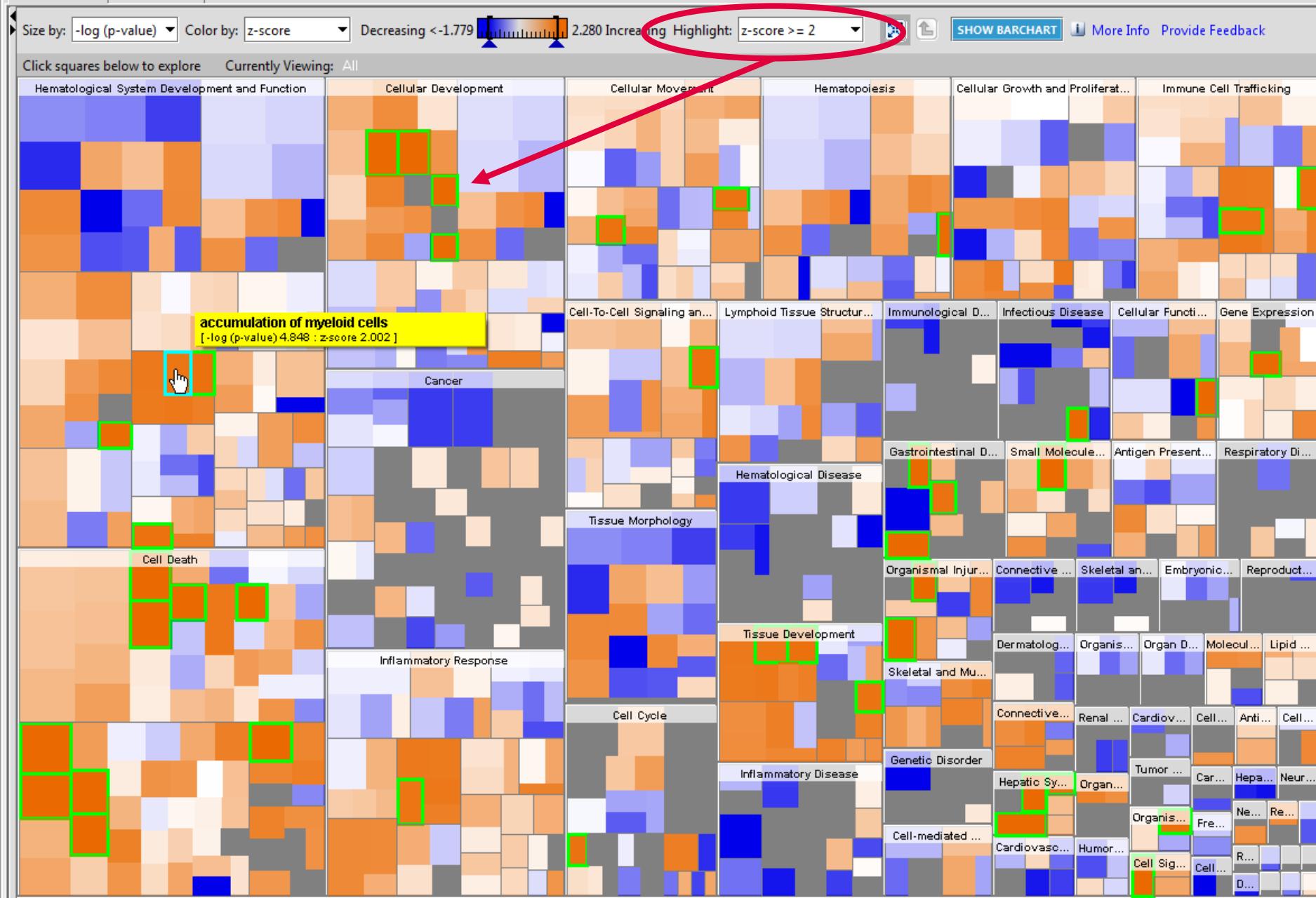
[ADD TO MY PATHWAY](#) [ADD TO MY LIST](#) [CUSTOMIZE TABLE](#) [CREATE DATASET](#) [More Info](#)

ID	Genes in dataset	Prediction (based on expression)	Fold Change	Findings
H62162	HPN	Decreased	↑3.016	Decreases (1)
AA464600	MYC	Increased	↑2.761	Increases (8)
AA030029	PRKCA	Increased	↑2.175	Increases (3)
H24650	LAMC1	Increased	↑1.917	Increases (1)
N71159	MTA1	Increased	↑1.873	Increases (1)
R19956	VEGFA	Increased	↑1.807	Increases (2)
N49517			↑1.787	Affects (1)
H31212	BCL2A1	Decreased	VEGFA: Known to increase proliferation of epithelial cells and is upregulated ↑ in the dataset therefore predicted to increase the function	(1)
AA459263	BCL2A1	Increased	↑1.722	(1)
AA488645	NAB1	Decreased	↑1.576	Decreases (1)
N74882	DLX5	Increased	↑1.572	Increases (1)
H65052	F2	Increased	↑1.564	Increases (2)
W48713	EGFR	Increased	↑1.515	Increases (10)
H84048	RBL1	Increased	↓-1.559	Decreases (5)
AA456439	SMAD4	Increased	↓-1.565	Decreases (4)
N67039	CDK6	Increased	↓-1.570	Decreases (1)
AA487589	METAP2	Decreased	↓-1.570	Increases (1)
AA489752	CCNG2	Increased	↓-1.578	Decreases (1)
N49517	TCER2		↑1.721	Affects (1)

Selected/Total rows : 0/42

Prediction Logic

Expression Value in Your Dataset





# Functional Analysis (FA) Workflow

Goal is understand biology and identify smaller subsets of genes that are of interest

Genes related to a particular function can be :

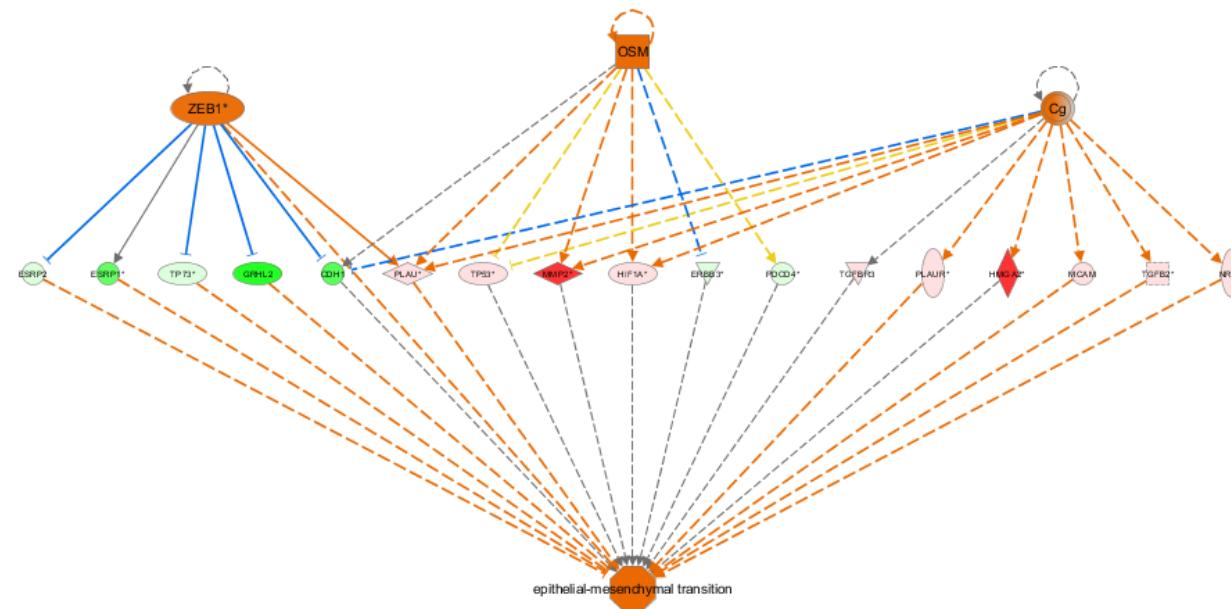
- sent to a pathway for building and/or overlay analysis
- saved as a new Data Set and sent to Core Analysis for additional categorization and segmentation

# Analyzing Results

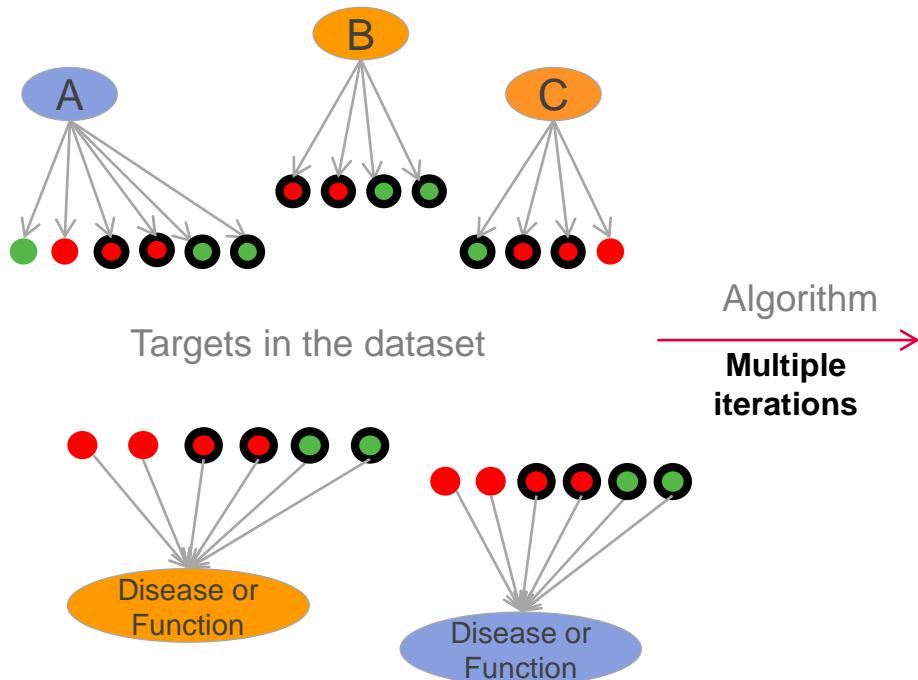
## Regulator Effects

# Regulator Effects

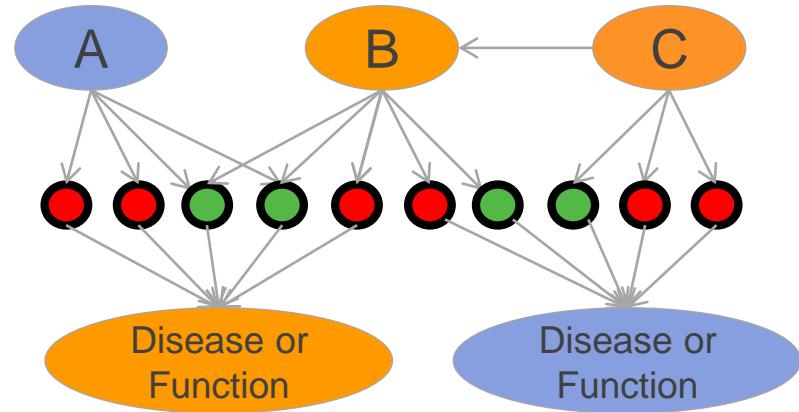
- Hypothesis for how a phenotype, function or disease is regulated in the dataset by activated or inhibited upstream regulators
- Explain impact of upstream molecules on downstream biology
- Explain potential mechanism for a phenotype or drug
- Define drug targets
- Discover novel (or confirm known) regulator → disease/phenotype/function relationships



## Upstream Regulator Analysis



## Functional Network Analysis



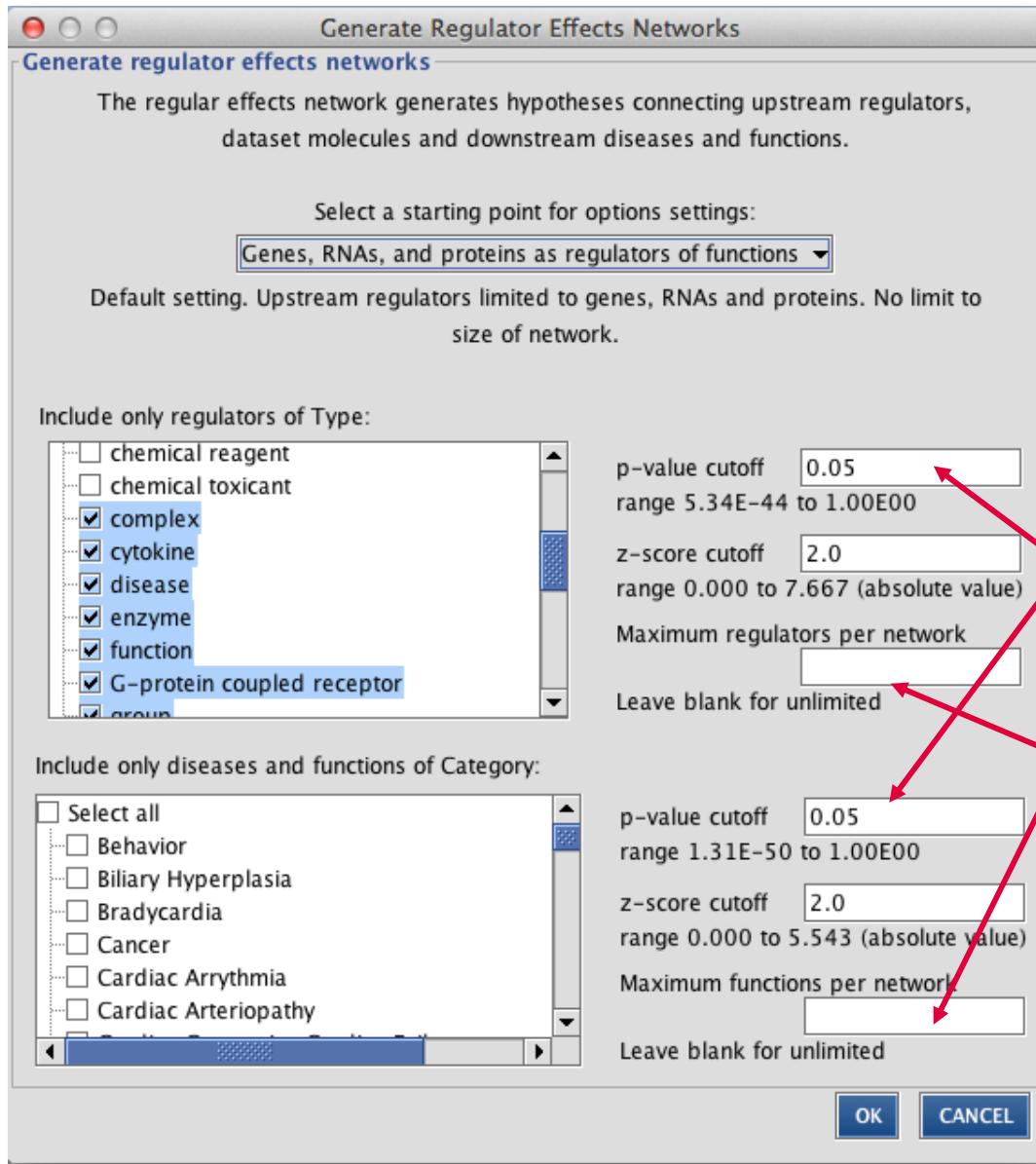
## Downstream Effects Analysis

# Customize Regulator Effects to answer your question

**Default:**  
Genes, RNAs,  
proteins as  
regulators

VS.

Any type of  
disease or  
function

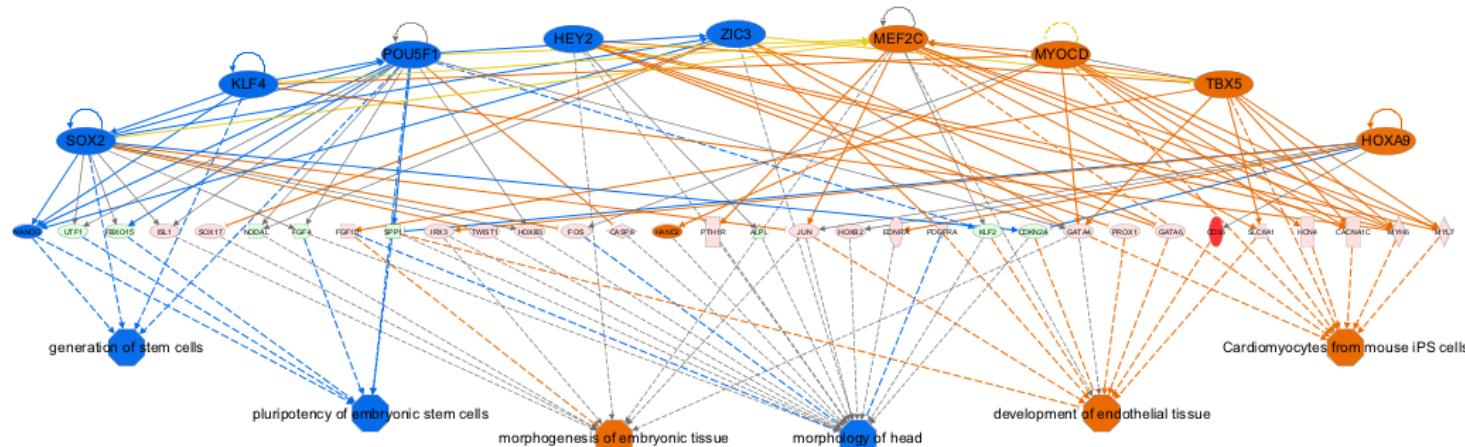


Recommend decreasing p-value cutoffs to 1.0E-3 or lower in most cases.

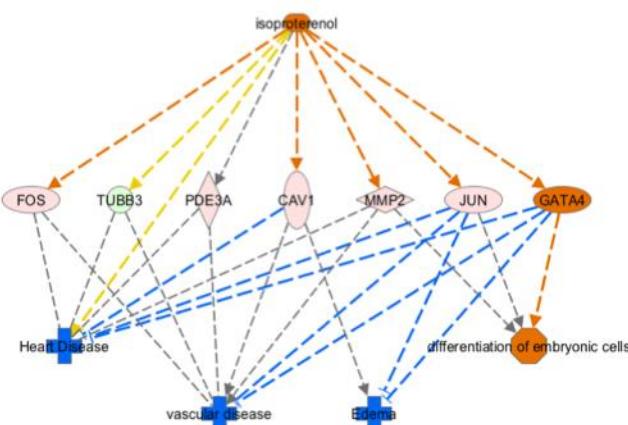
Recommend setting network size to 1 regulator and 1 function in addition to viewing with no value

Alternate presets to answer different questions

# Transcription factors and functions



## Compound as regulator of functions



Minimal regulator to function networks



## Analyzing Results Networks



- To show as many interactions between user-specified molecules in a given dataset and how they might work together at the molecular level
  
- Highly-interconnected networks are likely to represent significant biological function

- Networks are assembled based on gene/molecule connectivity with other gene/molecules.
  - Assumption: the more connected a gene/molecule, the more influence it has and the more “important” it is.
- Networks are assembled using decreasingly connected molecules from your data set.
- Genes/molecules from the Knowledge Base may be added to the network to fill or join areas lacking connectivity.
- A maximum of 35, 70, or 140 genes/molecules can comprise a network based on parameter settings.
- Networks are annotated with high-level functional categories.

# How Networks Are Generated

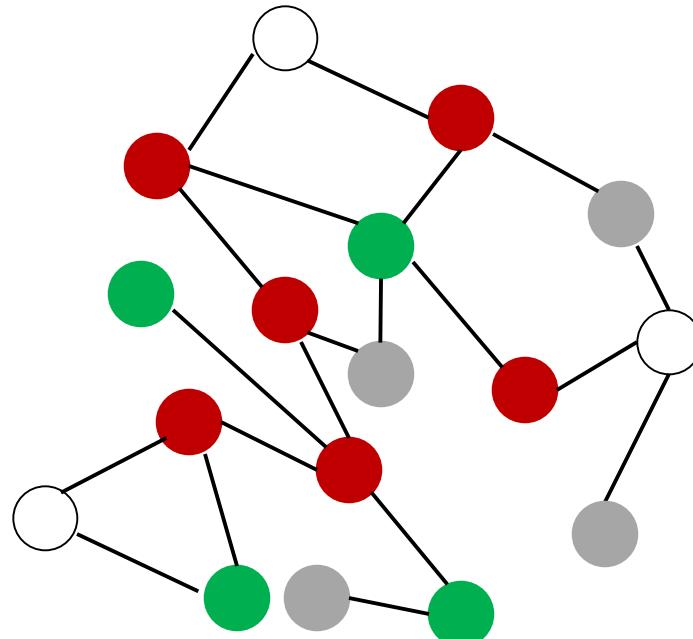
Focus molecules are “seeds”

Focus molecules with the most interactions to other focus molecules are then connected together to form a network

Non-focus molecules from the dataset are then added

Molecules from the Ingenuity's KB are added

Resulting Networks are scored and then sorted based on the score



# Using the information in Networks

Keep in mind..

- Networks may contain smaller networks of connectivity related to specific functions. It might make sense to subset a network. (What does this mean? Just focus on subportions of the network?)
- Larger cellular networks may span IPA assembled networks. Merging networks may allow you to visualize these larger networks.
- Networks should be treated as “starter pathways” that you then modify based on your biological understanding of the system and the questions that you want to answer. Use the pathway building ('Build' button) and Overlay tools to expand on your initial results.

# Getting Help

**support@ingenuity.com**

**support-ingenuity@qiagen.com**

+1 650 381-5111  
6am-5pm Pacific Time (M-F)

**QIAGEN Redwood City/Silicon Valley**

1700 Seaport Blvd., 3rd Floor  
Redwood City  
CA 94063, USA



**INGENUITY®  
PATHWAY ANALYSIS**

**INGENUITY®  
iREPORT**

**INGENUITY®  
VARIANT ANALYSIS**



For Help and Technical Support contact our Customer Support team by email to [support@ingenuity.com](mailto:support@ingenuity.com), or by phone to +1 650-381-5111

For getting started tutorials and training videos see the 'Tutorials' link on the help menu within IPA

To see case studies, application notes, and white papers visit  
[www.ingenuity.com/library](http://www.ingenuity.com/library)

To view our future scientific seminars, and to watch the series archive visit  
[www.ingenuity.com/science/scientific-seminar-series.html](http://www.ingenuity.com/science/scientific-seminar-series.html)

To see how IPA has been used and cited in over 9000 publications visit  
[www.ingenuity.com/science/search-pub.html](http://www.ingenuity.com/science/search-pub.html)

## IPA search and explore series videos:

- The Ingenuity Knowledge Base for IPA <http://youtu.be/4lFxsfMkpQg>
- Searching and accessing the Knowledge Base <http://youtu.be/iU9ihqzfeEY>
- Building a pathway: Filtering and growing <http://youtu.be/8rYEs8F0Cws>
- Building a pathway: Exploring the path of interaction <http://youtu.be/-TRmuMVP9E>
- Overlay contextual information  
<http://www.youtube.com/watch?v=rSp8X6Y6WIc>
- Editing a pathway for publication <http://youtu.be/yEJjqIUM4So>

## IPA data analysis series videos:

- Data analysis : Part 1 (Data upload) <http://youtu.be/XrdMN9eGWjg>
- Data analysis : Part 2 (Results interpretation) [http://youtu.be/PfF\\_Ru73-1o](http://youtu.be/PfF_Ru73-1o)
- Comparison analyses <http://youtu.be/JCanWpyfvQE>
- Analysis results <http://youtu.be/rrppI9OGPUY>
- Statistical calculation <http://www.youtube.com/watch?v=0xCQ9dOQIE>
- Canonical pathways <http://youtu.be/6iZdD9OjII0>
- Network Analysis <http://youtu.be/eReZrNE2bWY>
- Downstream effects analysis <http://youtu.be/CYMrhwuvVKs>
- Upstream regulator analysis <http://www.youtube.com/watch?v=X2bStYNJXm4>
- Human isoforms <http://youtu.be/Po07vk3pOVE>
- Molecular toxicology [http://youtu.be/m1nYDFdY\\_Zg](http://youtu.be/m1nYDFdY_Zg)
- Biomarker filter and comparison analysis <http://youtu.be/XQFUy0s6wCU>
- MicroRNA target filter <http://www.youtube.com/embed/06xoKQL9-KA>

<http://ingenuity.force.com/ipa/articles/Tutorial/Tutorials>

- Search for genes tutorial
- Analysis results tutorial
- Upload and analyze example data tutorial
- Upload and analyze your own expression data tutorial
- Visualize connections among genes tutorial
- Learn about specialized features
- Human isoforms view tutorial
- Transcription factor analysis tutorial
- Downstream effects analysis tutorial

# Comparing Core Analyses

# Comparison Analysis

- Multiple Comparison
  - Time course
  - Does response
- Multiple Platforms and Data Integration
  - Systems biology
  - Combining SNP, CNA, mRNA, microRNA, proteomics, etc.
- Analysis Comparisons work best with Canonical Pathways, Upstream Regulators, and Disease and Functions
- Regulator Effects and Mechanistic Networks are similarly difficult to compare because these networks are created in the context of the single analysis.
  - To compare these networks across analyses, open, view the network, and then use the OVERLAY -> “Data Sets, Analyses, and Lists” to overlay colored representation of gene changes.



# Comparing Analyses

IPA

File Edit View Window Help

Provide Feedback | Support

New

Open

Save Ctrl-S

Save As...

Upload Dataset... Ctrl-U

Batch Upload Datasets...

Search Datasets and Analyses...

Refresh Project Manager F5

View References

Export Data... Ctrl-E

Export Image... Ctrl+Shift-E

Send To

Share

Properties

Preferences

Print... Ctrl-P

Close IPA Ctrl-Q

Influenza Mouse Lung 2013\_07

Welding Gas Toxicity

Melanoma 2013-06-13

Diabetes and Rosi

E2 Treatment, MCF7 2013\_05

Summer 2013

Diabetes and Rosiglitazone

Influenza 2013\_04

Tox Lists

SEARCH Advanced Search

Core Analysis... Ctrl-N

Tox Analysis... Ctrl+Shift-T

Biomarker Filter... Ctrl+Shift-B

Metabolomics Analysis... Ctrl+Shift-A

Core, Tox or Metabolomics Comparison Analysis... Ctrl+Shift-C

Biomarker Comparison Analysis... Ctrl+Shift-K

microRNA Target Filter... Ctrl+Shift-I

My Pathway

Path Designer

Advanced Search

Project...

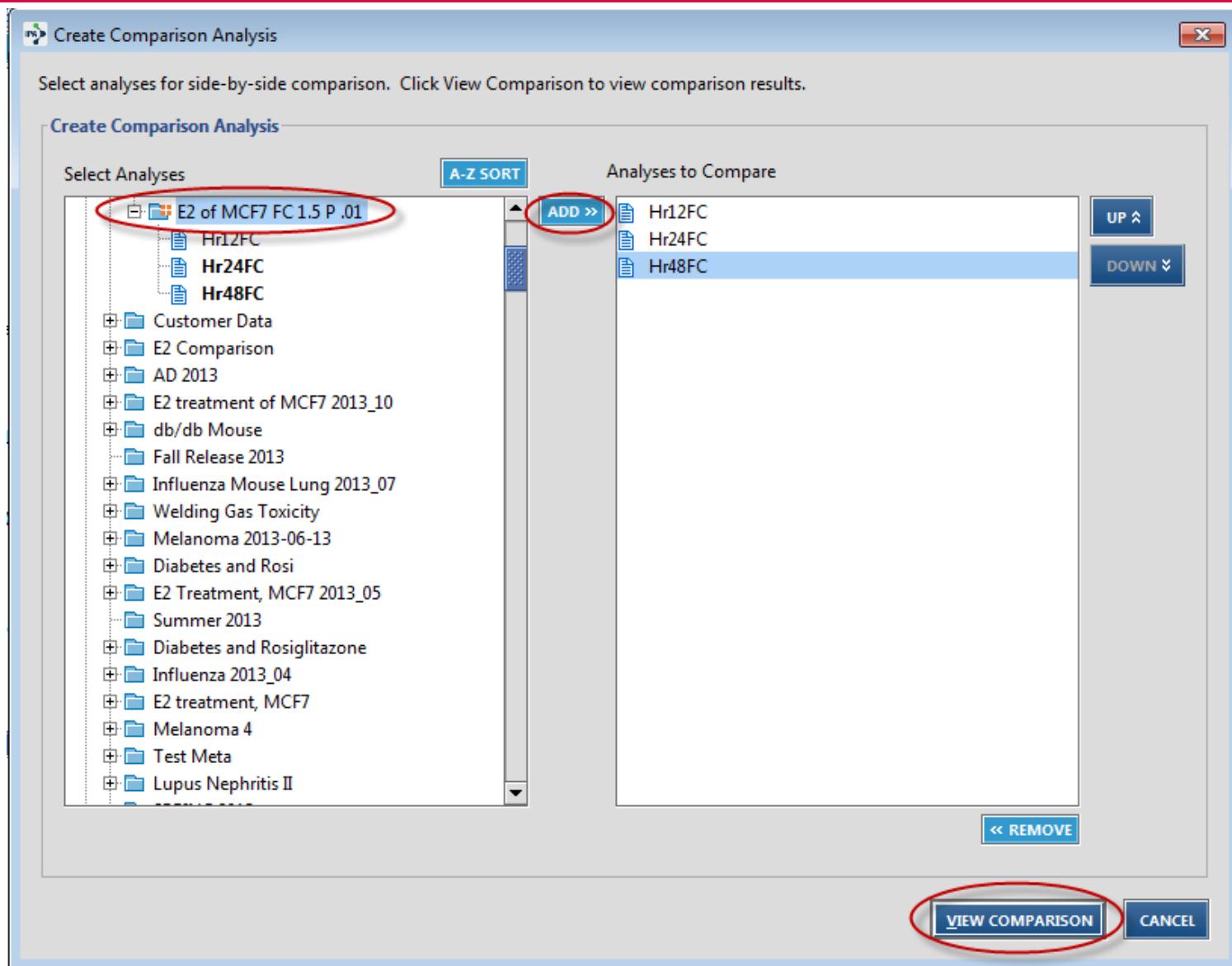
Compare...

Filtered Dataset... Ctrl-D

Import Pathway

A screenshot of the IPA software interface. The window title is 'IPA'. The 'File' menu is open, showing various options like 'New', 'Open', and 'Save'. A sub-menu is displayed under 'File' with many items, including 'Core Analysis...', 'Tox Analysis...', and 'Core, Tox or Metabolomics Comparison Analysis...'. The 'Core, Tox or Metabolomics Comparison Analysis...' option is highlighted with a yellow starburst cursor. On the right side of the interface, there's a search bar with 'Tox Lists' and a 'SEARCH' button, along with a link to 'Advanced Search'. At the bottom left, there's a tree view of project files, and at the bottom right, a footer with 'Sample to Insight' and the URL 'www.ingenuity.com'.

# Selecting Analyses to Compare

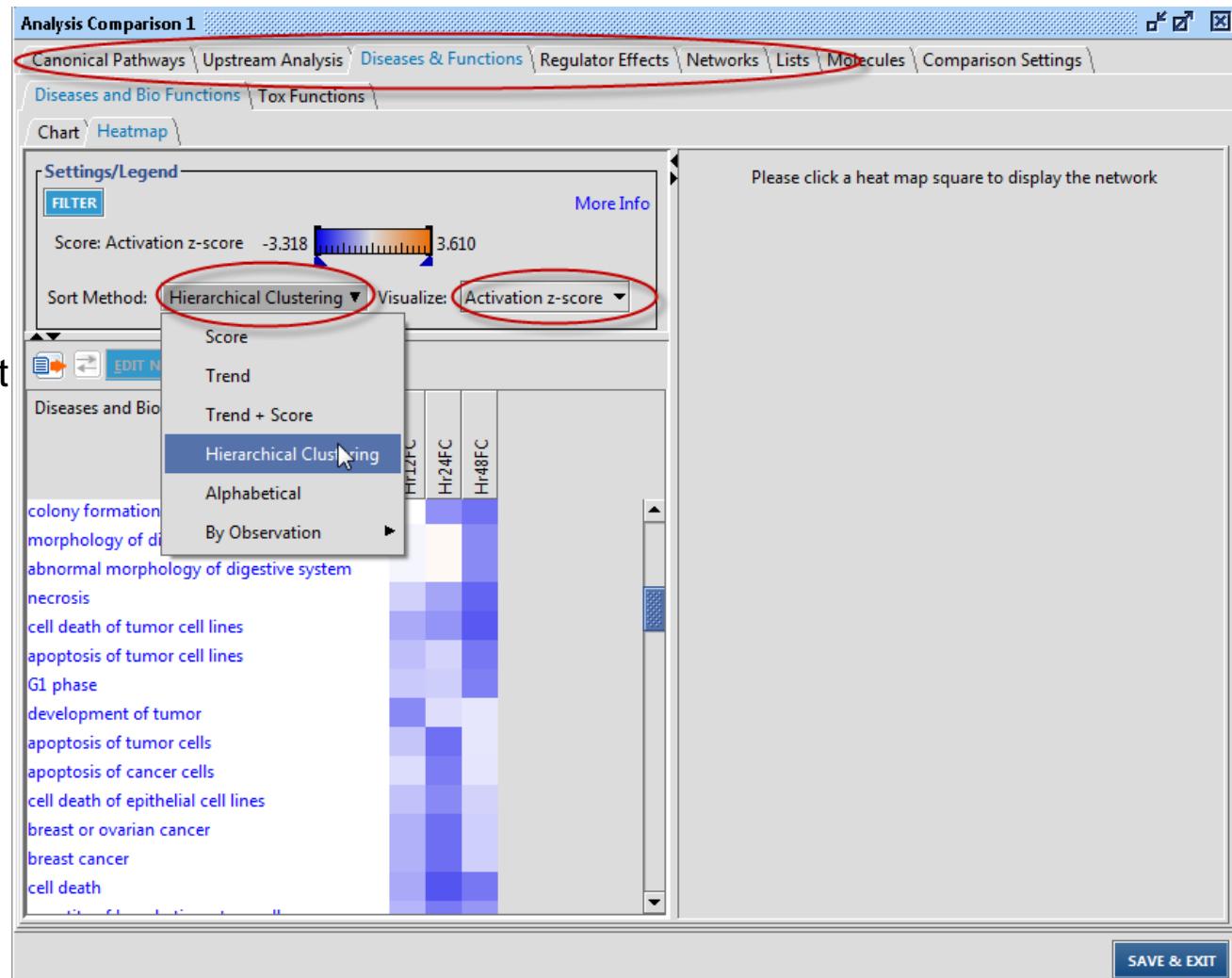


# Analyzing Comparison Results

Tabs at top navigate to the analysis-type of interest

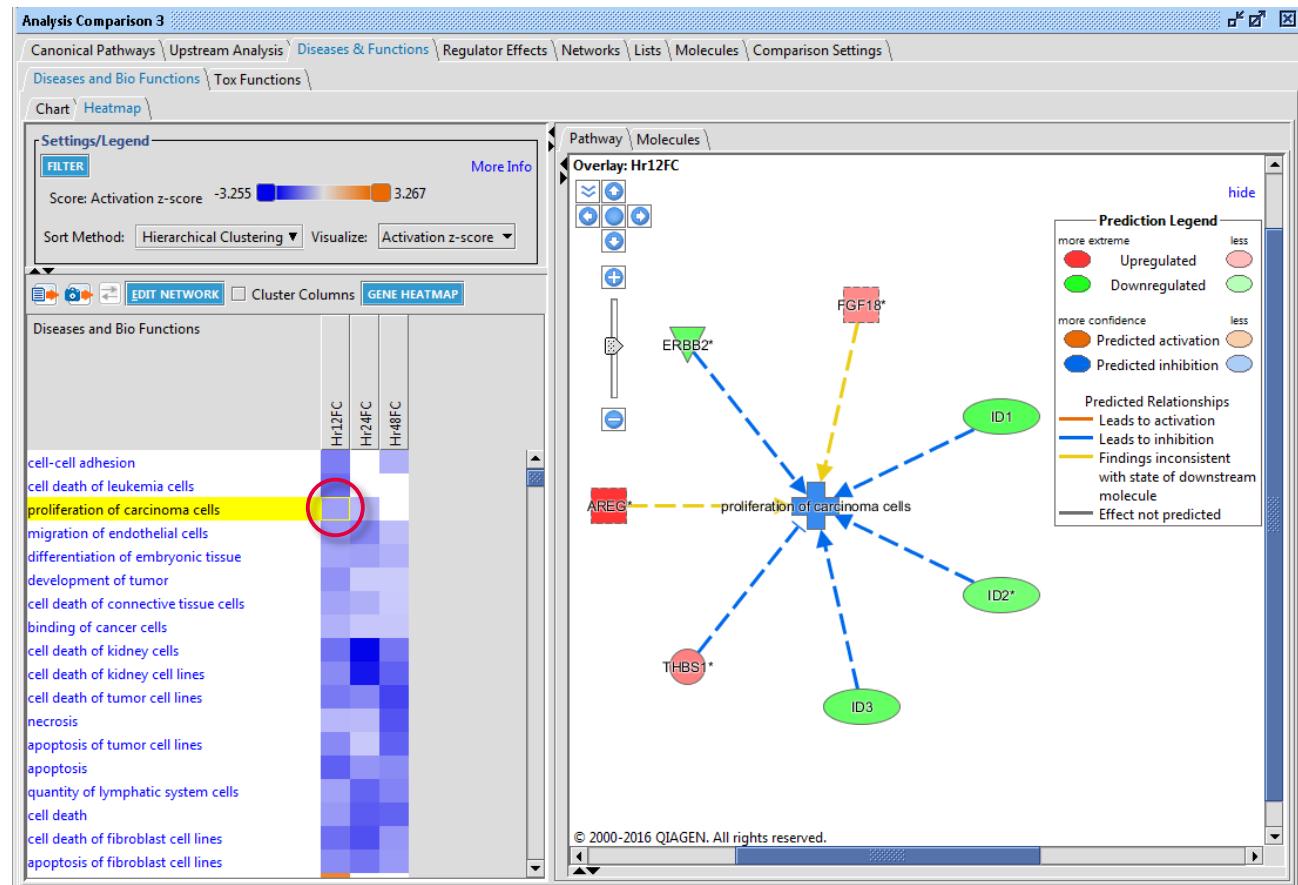
Heatmap can be generated using different calculation methods

Heatmap can be based on different metrics depending on analysis-type.



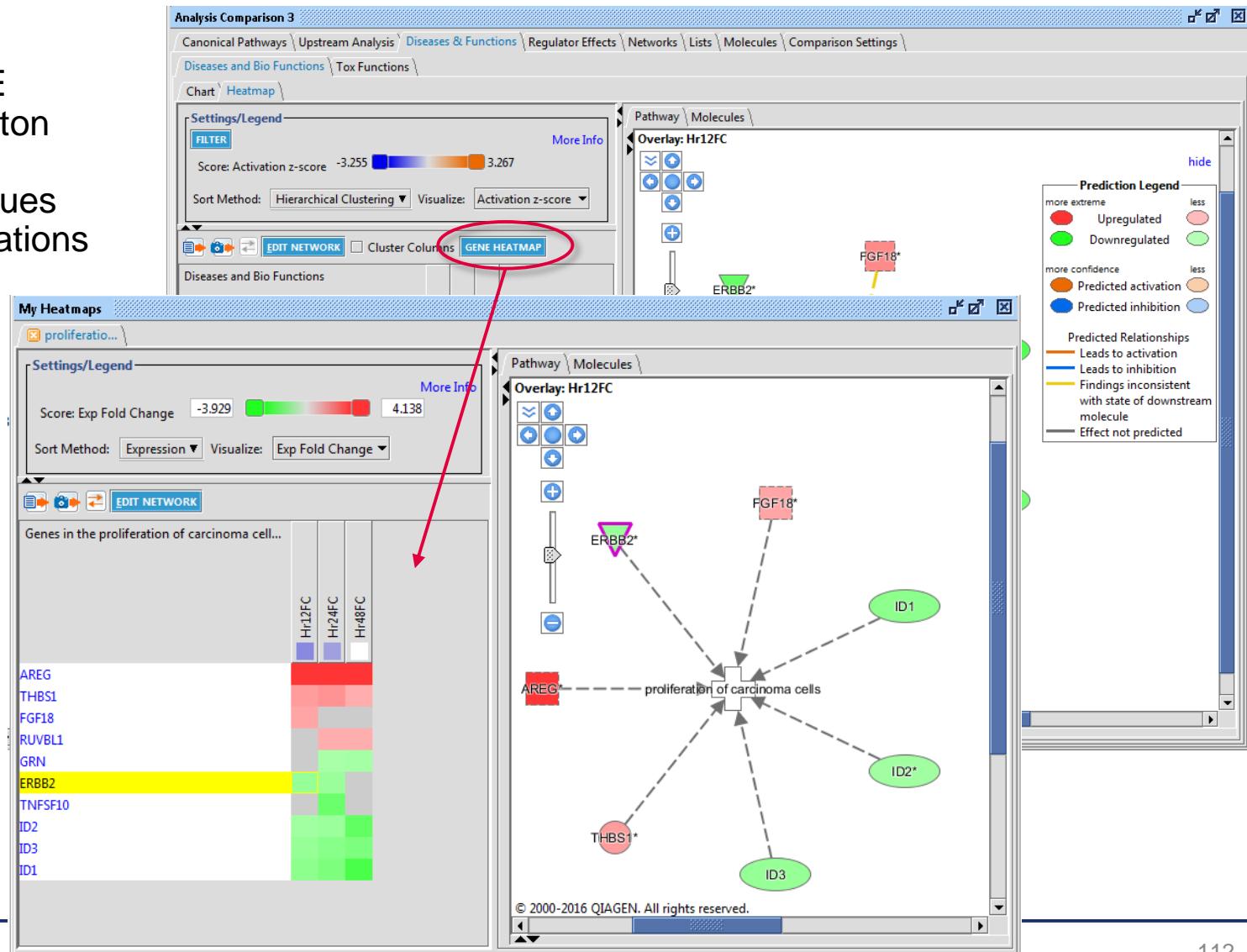
# Analyzing Comparison Results

Selecting heatmap element displays pathway or network with data-values overlay and MAP coloring (if applicable)



# Analyzing Comparison Results

Clicking GENE HEATMAP button displays gene expression values across observations



# Micro-RNA Target Filter

# Workflow for miRNA Target Filter

Filter miRNA differential expression data set (if corresponding mRNA differential expression data, filter as well)

- File -> New -> Filtered Data Set

Start microRNA Target Filter

- File -> New -> miRNA Target Filter
- Open miRNA filtered data set

Using funnel in column headers, filter mapping based on information type/confidence

Add annotation columns, if desired, by clicking the plus sign in column header and filter as desired

# Workflow for miRNA Target Filter

If corresponding mRNA, click “ADD/REPLACE MRNA DATA SET” to filter mRNA mappings to genes in the mRNA expression data set

- Click “EXPRESSION PAIRING” to pair the expression between the miRNA and mRNA
- Click the funnel in the column header of the expression pairing column to filter for the miRNA-mRNA pairing desired

Click to summary tab to view a summary of miRNA-mRNA mappings

For further analysis, select one or more miRNAs from the summary tab and add the miRNA and targets to a new pathway and perform overlays for interpretation of functions, pathways, drug targets, etc.

## microRNA Target Filter

68 microRNA families have targeting information available.  
Filtered to 51 microRNAs targeting 32 mRNAs.

ADD/REPLACE mRNA DATASET

EXPRESSION PAIRING ↑↓

Details | Summary

ADD TO MY PATHWAY | ADD TO MY LIST |

Rows: 1 - 131 |

Use to filter a column. Add data or more columns using 'Add column(s) '.  
Use to refresh the page.

<input checked="" type="checkbox"/>	microRNA dataset: melanoma_microRNA_data			Relationship			Add column(s)			mRNA dataset: mRNA Metastasis vs Normal - 2FC,0.05PV					Add column(s)	
	ID	Symbol	metastatic melanoma (Fold C...)	Source	Confidence	Expression Pairing		ID	Symbol	Fold Change	Molecular Type	Pathway				
<input checked="" type="checkbox"/>	hsa-let-7c	let-7	↓ -3.120	TargetScan Human	High (predicted)	↓↑		8072015	ADRBK2	↑ 3.394	kinase	Colorectal Cancer Met...				
<input checked="" type="checkbox"/>	hsa-let-7c	let-7	↓ -3.120	TargetScan Human	Moderate (predicted)	↓↑		8067167	AURKA	↑ 2.136	kinase	Molecular Mechanisms...				
<input checked="" type="checkbox"/>	hsa-let-7c	let-7	↓ -3.120	TargetScan Human	High (predicted)	↓↑		8105121	GHR	↑ 2.052	transmembrane receptor	Growth Hormone Signa...				
<input checked="" type="checkbox"/>	hsa-let-7c	let-7	↓ -3.120	TargetScan Human	Moderate (predicted)	↓↑		7994131	PRKCB	↑ 4.995	kinase	Breast Cancer Regulat...				
<input checked="" type="checkbox"/>	hsa-miR-206	mir-1	↑ 1.880	TargetScan Human	Moderate (predicted)	↑↓		7956301	LRP1	↓ -3.463	transmembrane receptor	Colorectal Cancer Met...				
<input checked="" type="checkbox"/>	hsa-miR-206	mir-1	↑ 1.880	TargetScan Human	High (predicted)	↑↓		8008201	NGFR	↓ -2.917	transmembrane receptor	PTEN Signaling				
<input checked="" type="checkbox"/>	hsa-miR-122	mir-122	↑ 1.970	TargetScan Human	High (predicted)	↑↓		7963670	MAP3K12	↓ -3.119	kinase	Germ Cell-Sertoli Cell...				
<input checked="" type="checkbox"/>	hsa-miR-122	mir-122	↑ 1.970	TargetScan Human	Moderate (predicted)	↑↓		8157524	TLR4	↓ -6.290	transmembrane receptor	Colorectal Cancer Met...				
<input checked="" type="checkbox"/>	hsa-miR-125a-5p	mir-125	↓ -1.450	TargetScan Human	Moderate (predicted)	↓↑		7985213	CHRNA5	↑ 2.965	transmembrane receptor	AMPK Signaling				
<input checked="" type="checkbox"/>	hsa-miR-125a-5p	mir-125	↓ -1.450	TargetScan Human	High (predicted)	↓↑		7994131		↑ 4.995						

miRNA  
datamiRNA  
Target  
FilterMolecule  
TypePathways  
(Cancer/  
Growth)

mRNA

?

88 data  
points13,690  
targets1,090  
targets333  
targets39  
targets32  
targets

Use Pathway tools to build hypothesis for microRNA – target association to melanoma metastasis.



Ingenuity®  
Knowledge Base

IPA has high-quality microRNA-related findings (including both experimentally validated and predicted interactions)

- **TarBase**: experimentally validated microRNA-mRNA interactions
- **Target Scan**: predicted microRNA-mRNA interactions (low-confidence interactions were excluded)
- **miRecords**: experimentally validated human, rat, and mouse microRNA-mRNA interactions
- **Literature Findings**: microRNA-related findings manually curated from published literature by scientific experts and structured into the Ingenuity® Knowledge Base

Single source for microRNA content plus related biology enables biologically relevant target prioritization in minutes vs. weeks

Extensive human, mouse, and rat coverage



- For Searching, IPA Supports:
  - miRBase Identifiers
  - Entrez Gene Symbols and Entrez Gene IDs
  - Other synonyms used in the literature
  
- For Data Upload, IPA Supports:
  - miRBase Identifiers for mature miRNAs
  - miRBase Accession Numbers (format MIMAT#####) are preferred. These are stable identifiers.
  - miRBase Name Identifiers (format: mmu-miR-###) are allowed. Since some miRNA arrays provide annotations only with the name, we have provided mappings for them. These change over time so use MIMAT instead if available.
  - Precursor identifiers are NOT supported
  - Entrez Gene IDs (not Entrez Gene Symbols)
  - HUGO gene symbols (human only)



## Mapping microRNA IDs in IPA during Data Upload

- A given ID can only map to a single node in IPA
- miRNA identifiers each map to either a group node or a locus-specific node:
  - miRNA identifiers that correspond to mature miRNAs that do NOT appear in a group (ie, they arise from only one known precursor, and that precursor has no more than one known Entrez Gene ID/locus) are mapped to a locus-specific node.
  - miRNA identifiers corresponding to mature miRNAs that ARE in a group map to that group.
  - No miRNA maps to more than one group node in IPA.

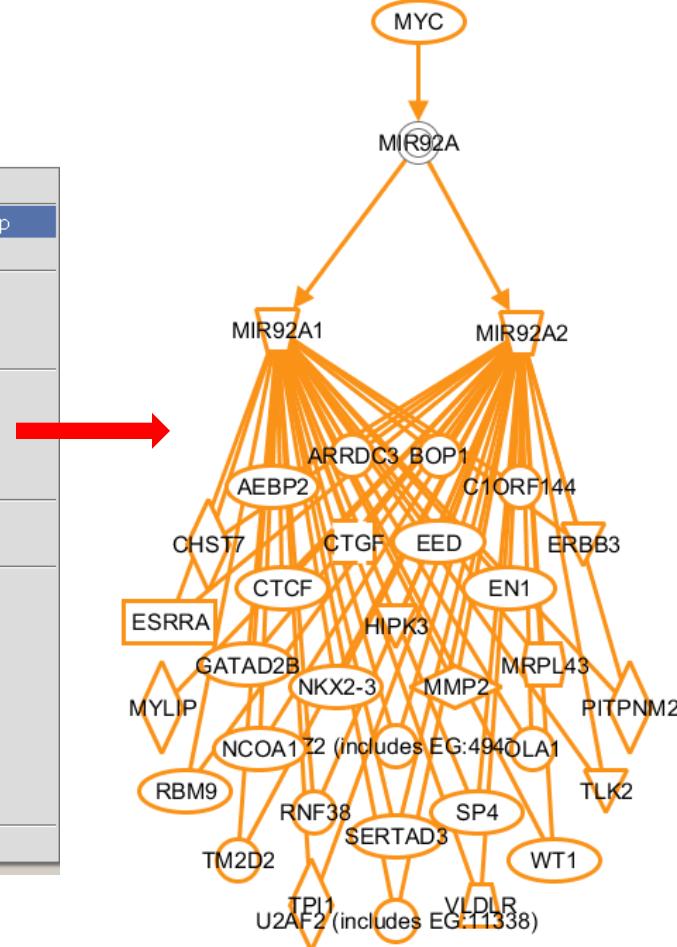
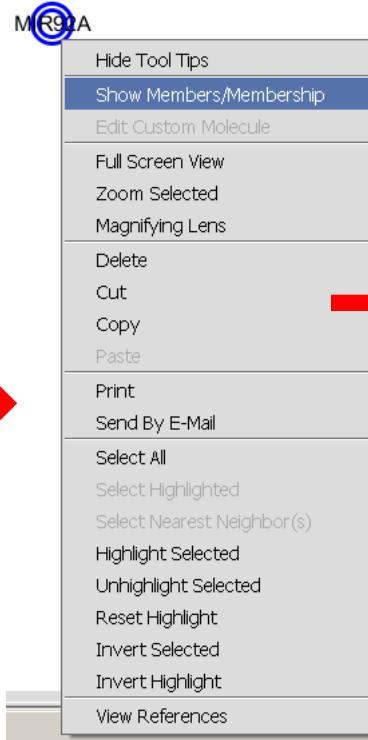


# Working with miRNA Groups

- Mature miRNAs may arise from multiple precursors:
  - A given mature form may arise from multiple distinct miRNA precursors.
  - A given precursor may arise from multiple distinct loci.
- Groups are created in the knowledge base to represent mature miRNA's that may arise from multiple precursors or multiple loci.
  - When authors refer to a particular mature miRNA form that may arise from multiple distinct precursors and/or multiple genetic loci, the finding is mapped to a group concept that contains all potential “parent” precursors.
- Groups might have different network connections compared to the individual members of the group.
  - Findings might be mapped to the individual members or to the group, depending on information provided by the author.
  - ‘Grow’ functionality does not ‘look inside’ the groups.
  - Additional steps will ensure that all members of group will be considered when applying ‘Grow’



Expanding groups prior to Growing will provide information on known molecular interactions for all members of the group.





## Biomarker Filter

IPA-Biomarker™ analysis filters/refines candidate lists based on biological criteria such as association to a disease, normal presence in a fluid, or normal expression in a tissue/cell type/cell line and/or clinical usage.

- Species
- Tissues and Cell Lines
- Biofluids
- Diseases
- Clinical Biomarkers

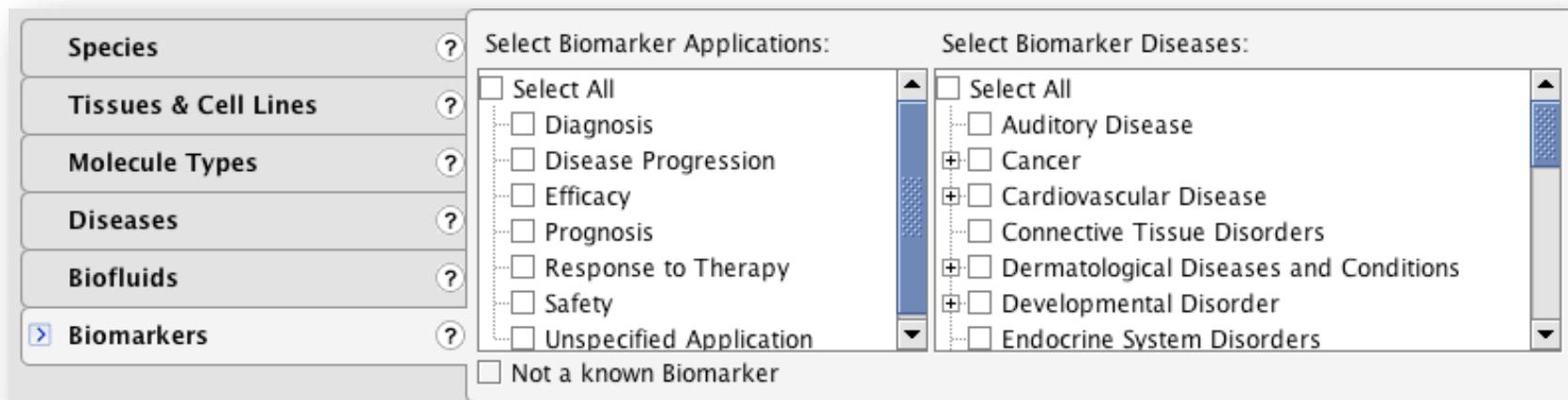
The output is a refined list of candidates

- It does not calculate functions, Canonical Pathways, or networks

Different observations or datasets can be compared using the Comparison Biomarker Analysis

- Calculates unique and common molecules

The Biomarker Filter capability rapidly prioritizes biomarker candidates based on biological characteristics and clinical usage.



## Clinical Usage (Biomarkers):

Identify biomarkers by their specific application, including markers for Disease Diagnosis and Prognosis, Disease Progression, markers of Drug Efficacy and Safety, and Patient Response to Therapy

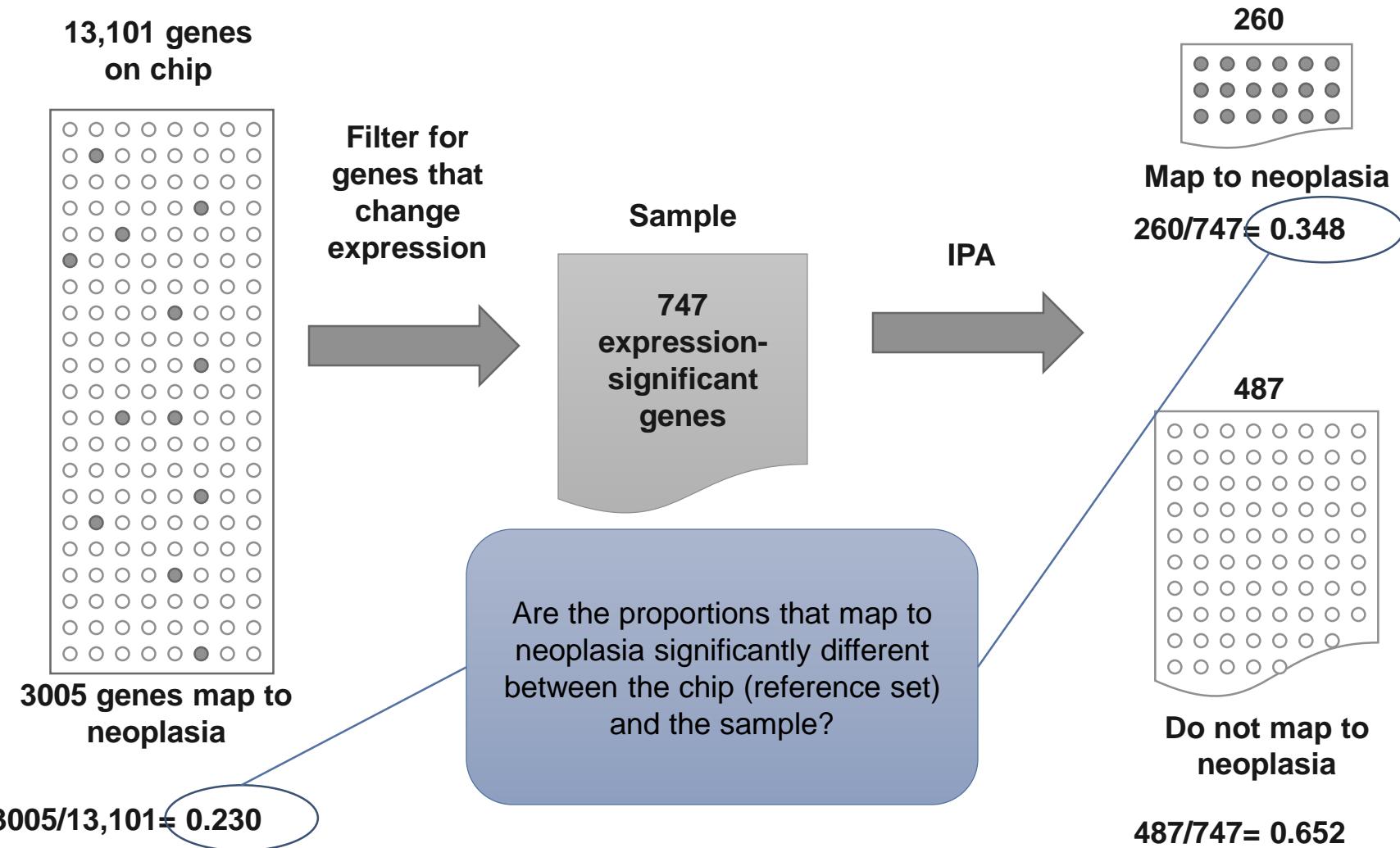
# Statistics in IPA



## How the Fisher's Exact Test is Calculated

- Is the proportion of genes in my sample mapping to a gene set (those that are significant) similar to the proportion of all measurable genes (reference set) that map in the gene set?
  - If the proportions are similar, there is no biological effect

# Mapping Colorectal Cancer Expression Data to the Function “Neoplasia”



# Calculating the Fisher's Exact Test

A 2x2 contingency table is created based on the total population, the sample, and how many genes map to the function/pathway. This table is used to calculate the Fisher's Exact Test

	Neoplasia	Not Neoplasia	
In Sample	k	n - k	n
Not in Sample	m - k	N + k - n - m	N - n
	m	N - m	N

$m$ = Total that map to function/pathway

$N$ = Total

$k$ = Number that map to function/pathway in sample

$n$ = Total sample

# Calculating the Fisher's Exact Test

Numbers based on the colorectal cancer data mapping to neoplasia

	Neoplasia	Not Neoplasia	
In Sample	260	487	747
Not in Sample	2745	9609	12354
	3005	10096	13101

3005 = Total that map to neoplasia on chip

13101 = Total on chip

260 = Number that map to neoplasia in sample

747 = Total sample

Fisher's Exact Test p-value = 2.13 E-14



# What Can We Say About Our Colorectal Cancer Data Set And Neoplasia?

---

- We can conclude that the proportion, or over representation, of genes mapping to neoplasia is not likely the result of sampling (and is likely an effect of the disease)



## How Do I Choose The Reference Set?

---

If you are using a standard vendor platform supported by IPA, then that platform should be selected as your reference set.

If you do not know the platform or the data were taken from different platforms, select a reference set that best estimates the entire population you evaluated.

- For gene expression data, select the “Ingenuity Knowledge Base (genes only)”
  - This setting uses all function- and pathway-eligible genes in the knowledge base.
- For metabolomics, select the “Ingenuity Knowledge Base (endogenous chemicals only)”
- You have the option to having your uploaded data set used as the reference set (User Data Set)



## What About TaqMan or Similar Focus Array?

Low density arrays are problematic because the genes that are being measured are usually not randomly chosen to start with, but are typically selected based on *a priori* function or pathway knowledge

Let's assume a inflammatory cytokine array

- If you select the Ingenuity Knowledge Base as your reference, your p-values for inflammation functions and pathways will be artificially low (significant) because the array was heavily biased for these genes.
- If you upload every gene on the array, and select the “User Data Set” reference option, your p-values are statistically accurate, but inflammatory functions and pathways may not appear significant because the likelihood of having a random sample with similar proportions to inflammation processes is extremely high.

Benjamini-Hochberg method of multiple testing correction

Based on the Fisher's exact test p-value

Calculates false discovery Rate

- Threshold indicates the fraction of false positives among significant functions



5% (1/20) may be a false positive



# Which p-Value Calculation Should I Use?

“What is the significance of function X in relation to my dataset?”

- Use Fisher’s Exact test result

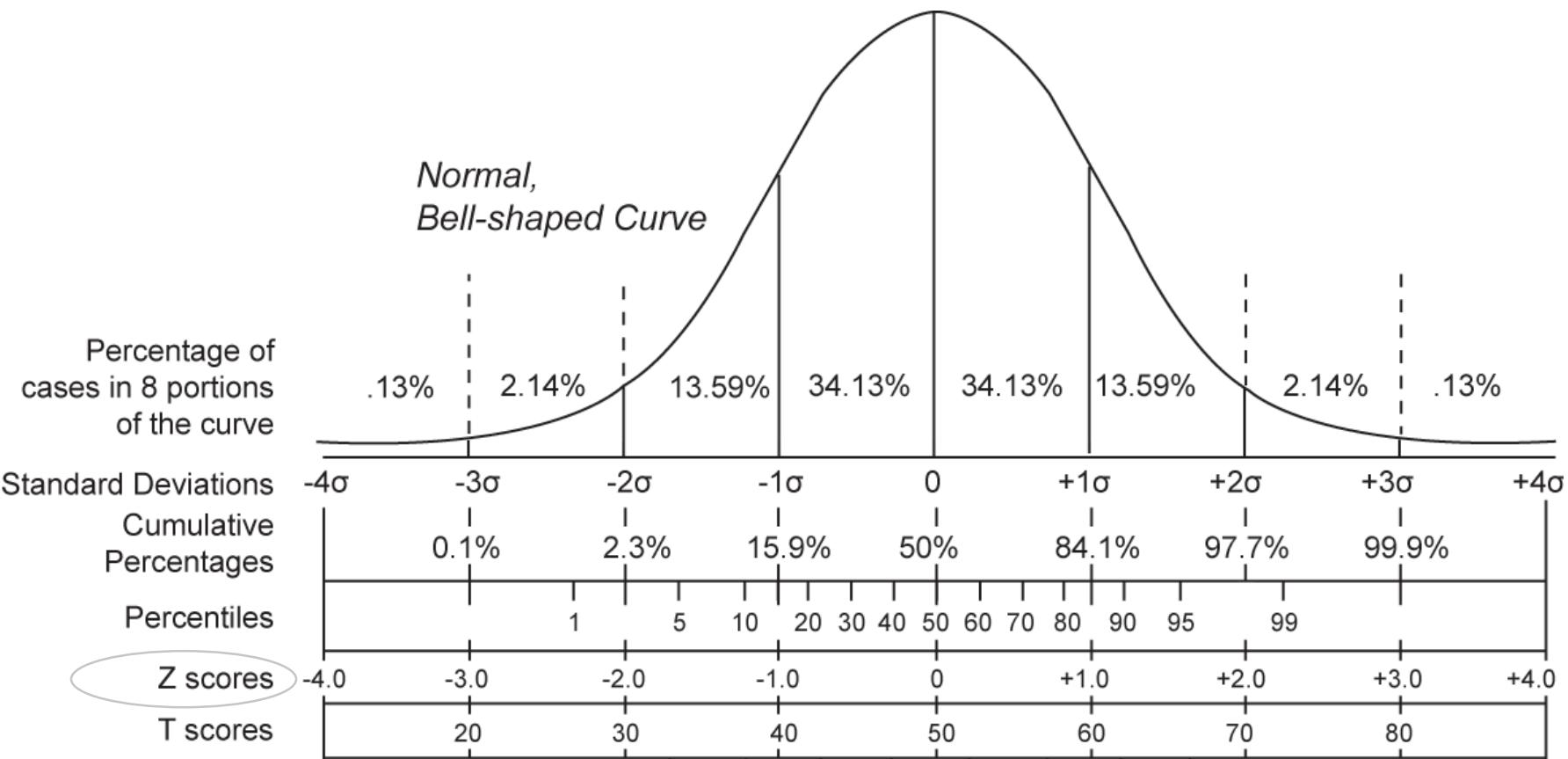
“What are all significant functions within this dataset?”

- Use Benjamini-Hochberg multiple testing correction

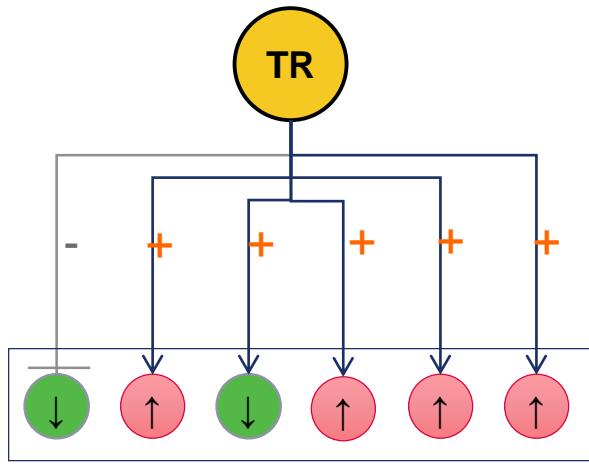
# z-scores and Normal Distribution

A set of genes chosen at random should be about equally likely to have an increasing or decreasing effect, thus, about 50% each direction, or a z=0.

A z-score represents the non-randomness of directionality within a gene set



# Activation z-score



Every TR is analyzed

Literature-based effect TR has on downstream genes

Differential Gene Expression (Uploaded Data)

Predicted activation state of TR:

1: activated (correlated), -1: inhibited (anti-correlated)

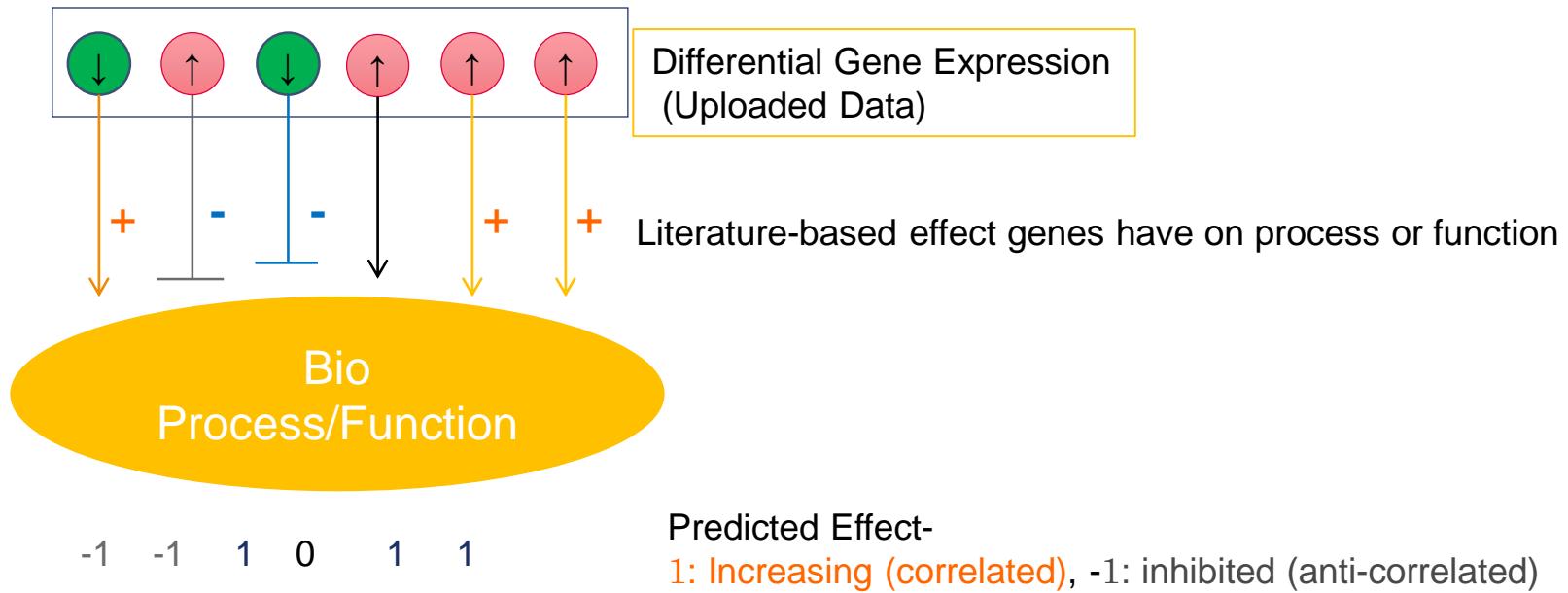
1    1    -1    1    1    1

$$z = \frac{x}{\sigma_x} = \frac{\sum_i x_i}{\sqrt{N}} = \frac{N_+ - N_-}{\sqrt{N}} = \frac{4}{\sqrt{6}} = 2.04$$

- z-score is statistical measure of correlation between relationship direction and gene expression.
- z-score  $> 2$  or  $< -2$  is considered significant

Actual z-score can be weighted by relationship, relationship bias, data bias

# Downstream Effect z-score



$$z = \frac{x}{\sigma_x} = \frac{\sum_i x_i}{\sqrt{N}} = \frac{N_+ - N_-}{\sqrt{N}} = \frac{1}{\sqrt{5}} = .447$$

- “z-score” is statistical measure of correlation between relationship direction and gene expression.
- z-score > 2 or < -2 is considered significant

Actual z-score is weighted by relationship, relationship bias, data bias