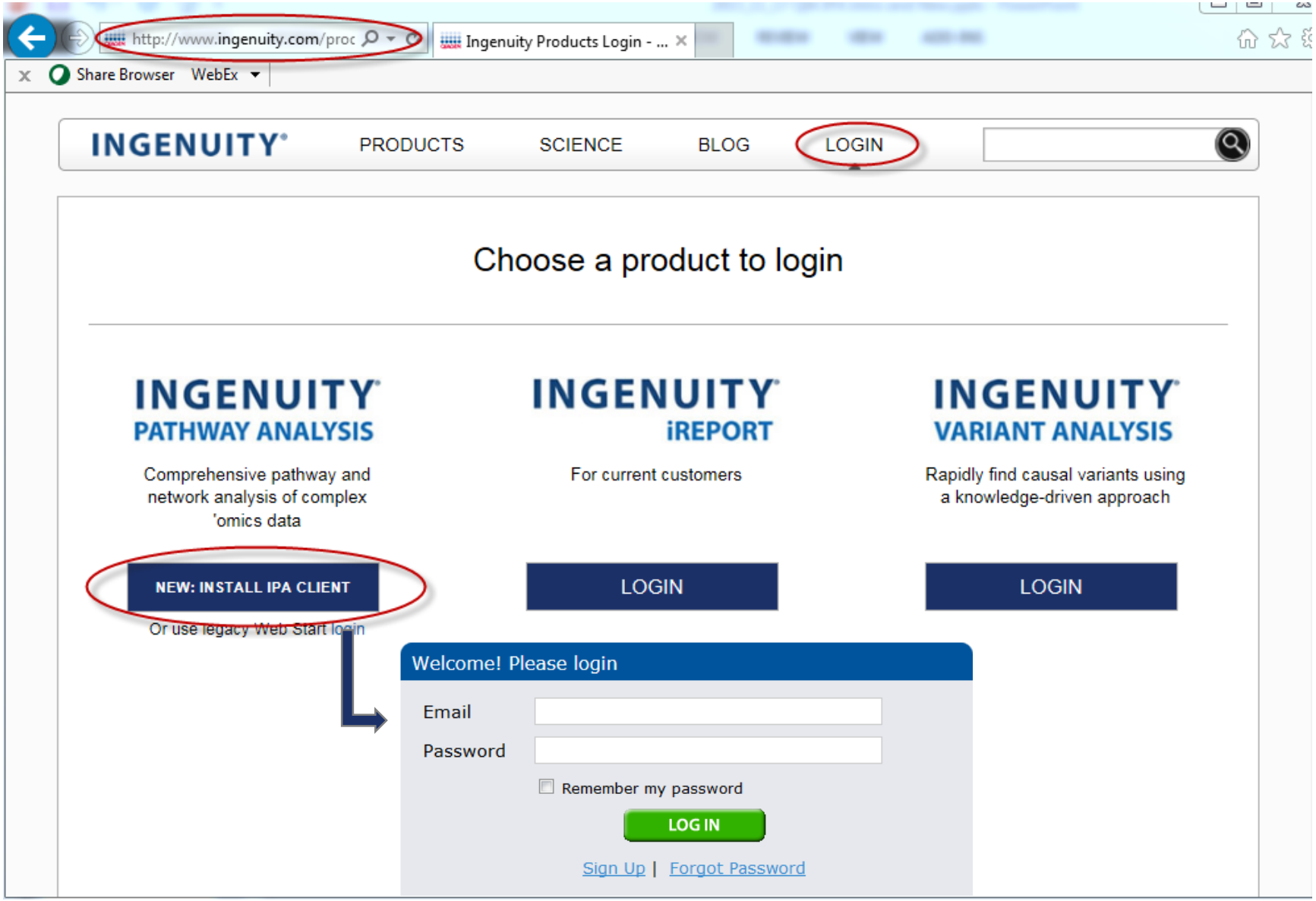
IPA Training:
Maximizing the Biological Interpretation of Gene, Transcript & Protein Expression Data with IPA

Log into IPA: www.ingenuity.com



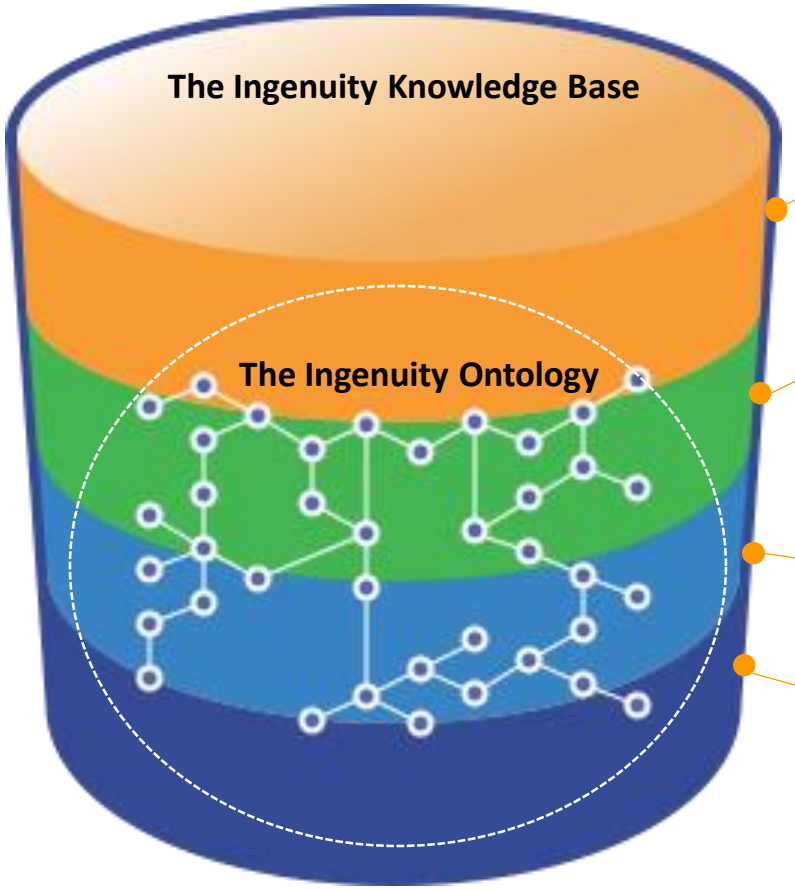
How can IPA help you?

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
		Log2Ratio
	_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
12	NM_016161	2.02
13	NM_015665	-0.27

+significance stat

	A	B	C
		Log2Ratio	p-value
	36	0.14	8.68E-01
	0	-0.99	2.24E-01
	32	-0.02	9.83E-01
	76	-0.02	9.85E-01
	33	0.02	9.79E-01
	14	-4.79	1.02E-01
	1	-0.67	6.17E-01
9	NM_144670	-5.96	1.30E-01
10	NM_001080438	-1.97	3.47E-01
11	NM_017436	-1.09	5.02E-01
12	NM_016161	2.02	5.97E-02
13	NM_015665	-0.27	5.68E-01

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

	A	B	C	D
		Log2Ratio	p-value	Intensity/ RPKM/FPKM
	86	0.14	8.68E-01	2931.69
	30	-0.99	2.24E-01	1649.26
	32	-0.02	9.83E-01	1.67
	76	-0.02	9.85E-01	1.77
	33	0.02	9.79E-01	1.83
	14	-4.79	1.02E-01	239.75
	71	-0.67	6.17E-01	213.79
9	NM_144670	-5.96	1.30E-01	610.64
10	NM_001080438	-1.97	3.47E-01	3.91
11	NM_017436	-1.09	5.02E-01	6186.83
12	NM_016161	2.02	5.97E-02	149.85
13	NM_015665	-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

Observation 1 Observation 2

	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	-1.42	1.03E-02	22828.45	-0.17	7.48E-01	3339.36

Verify the differential expression calculation

- Recommend Log_2(ratio) differential expression

$$Log_2\left(\frac{Experimental\ Condition\ Exp.}{Control\ Exp}\right)$$

- Ratio differential expression

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

- Fold Change
  - If increased differential expression

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

  - If decreased differential expression

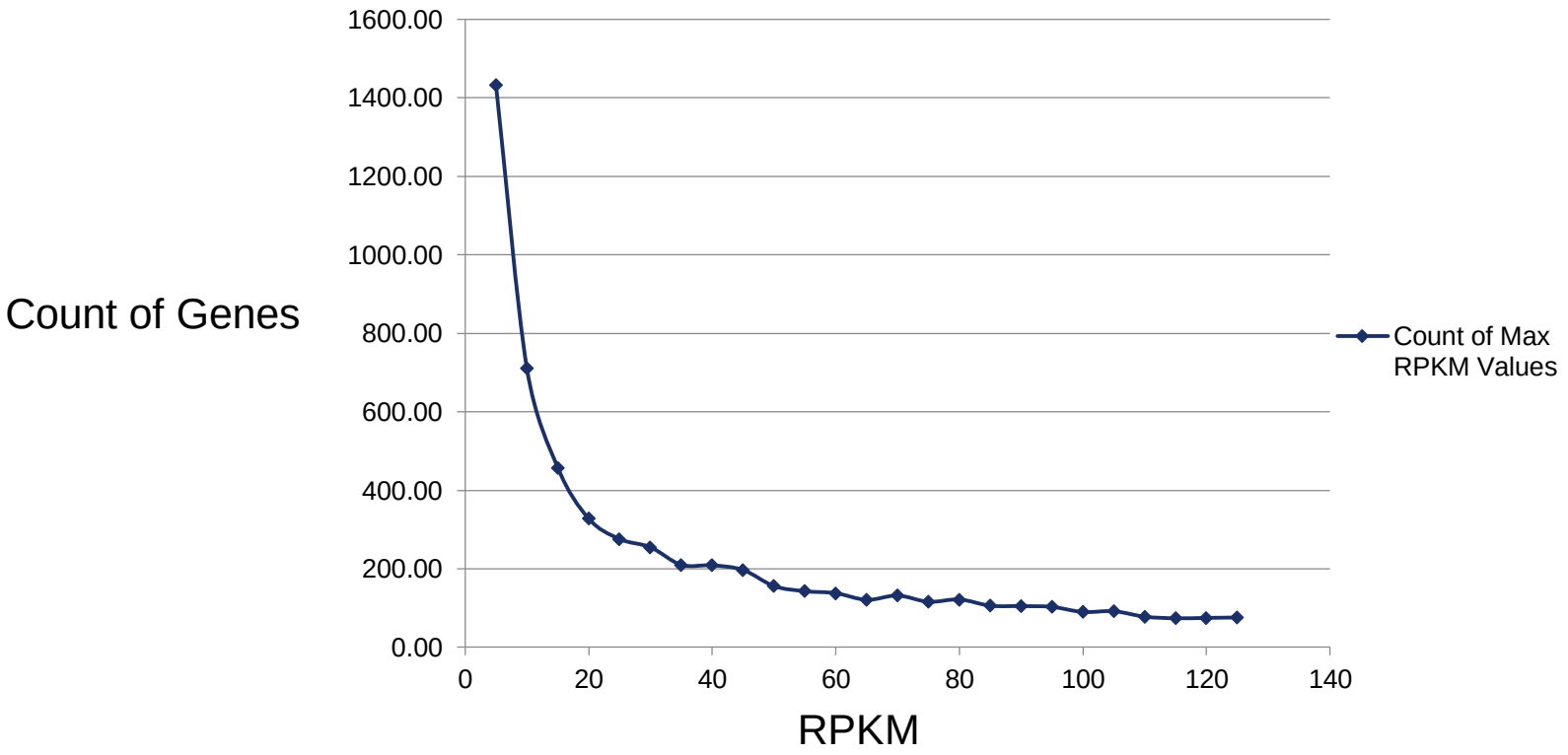
$$-1\left(\frac{Control\ Exp.}{Experimental\ Condition\ Exp.}\right)$$

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



Data Upload

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

Core Analysis Data Types

- 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.

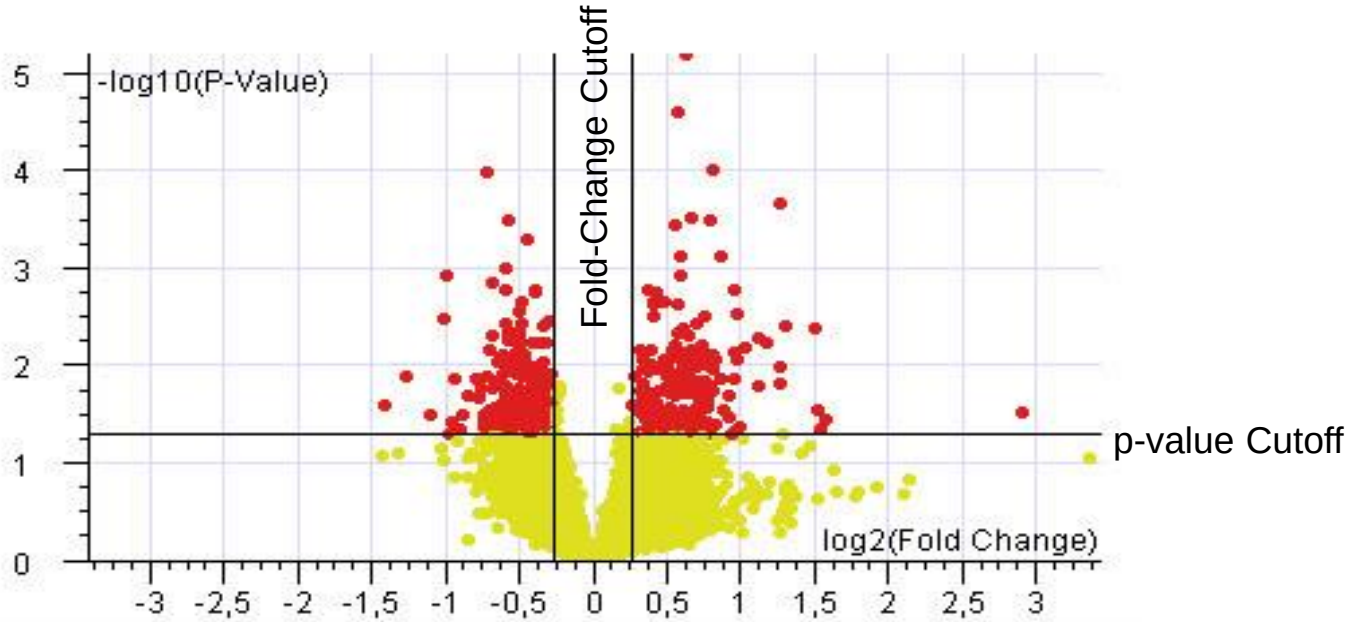
- For gene expression data, select the “Ingenuity Knowledge Base (genes only)”
- For metabolomics, select the “Ingenuity Knowledge Base (endogenous chemicals only)”
- You have the option to having your uploaded data set used as the reference set (User Data Set)

Core Analysis Set-up

Use cutoffs to refine set size per observation

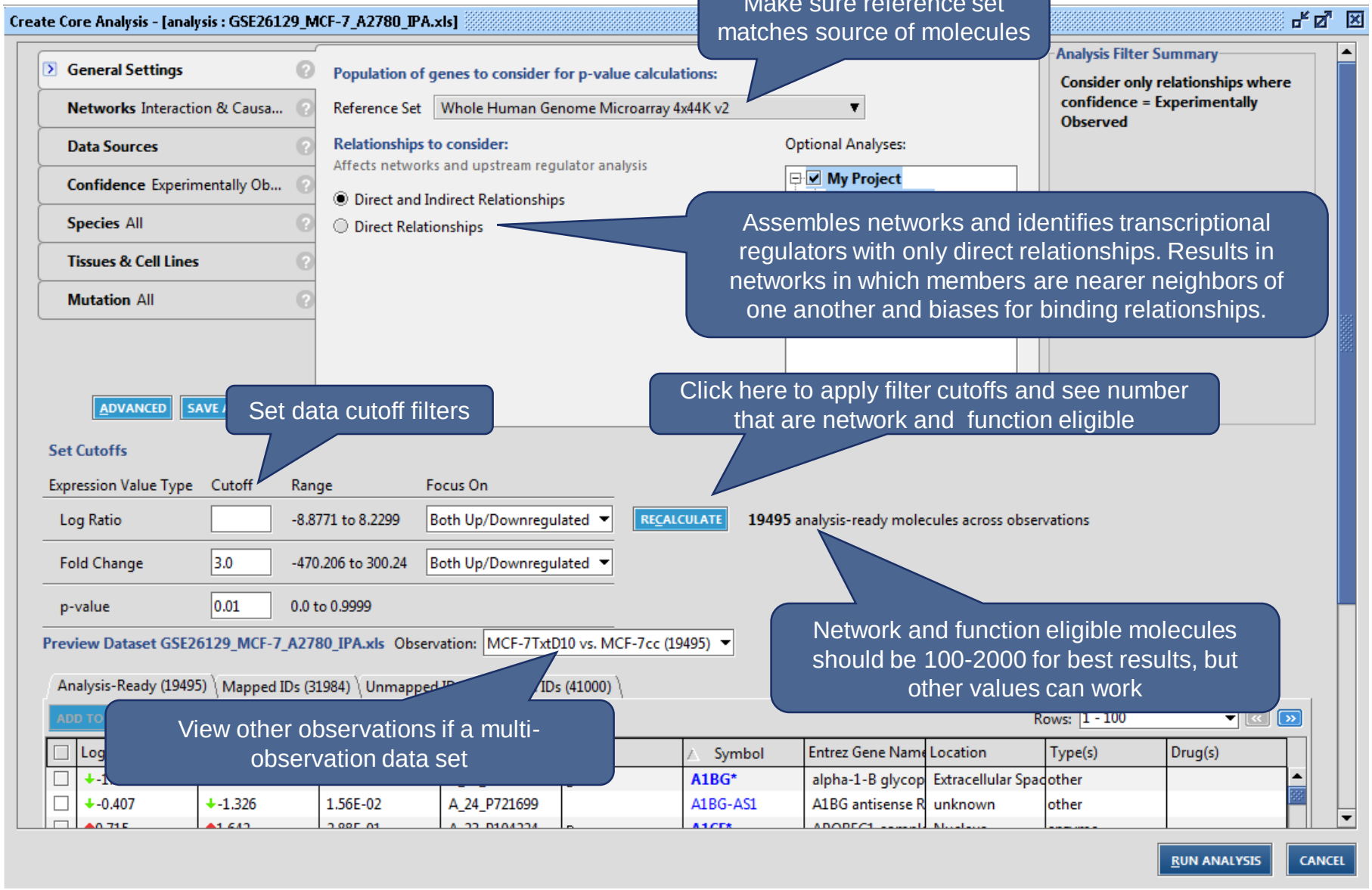
‘Ideal’ set size for IPA core analysis from gene expression data is typically 200-3000

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

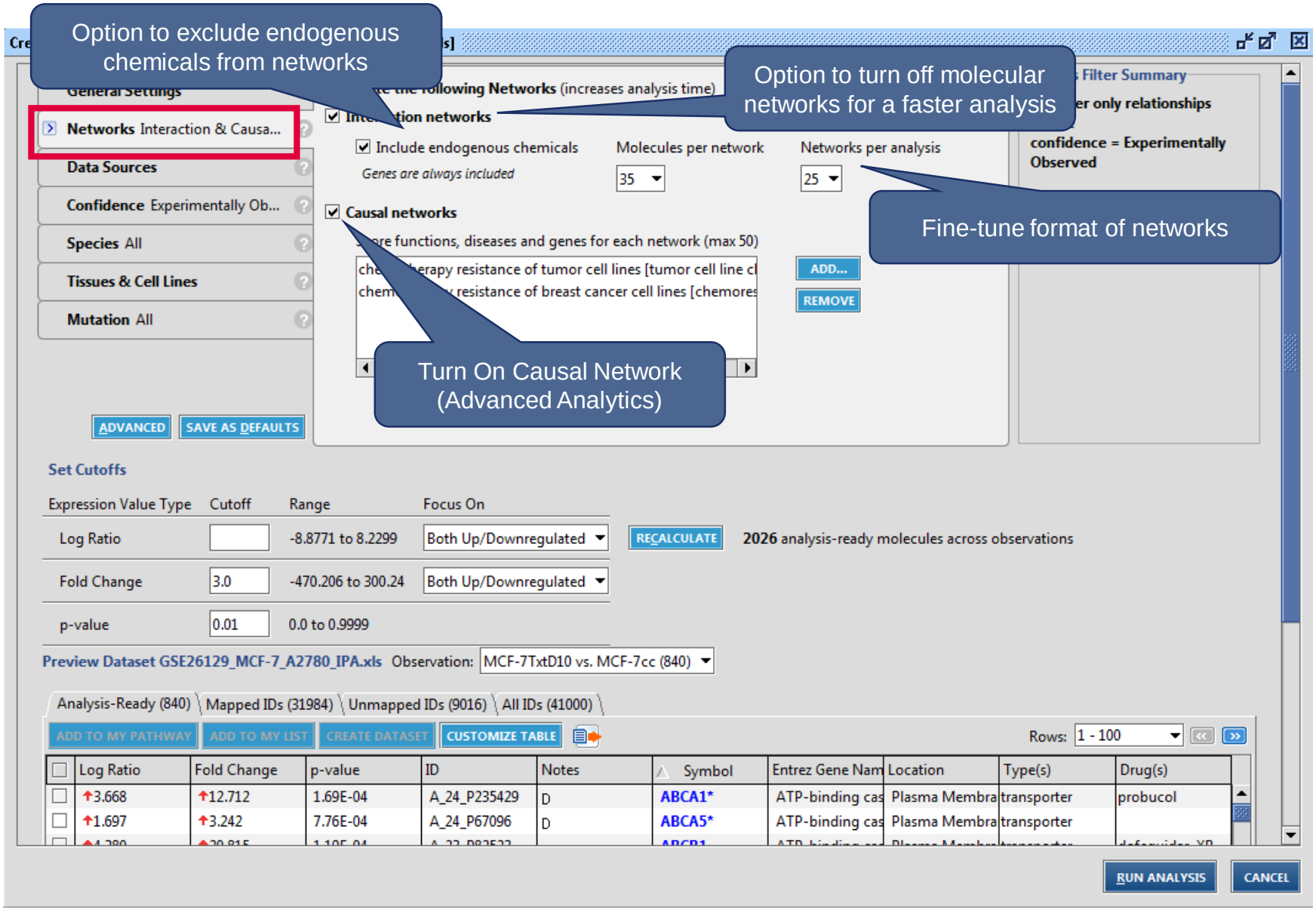


Adapted from Conti *et al* (2007), *BMC Genomics*, **8**, p268

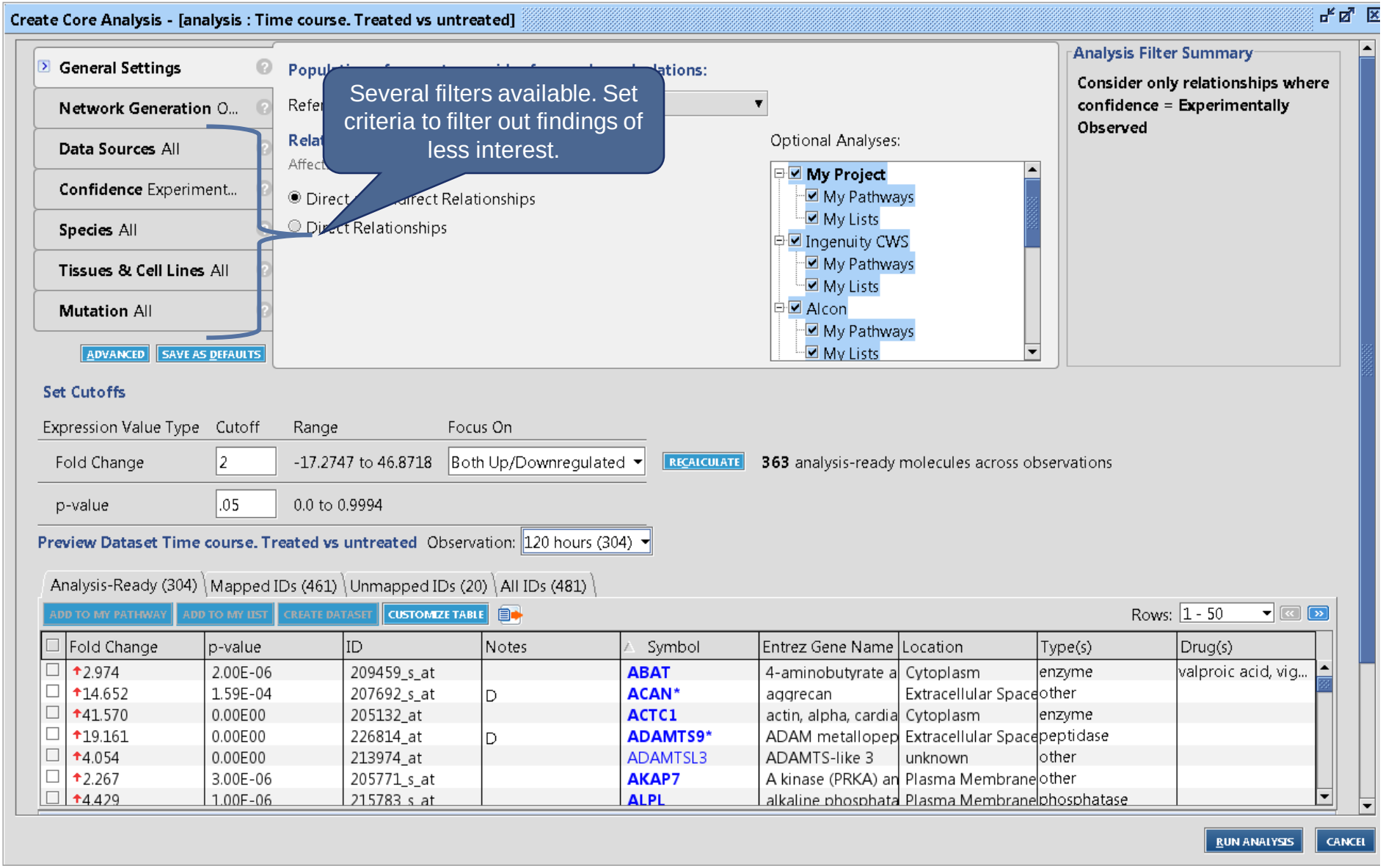
Creating an IPA Core Analysis



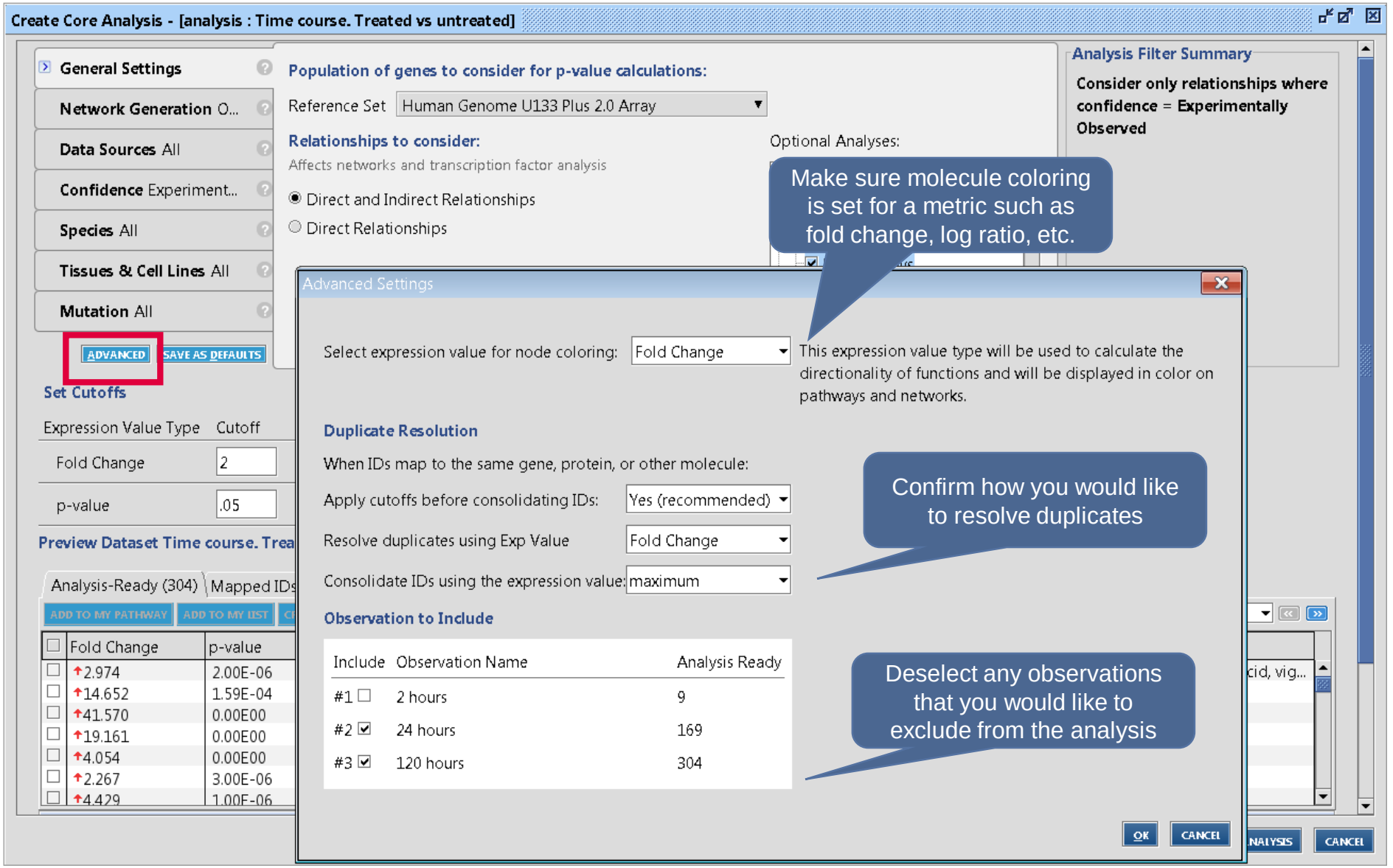
Creating an IPA Core Analysis- Network Generation



Creating an IPA Core Analysis: Using Filters



Creating an IPA Core Analysis- Advanced Settings



Using Core Analysis Pre-filters

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.

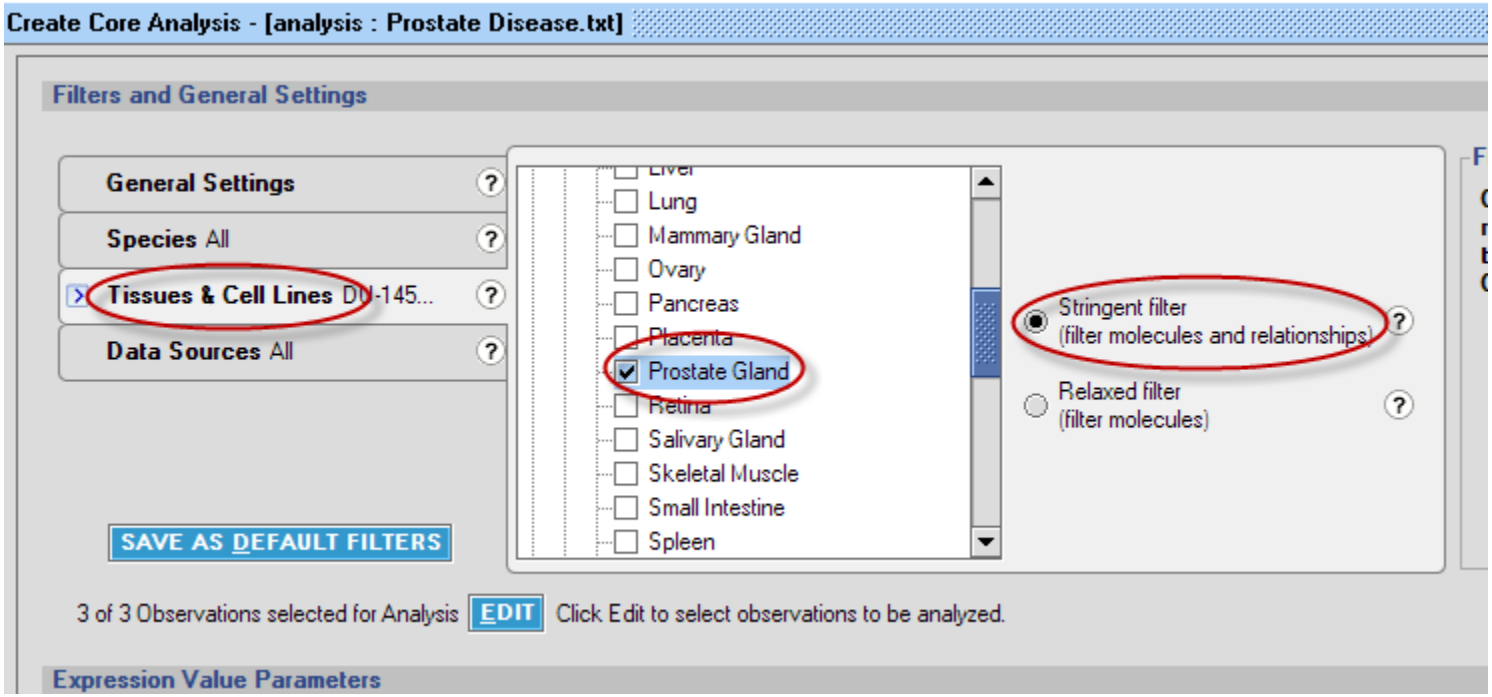
Unspecified refers to findings or molecules where cell/tissue/organ is not specified or classified

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
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)
SAVE AS DEFAULT FILTERS
3 of 3 Observations selected for Analysis
EDIT
Click Edit to select observations to be analyzed.
Expression Value Parameters

Large Scale Data Analysis

Gene/Protein Expression Analysis, Overview

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

Canonical Pathways

IL-10 Signaling

Edit: BUILD OVERLAY PATH DESIGNER View: Zoom: Export:

Overlay: MAP (Molecule Activity Predictor)

CLEAR

You can predict the upstream and downstream effects of activation or inhibition on other molecules. Begin by applying expression values from a dataset or analysis, or interactively in silico.

Predict effect of dataset or in silico changes

PREDICTION ON

Display prediction legend

Predict effects:

Upstream and Downstream

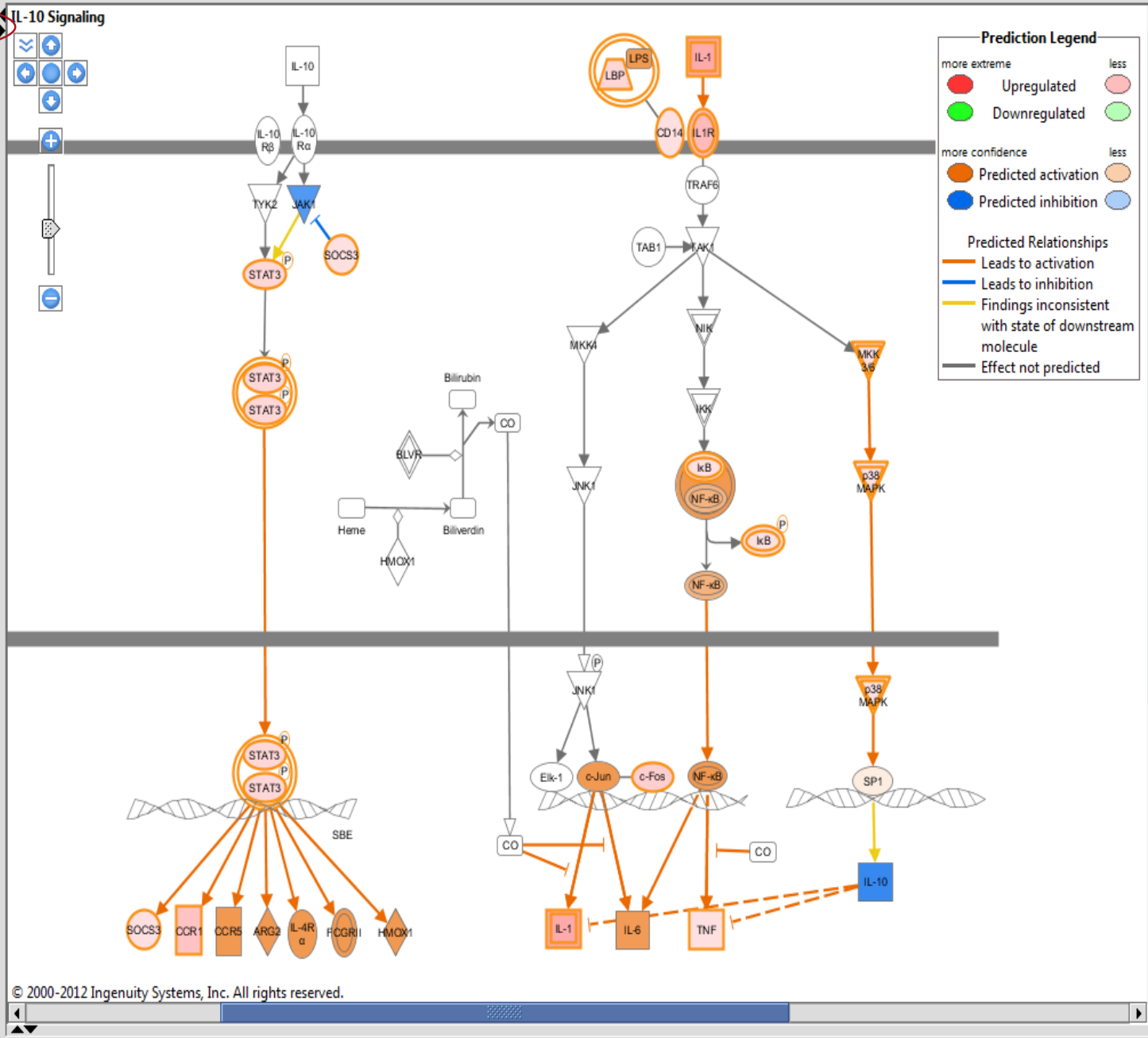
Activate or inhibit molecules interactively in silico

Select the value to apply and then click the molecules you wish to apply them to.

Use expression values from a Dataset or Analysis

Current Analysis/Dataset/List: Day 10

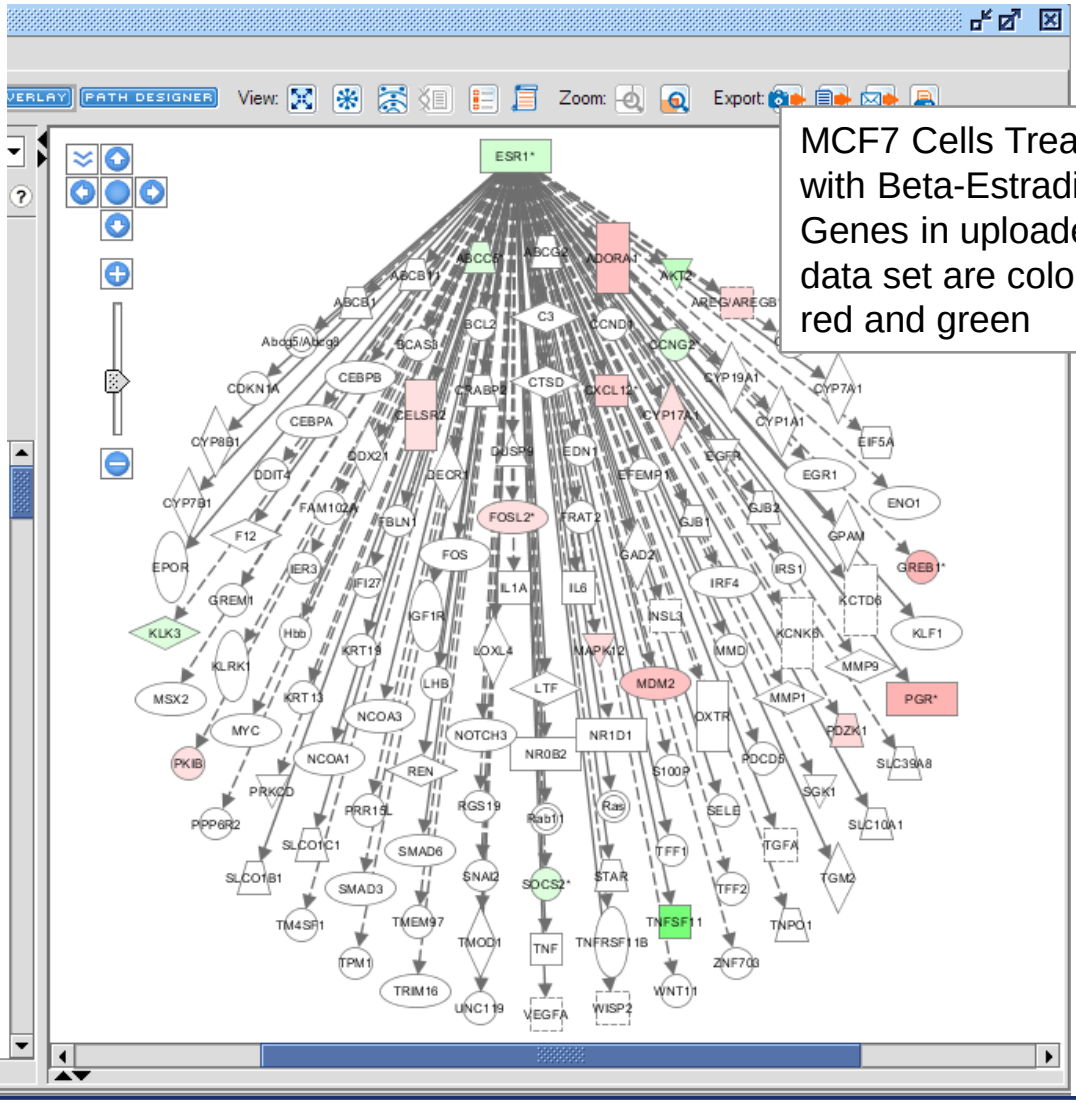
Change Analysis/Dataset/List



Identify likely upstream regulators and their activity state

IPA Upstream Regulator Analysis

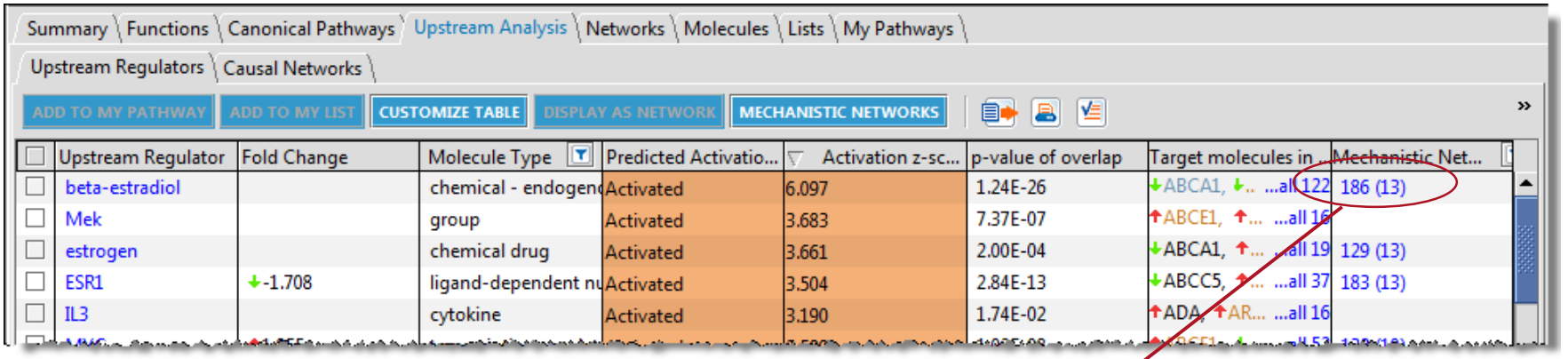
- Use published experimental molecular interactions to identify upstream regulators
- . Identify upstream regulators by determining gene enrichment in downstream genes
- Predict the activity state of regulators by correlating literature reported effects with observed gene expression



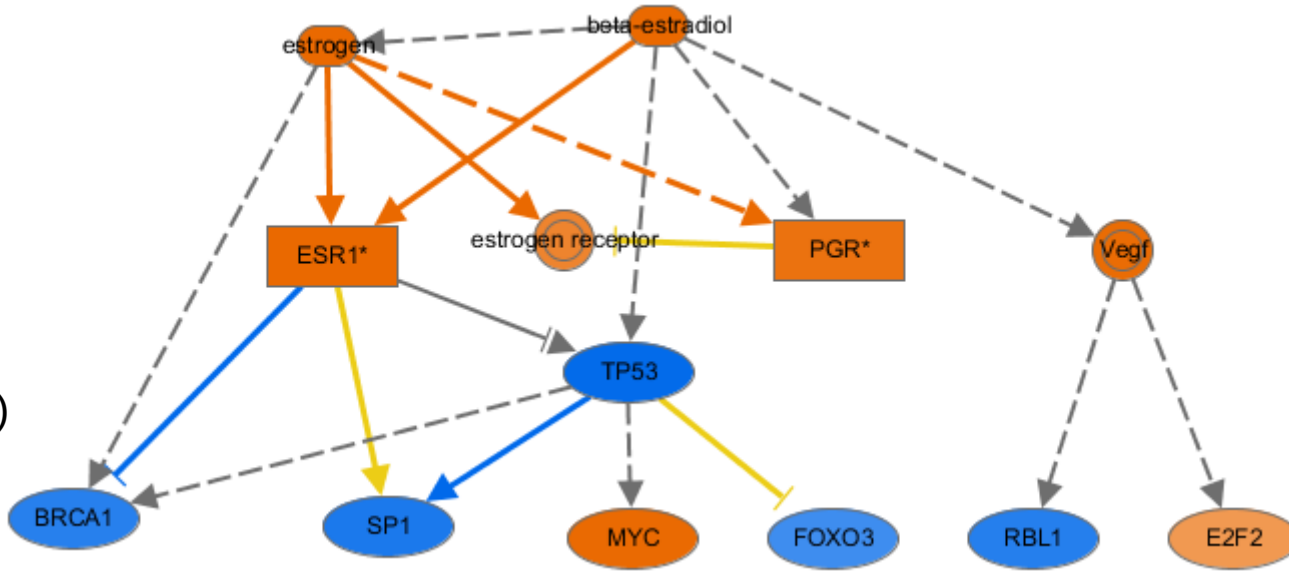
MCF7 Cells Treated with Beta-Estradiol. Genes in uploaded data set are colored red and green

Create *de novo* pathways of regulators and genes

IPA Mechanistic Networks



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



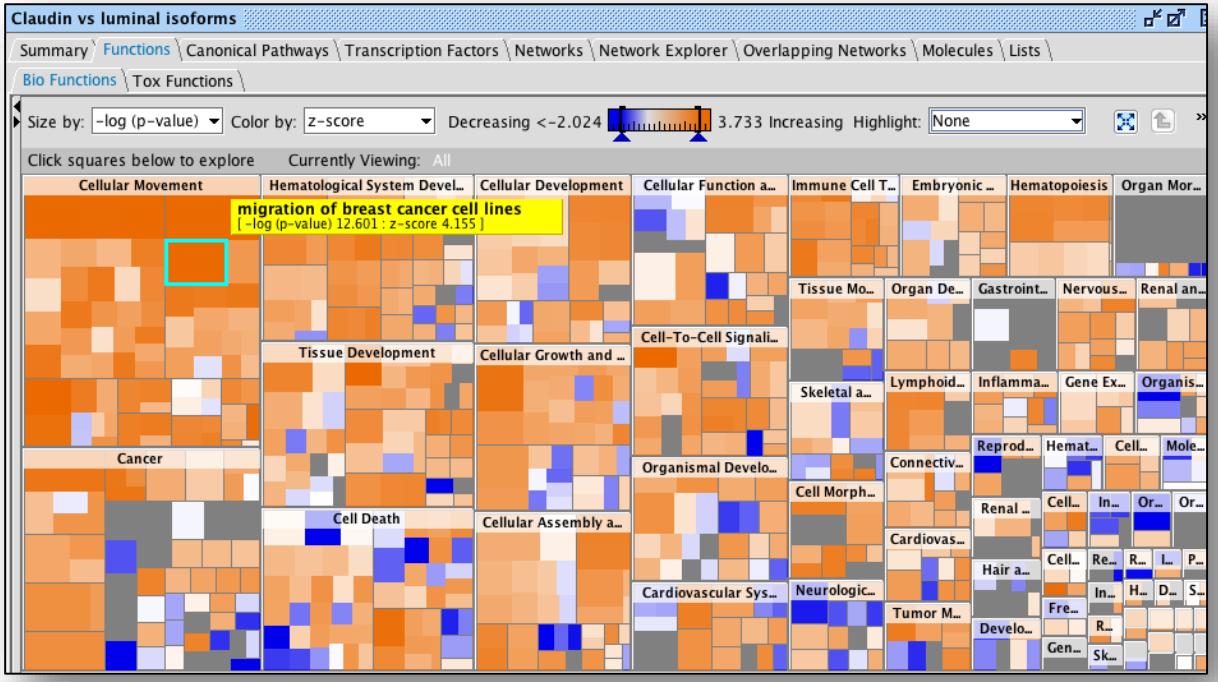
IPA Winter Release 2012

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

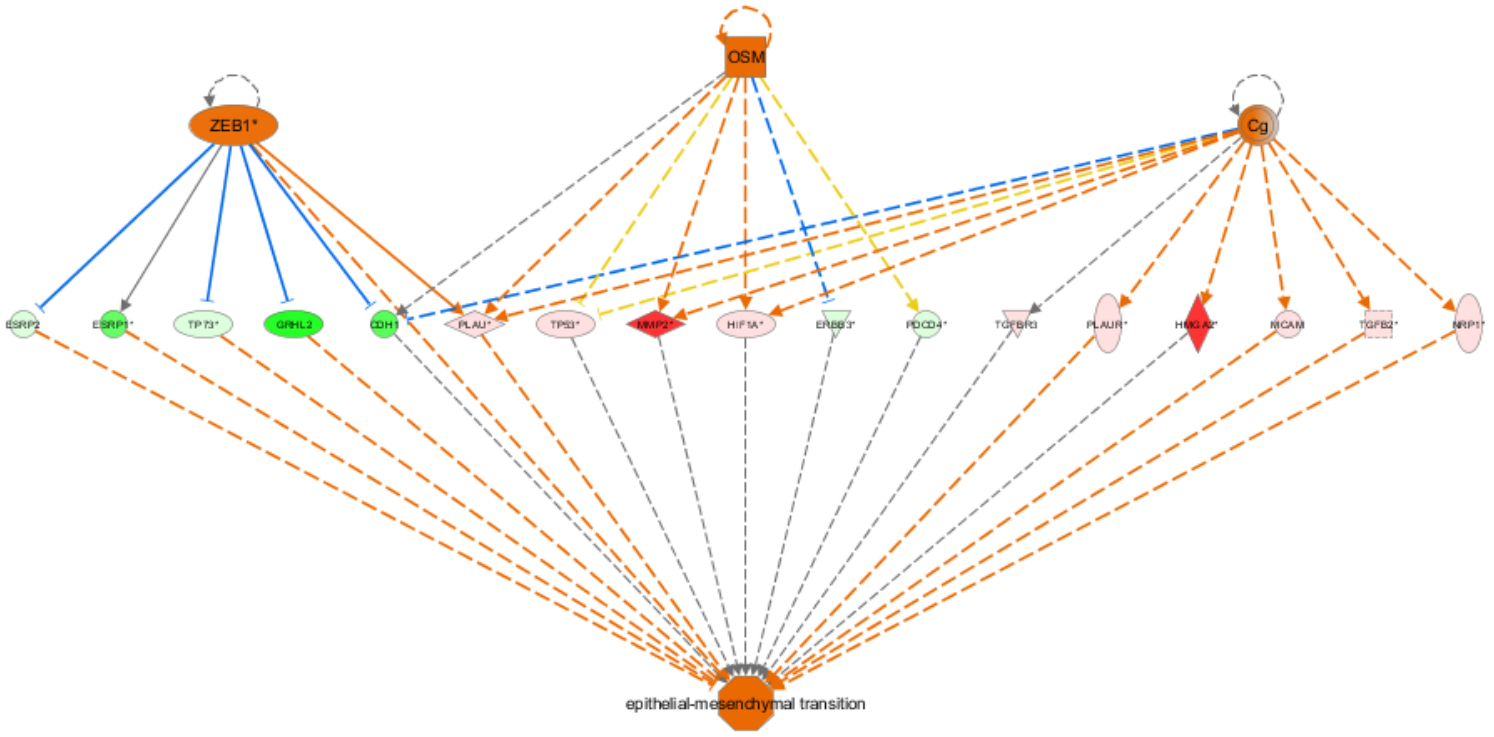


- Each box represents a biological process or disease
- The size of the box represents gene enrichment
- The color of the box indicates the predicted increase or decrease

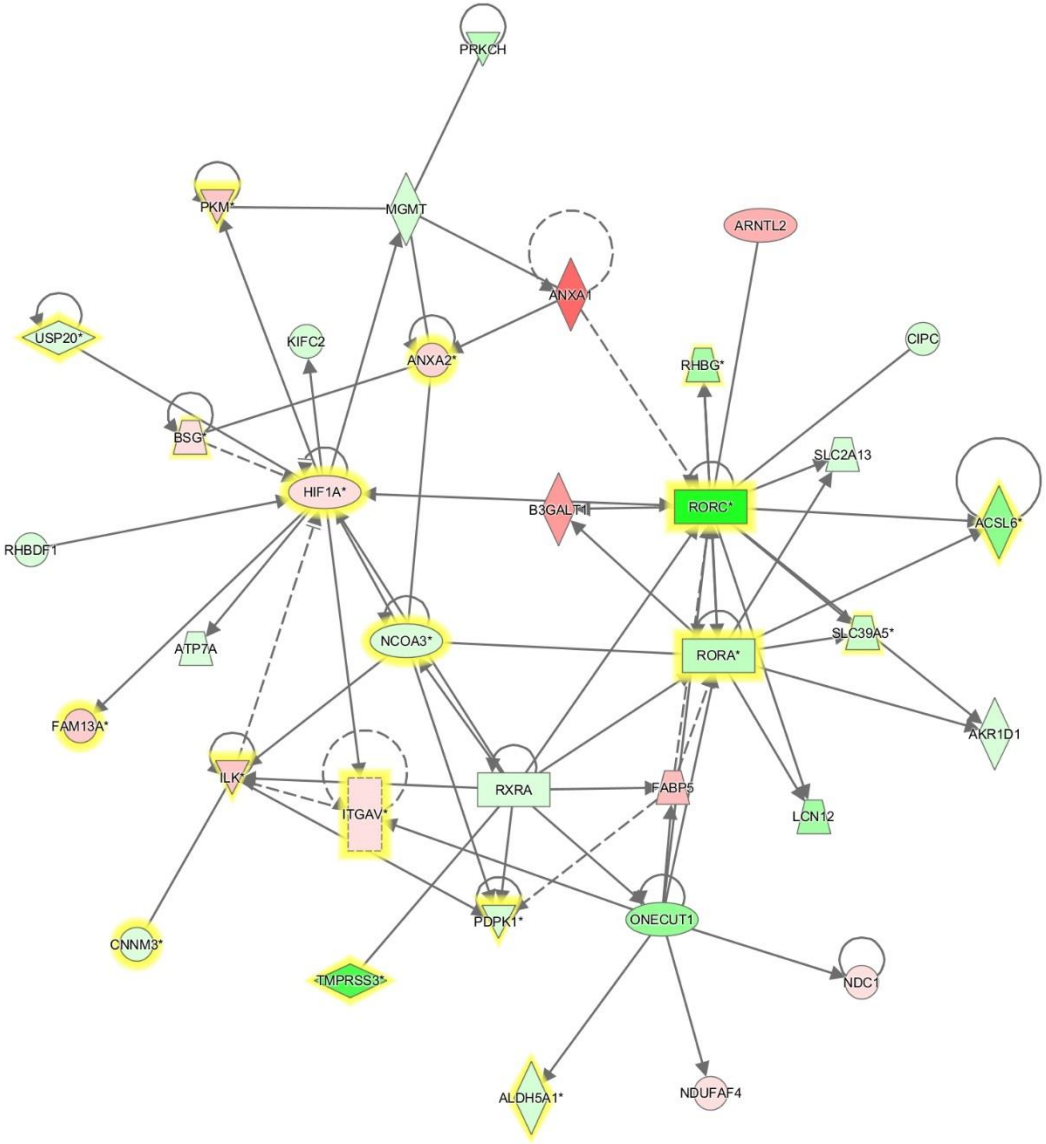
Regulator Effects

- 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

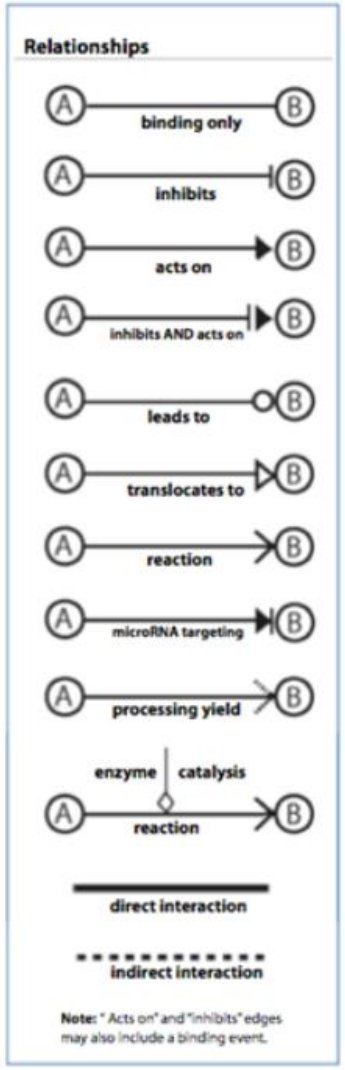
Pathway Analysis

- 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?

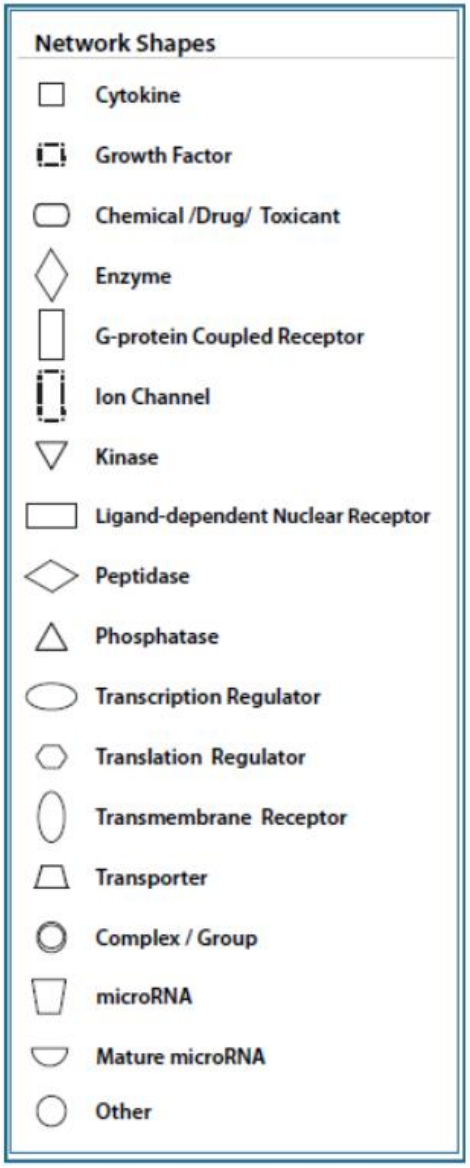
Pathway Analysis Interpretation

- 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



Canonical Pathway (CP) Workflow

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

Pathway Navigation

- Scroll-wheel on mouse controls zoom, or use toolbar zoom butt
- 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

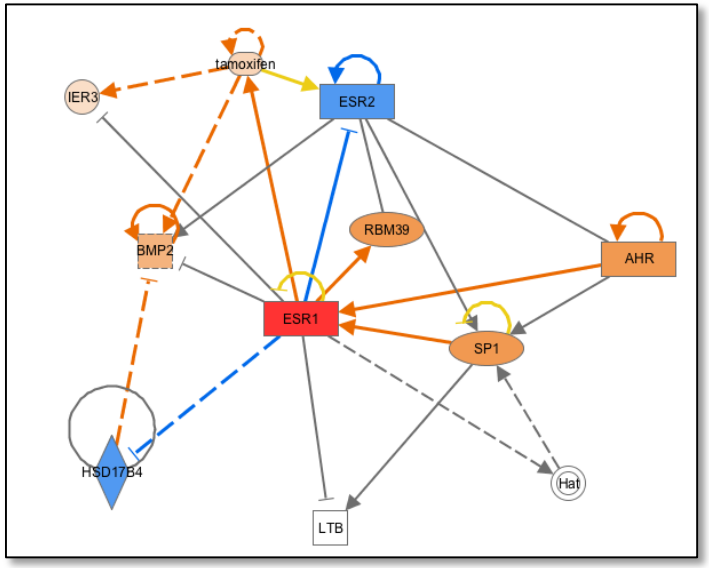
Navigation Control

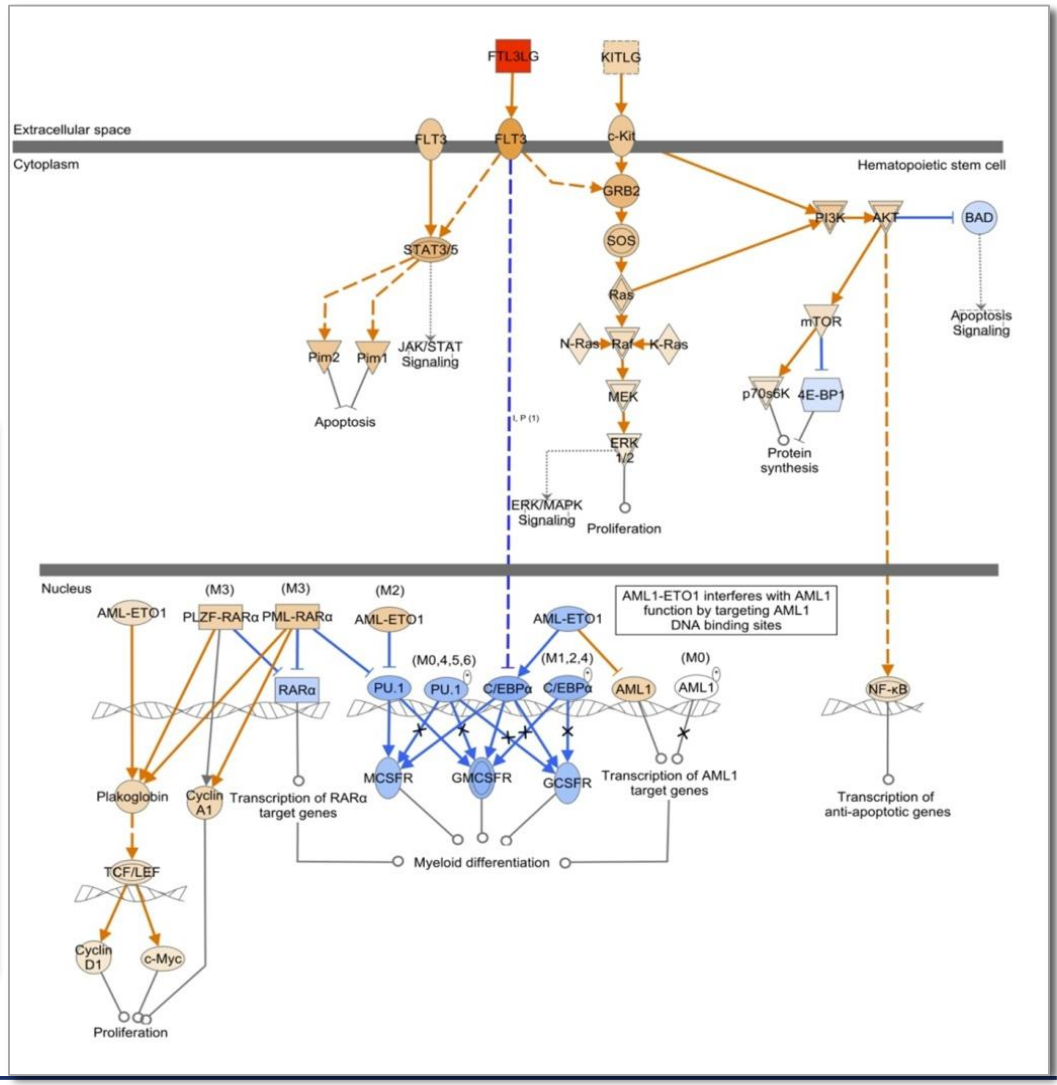
Pathway Navigation, continued

- 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**

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

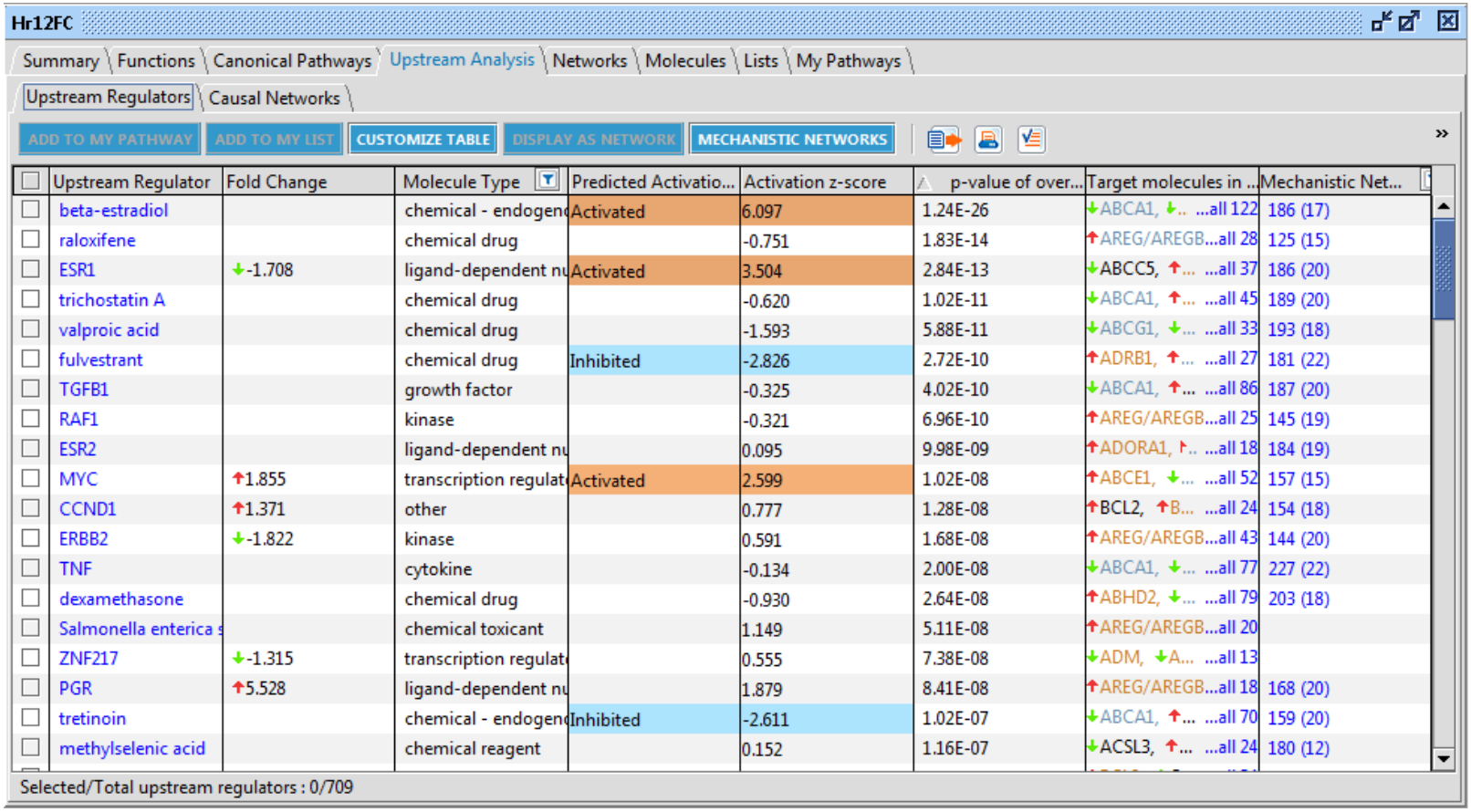
Upstream Regulators

Upstream Regulators and Networks

- 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

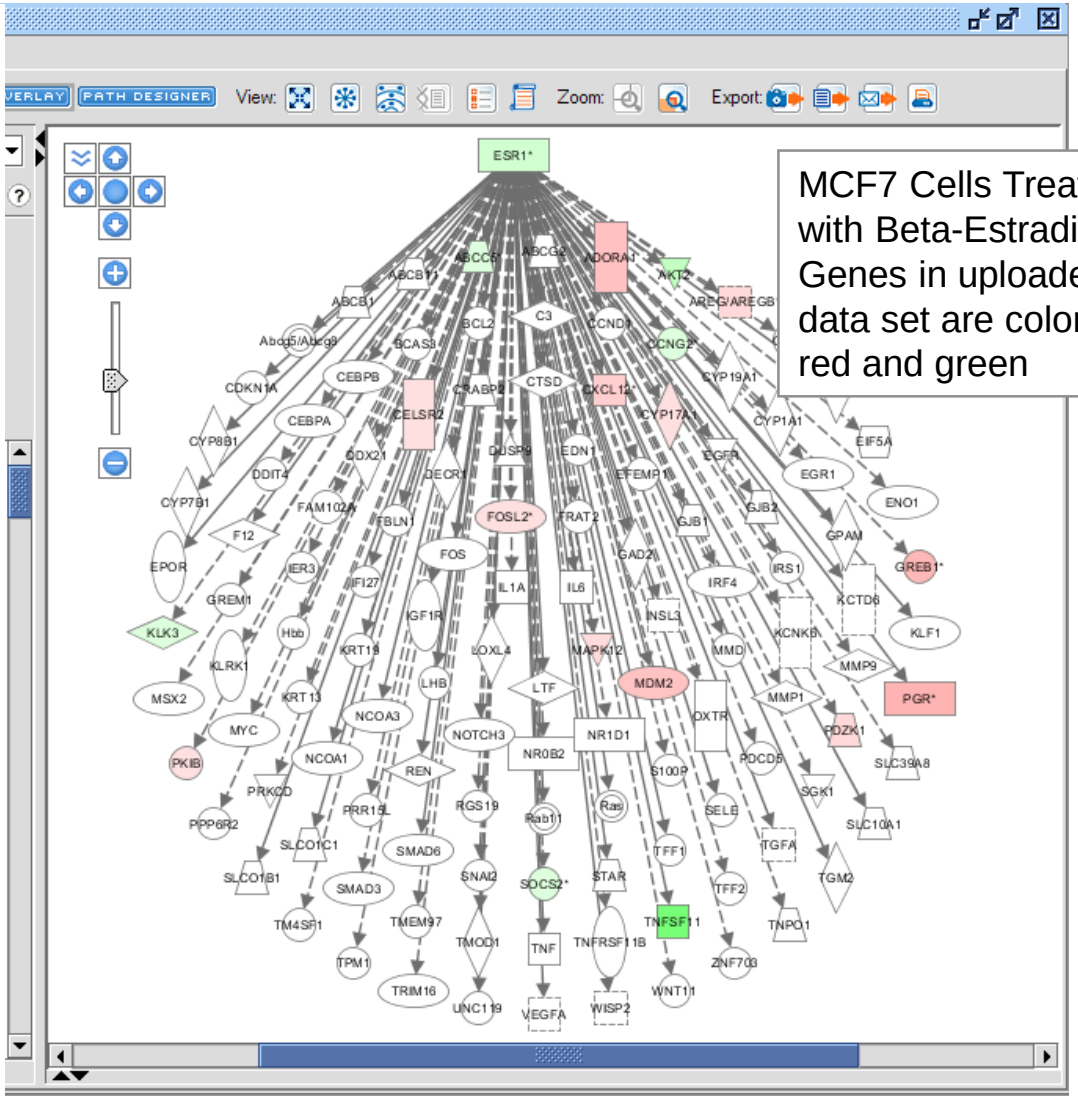


- Quickly filter by molecule type
- Filter by biological context
- Generate regulators-targets Network to identify key relationships

IPA Upstream Regulator Analysis

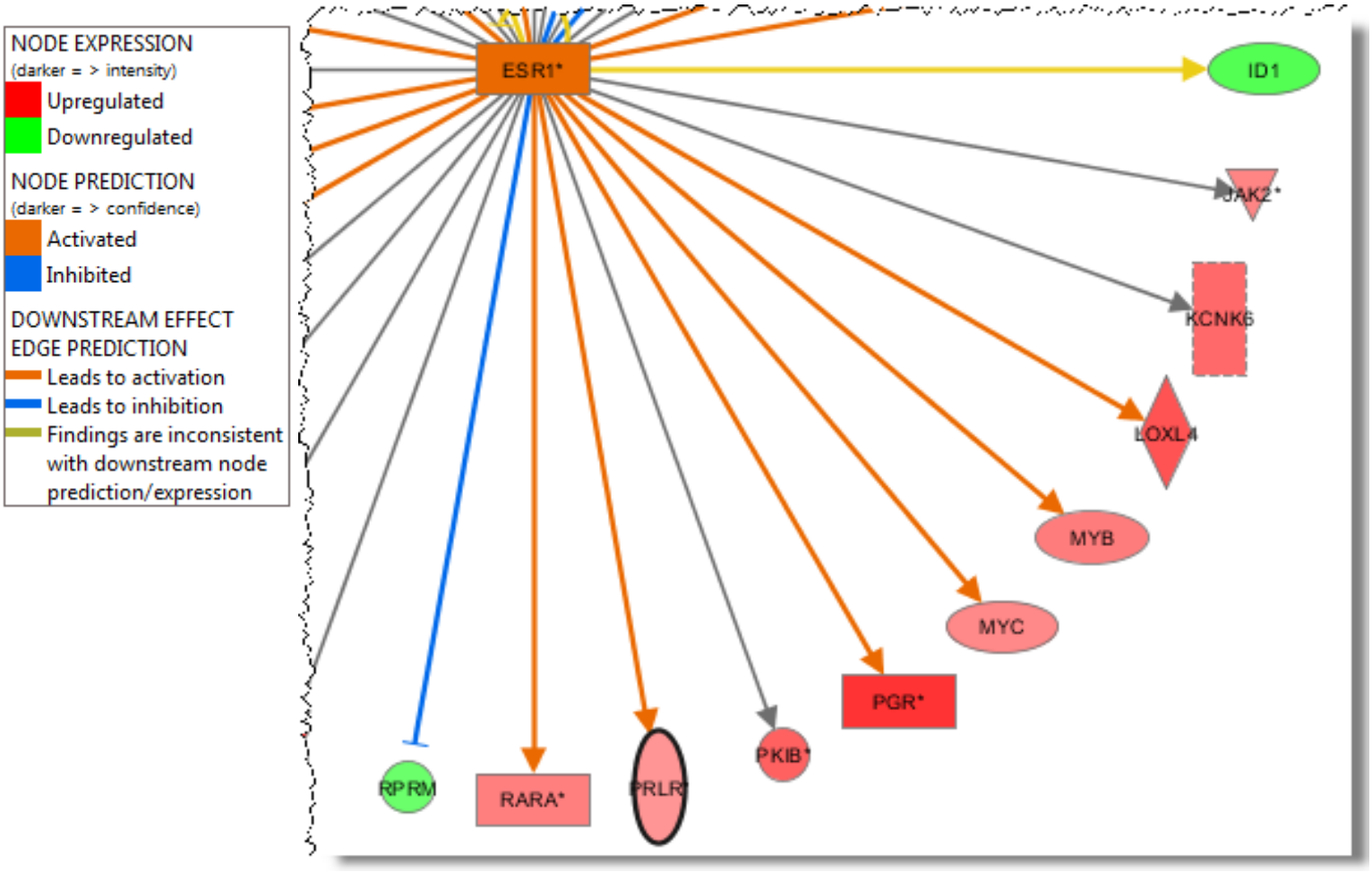
- Use experimentally observed relationships (not predicted binding) between regulators and dataset genes to predict upstream transcriptional regulators.

- Calculate z-score to predict activation or inhibition of regulators based on relationships with dataset genes and direction of change of dataset genes.



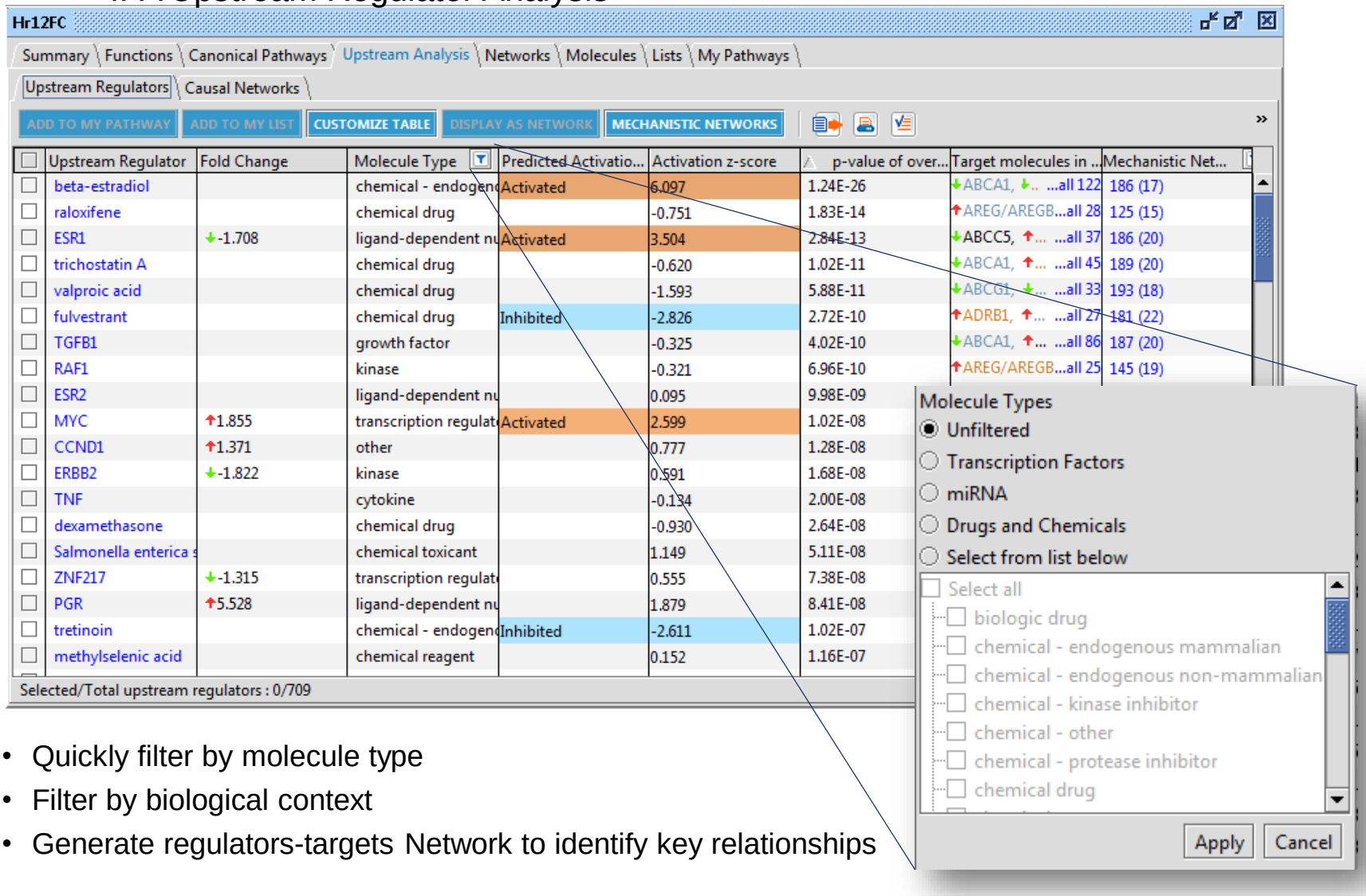
IPA Upstream Regulator Analysis

Directional Effects: Molecule Activity Predictor
Examine Expression Relationship Consistency



IPA Winter Release 2012

IPA Upstream Regulator Analysis



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

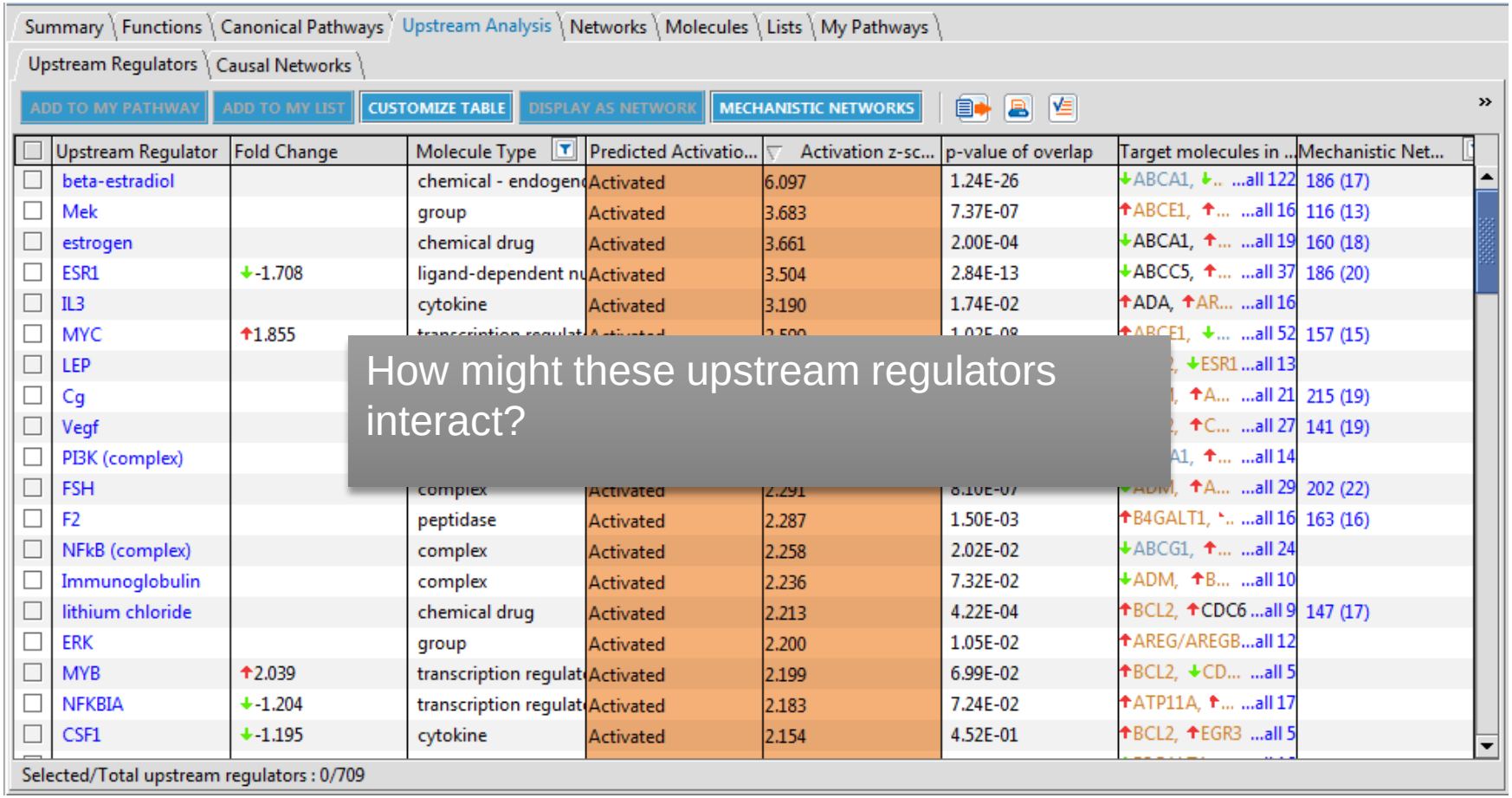
Considerations

- 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



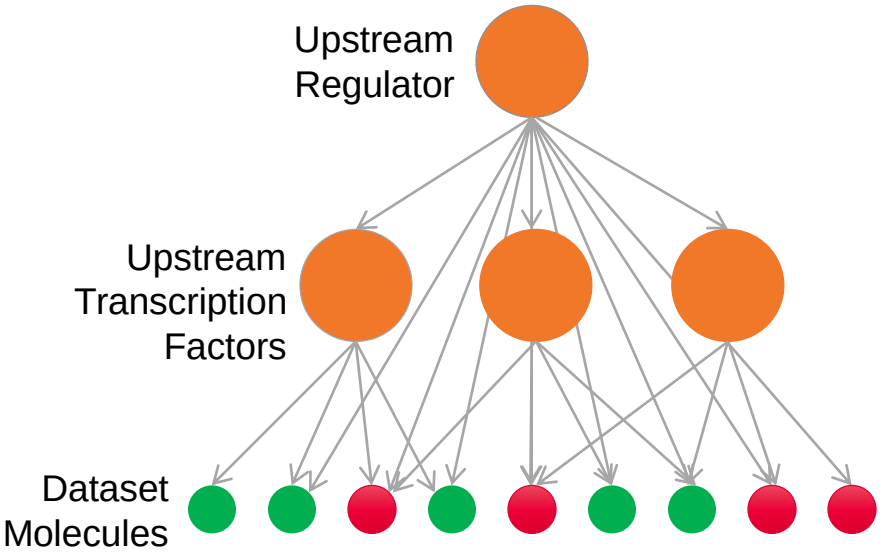
Upstream Regulator	Fold Change	Molecule Type	Predicted Activatio...	Activation z-sc...	p-value of overlap	Target molecules in ...	Mechanistic Net...
beta-estradiol		chemical - endogen	Activated	6.097	1.24E-26	ABCA1, ...all 122	186 (17)
Mek		group	Activated	3.683	7.37E-07	ABCE1, ...all 16	116 (13)
estrogen		chemical drug	Activated	3.661	2.00E-04	ABCA1, ...all 19	160 (18)
ESR1	-1.708	ligand-dependent nu	Activated	3.504	2.84E-13	ABCC5, ...all 37	186 (20)
IL3		cytokine	Activated	3.190	1.74E-02	ADA, AR... ...all 16	
MYC	1.855					...all 52	157 (15)
LEP						ESR1 ...all 13	
Cg						...all 21	215 (19)
Vegf						...all 27	141 (19)
PI3K (complex)						...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 regulat	Activated	2.199	6.99E-02	BCL2, CD... ...all 5	
NFKBIA	-1.204	transcription regulat	Activated	2.183	7.24E-02	ATP11A, ...all 17	
CSF1	-1.195	cytokine	Activated	2.154	4.52E-01	BCL2, EGR3 ...all 5	

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

IPA Mechanistic Networks

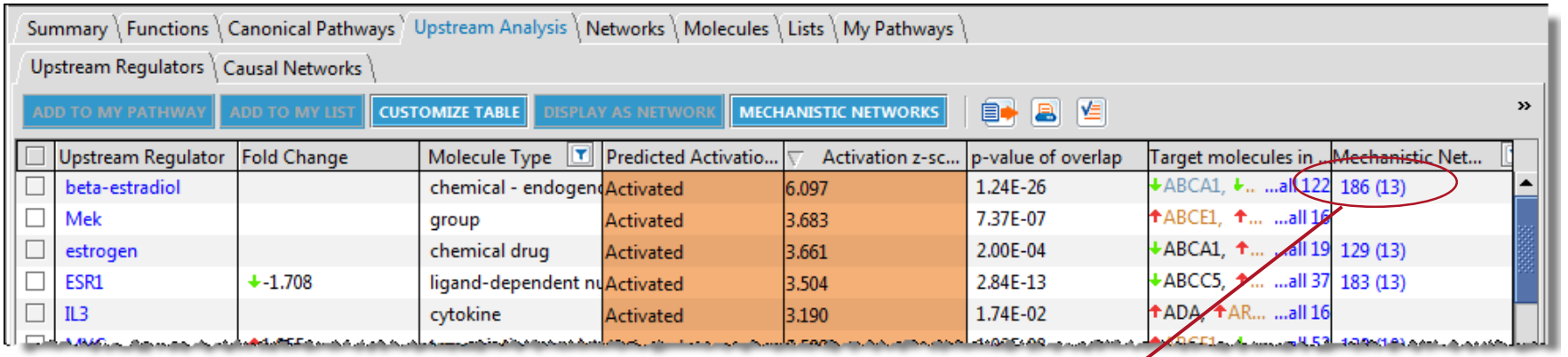
Goal: To discover plausible sets of connected upstream regulators that can work together to elicit the gene expression changes observed in a dataset

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:



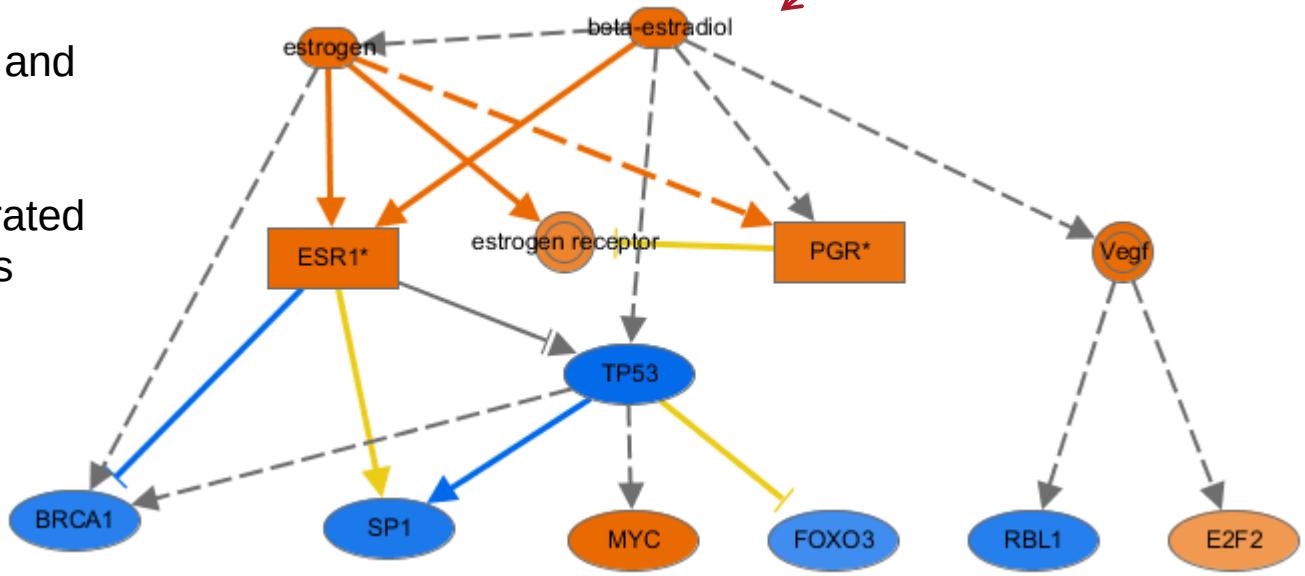
IPA Winter Release 2012

How might the upstream molecule drive the observed expression changes?



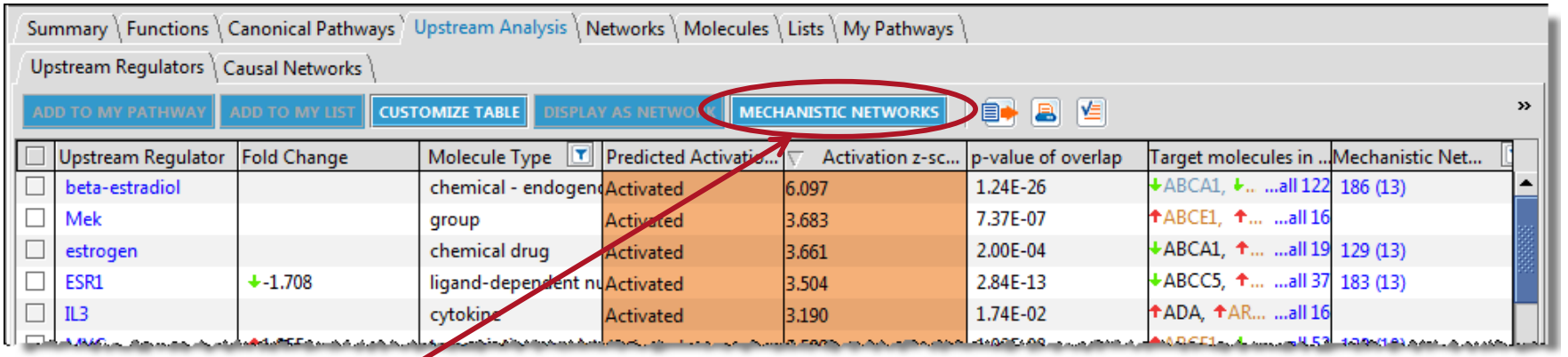
Hypothesis generation and visualization

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

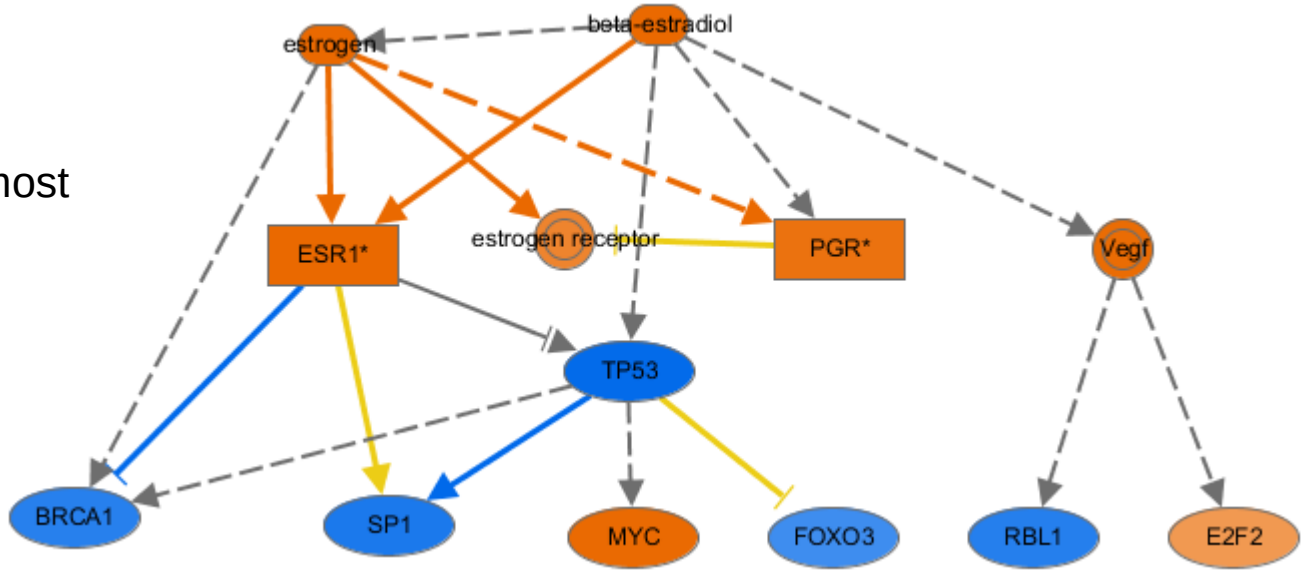


IPA Winter Release 2012

Mechanistic Networks



Recommend increasing stringency of criteria in most cases



Advanced Analytics

Causal Networks

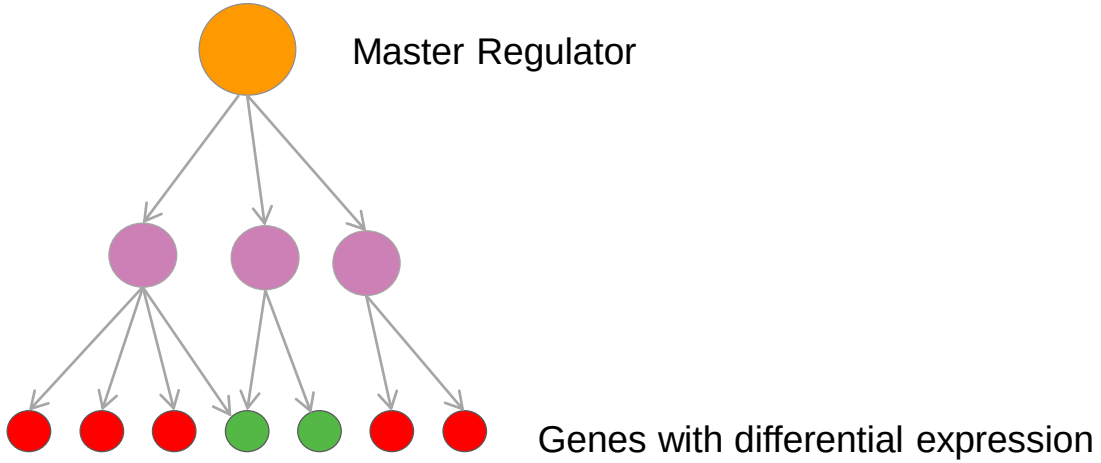
Advanced Analytics requires an additional subscription fee

Causal Network Analysis

Advanced Analytics

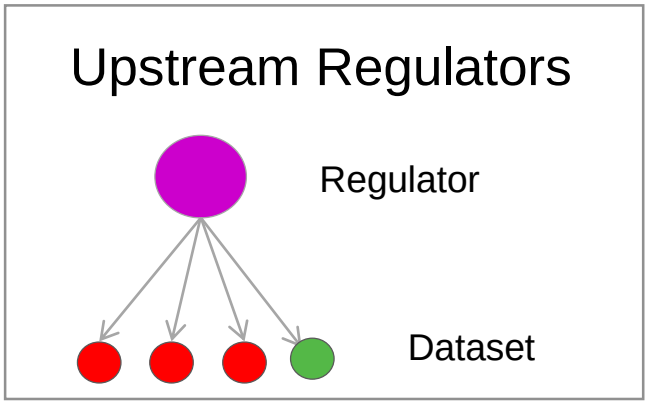
- Alternate method of predicting upstream regulators based on causal relationships and allowing multiple interaction steps to gene expression changes
- Identify potential novel master-regulators of your gene expression by creating pathways of literature-based relationships
- Expands predictions to include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

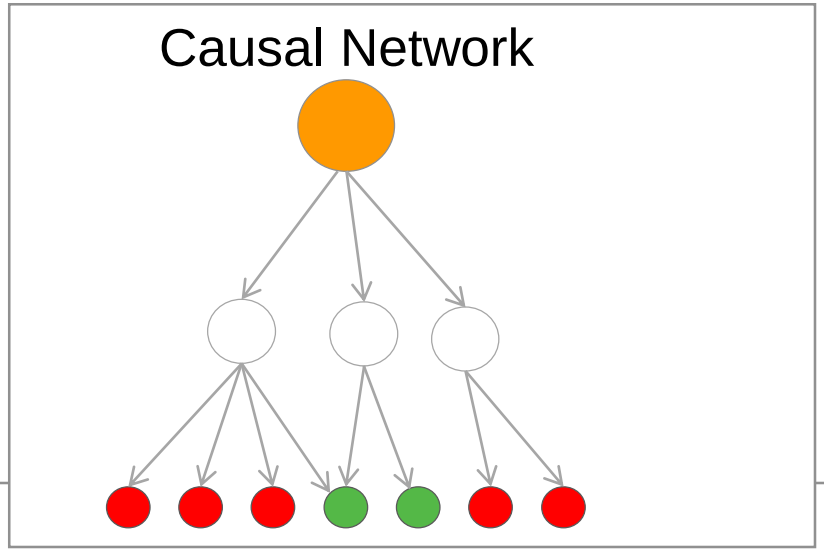
Causal Networks

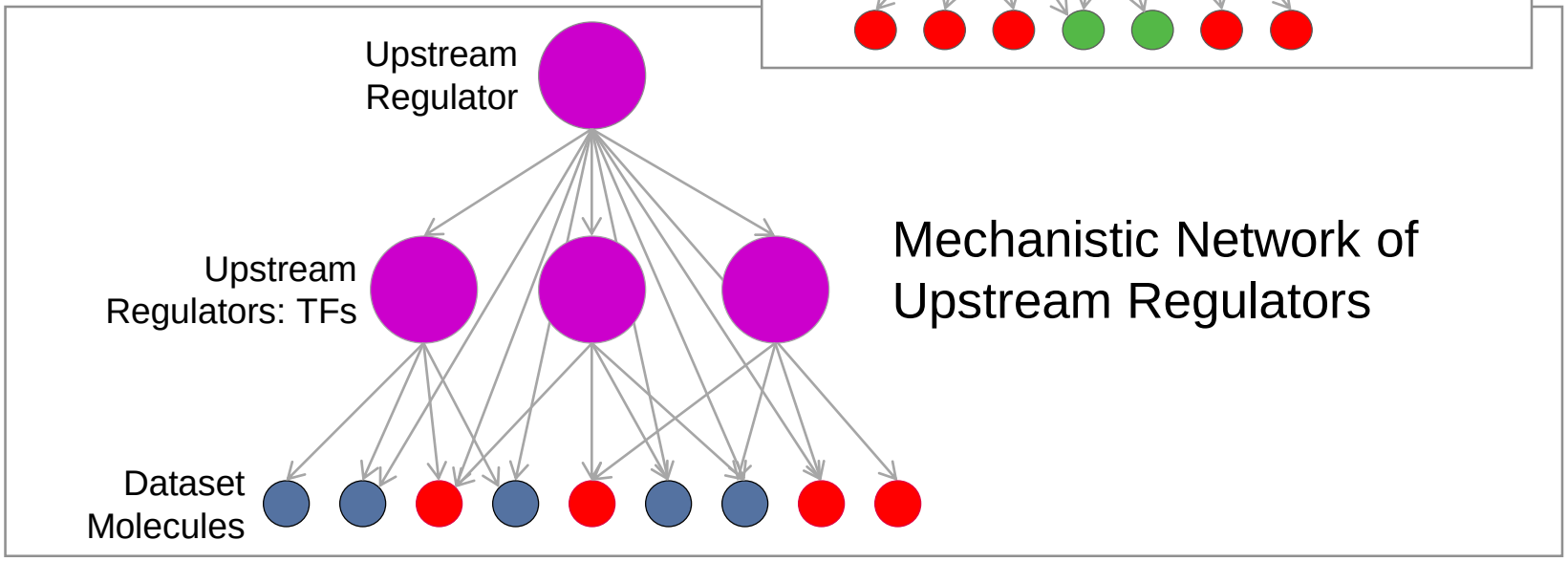


Single- vs. Mechanistic- vs. Causal Networks
Leveraging the network to create more upstream regulators

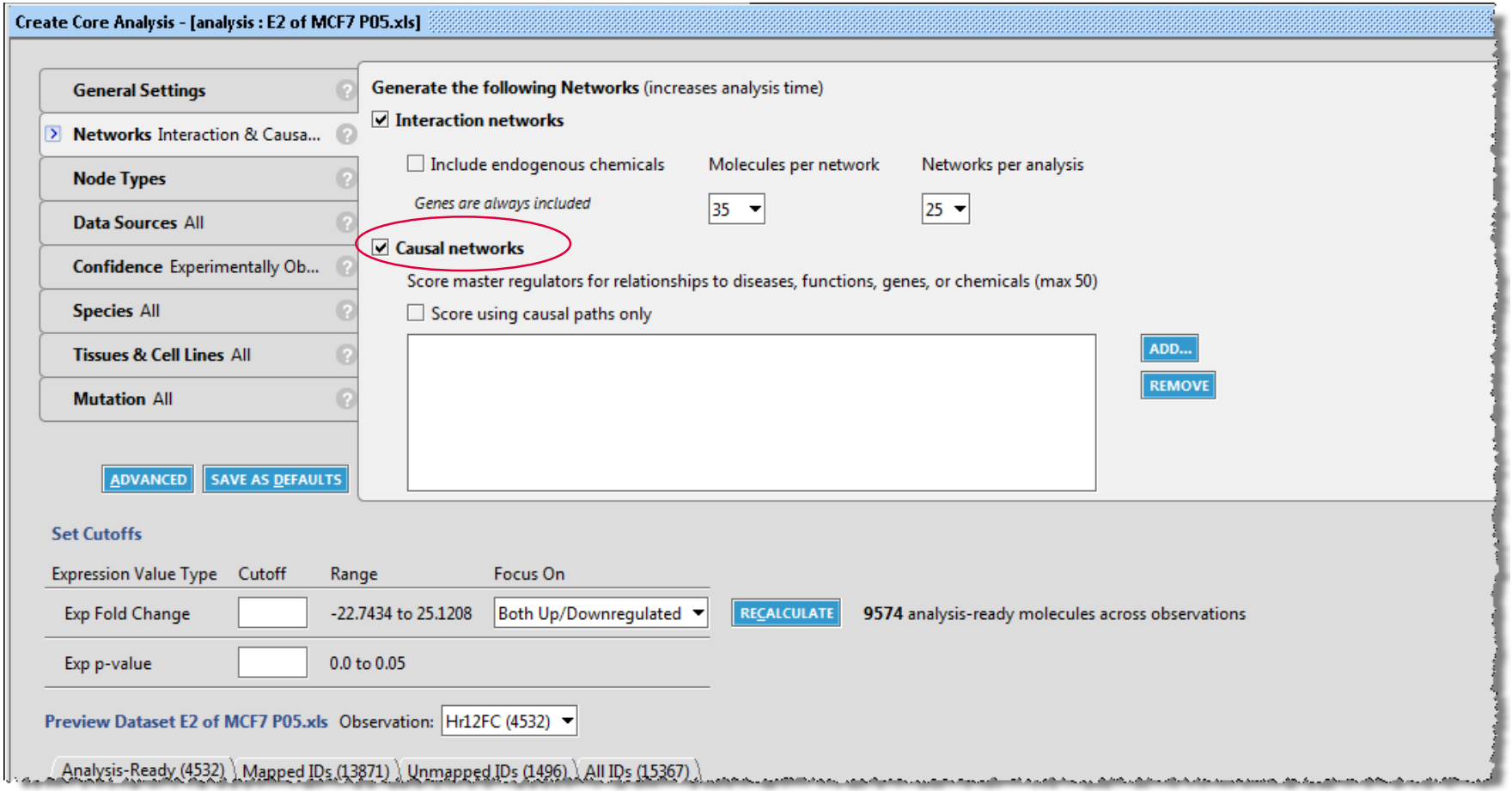
Advanced Analytics: Causal Network Analysis



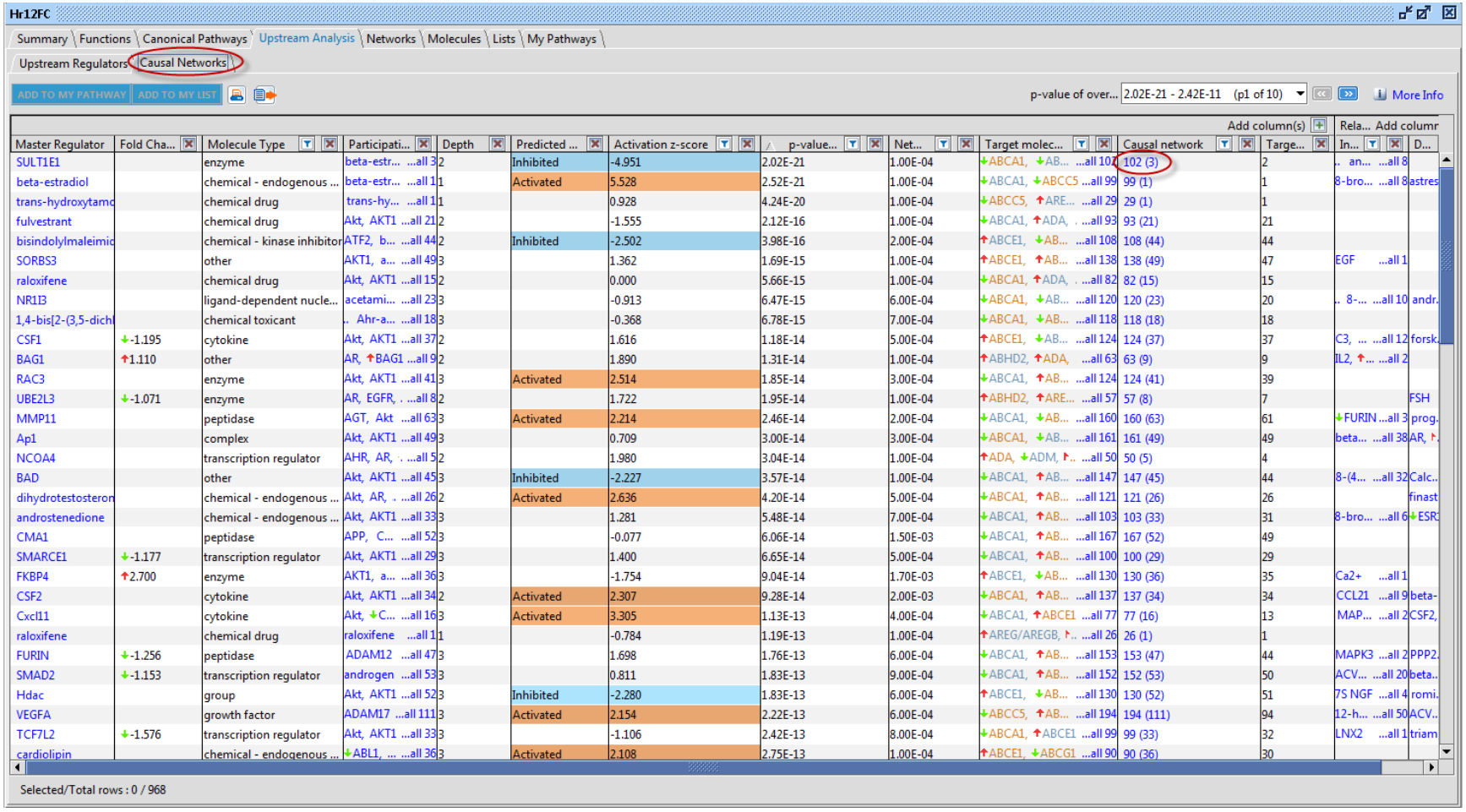




Turning on Causal Networks (with Advanced Analytics)



Advanced Analytics: Causal Network Analysis

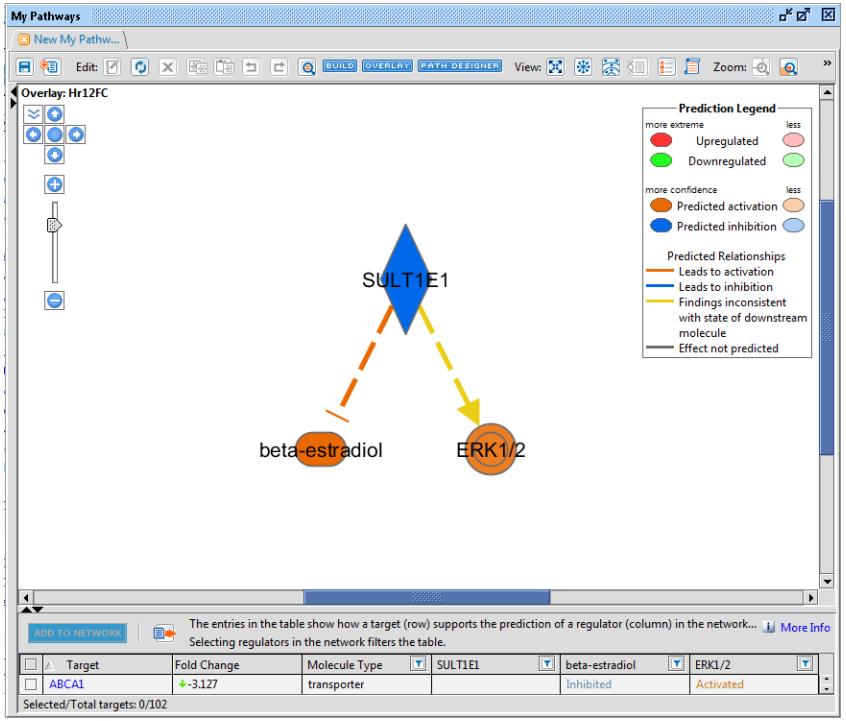


Beta-estradiol of MCF7 cells at 12 hr.
SULT1E1 is top master regulator, but does not appear in upstream regulator table

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

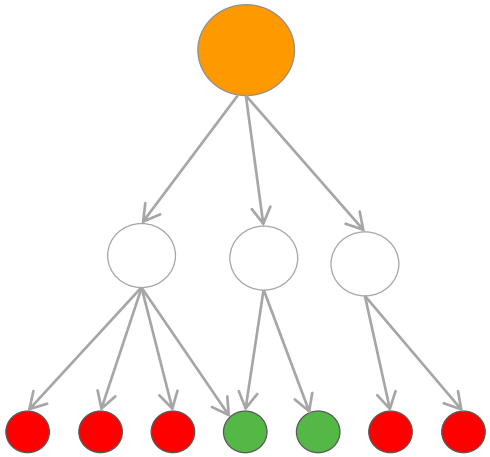
Edges that cannot be assigned a direction of regulation, including all types of binding edges are excluded

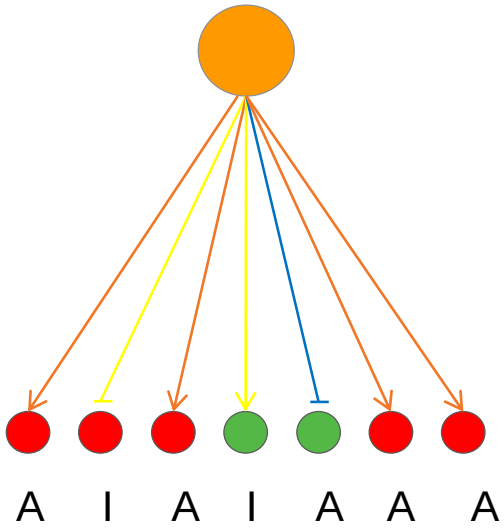
- Included relationship types
  - activation (A)
  - inhibition (I)
  - expression (E)
  - transcription (T)
  - group/complex
  - membership edges (MB, considered activating)
  - phosphorylation (P)

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

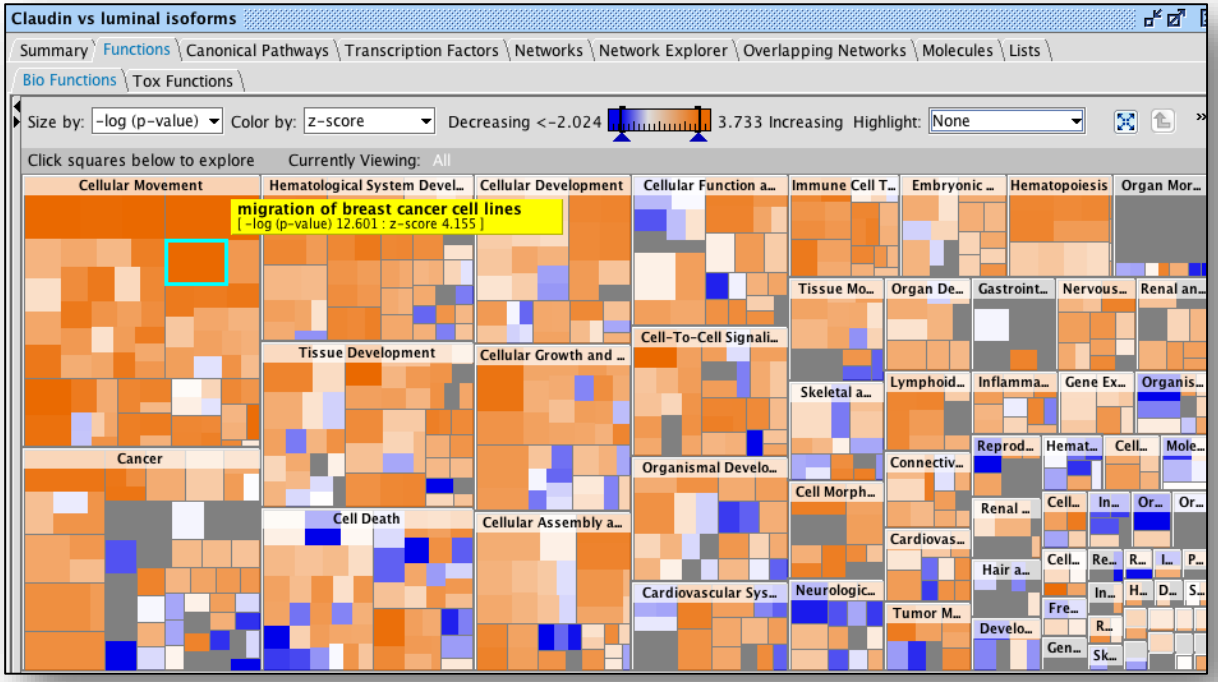
- How are cellular processes are predicted to be changing based on my gene expression data?
- What genes are driving these directional changes?

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

Downstream Effects Analysis

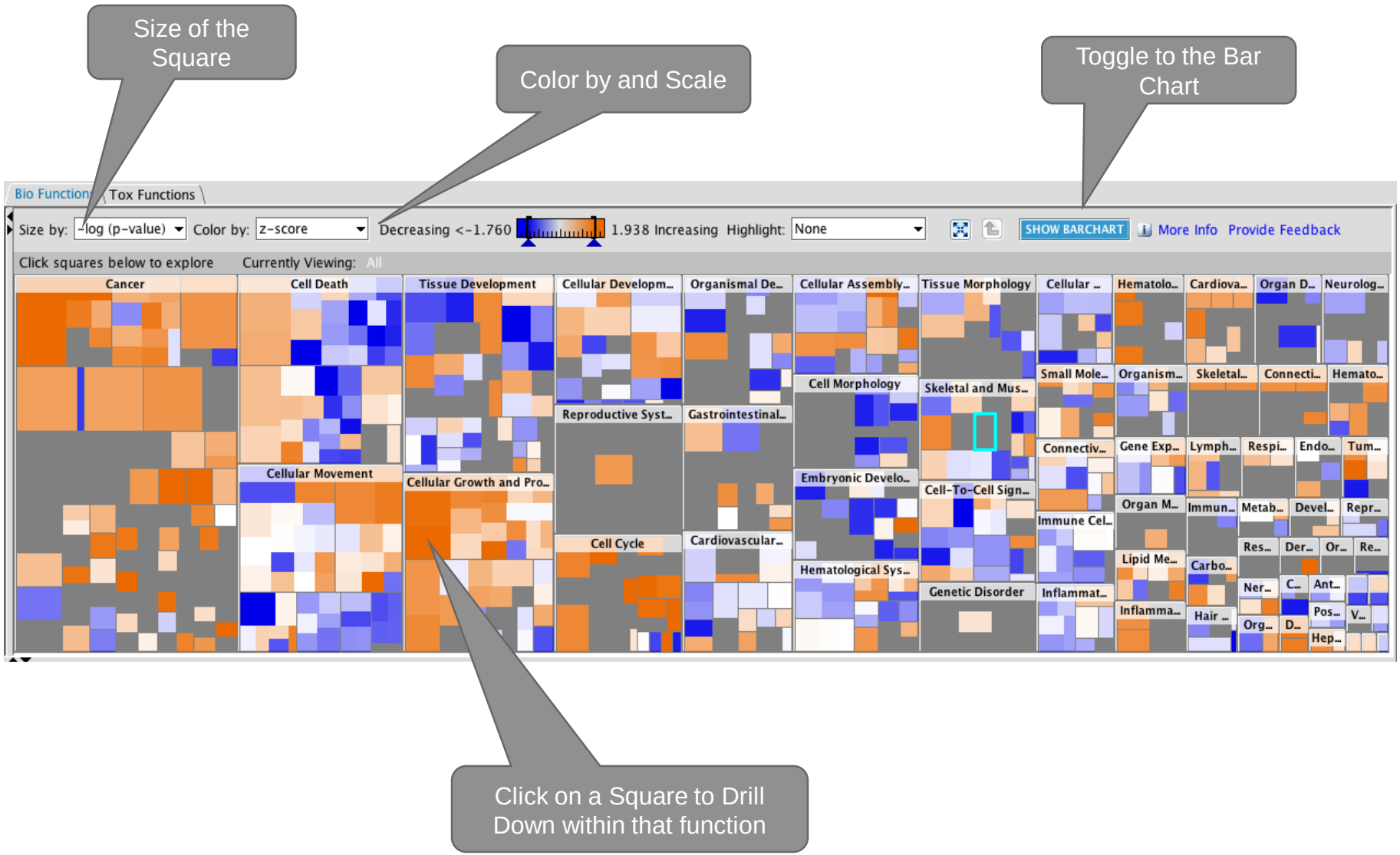
Identify key biological processes influenced by differentially expressed genes

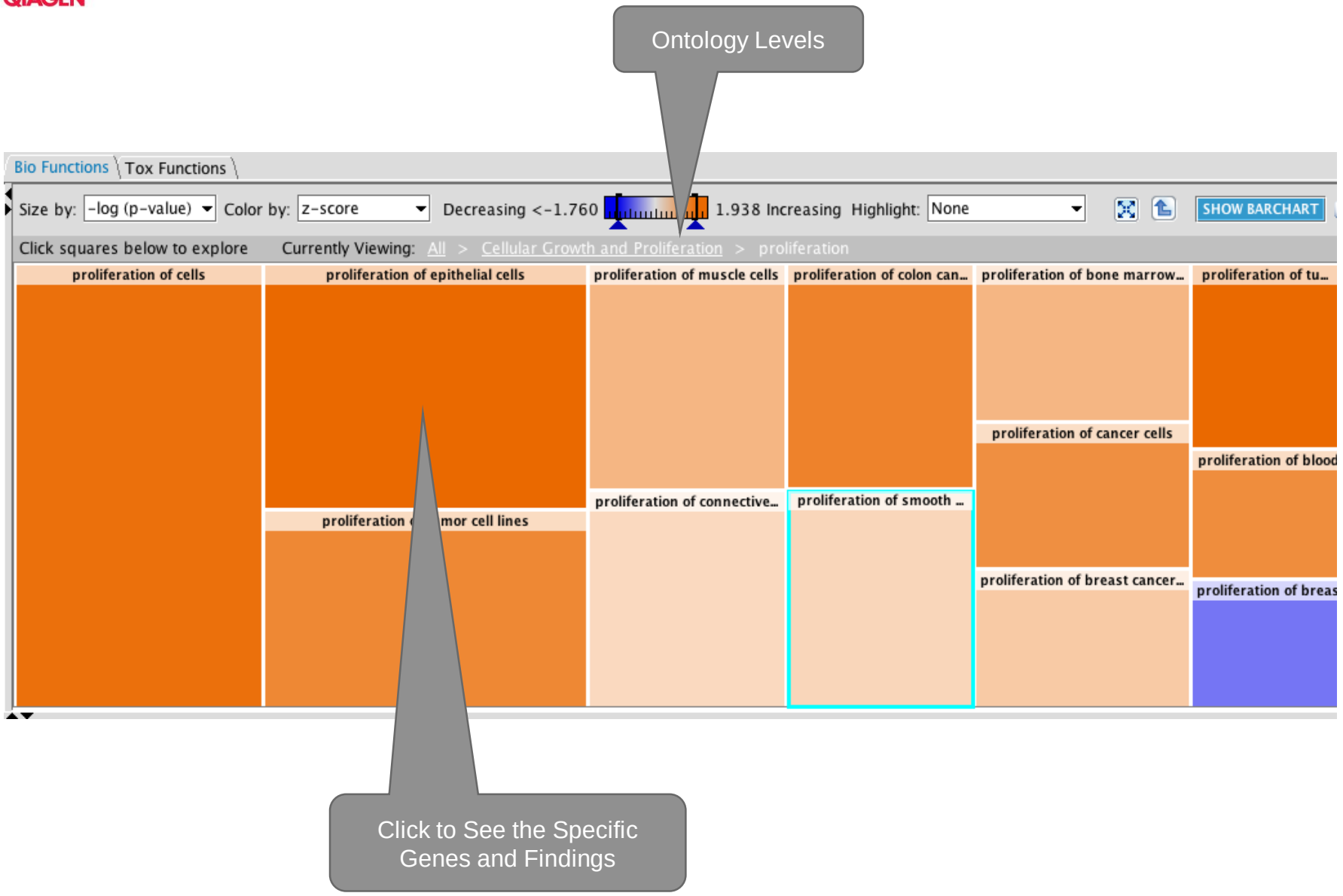
Understand whether cellular processes are being driven up or down by correlating observed expression with reported experimental gene effects

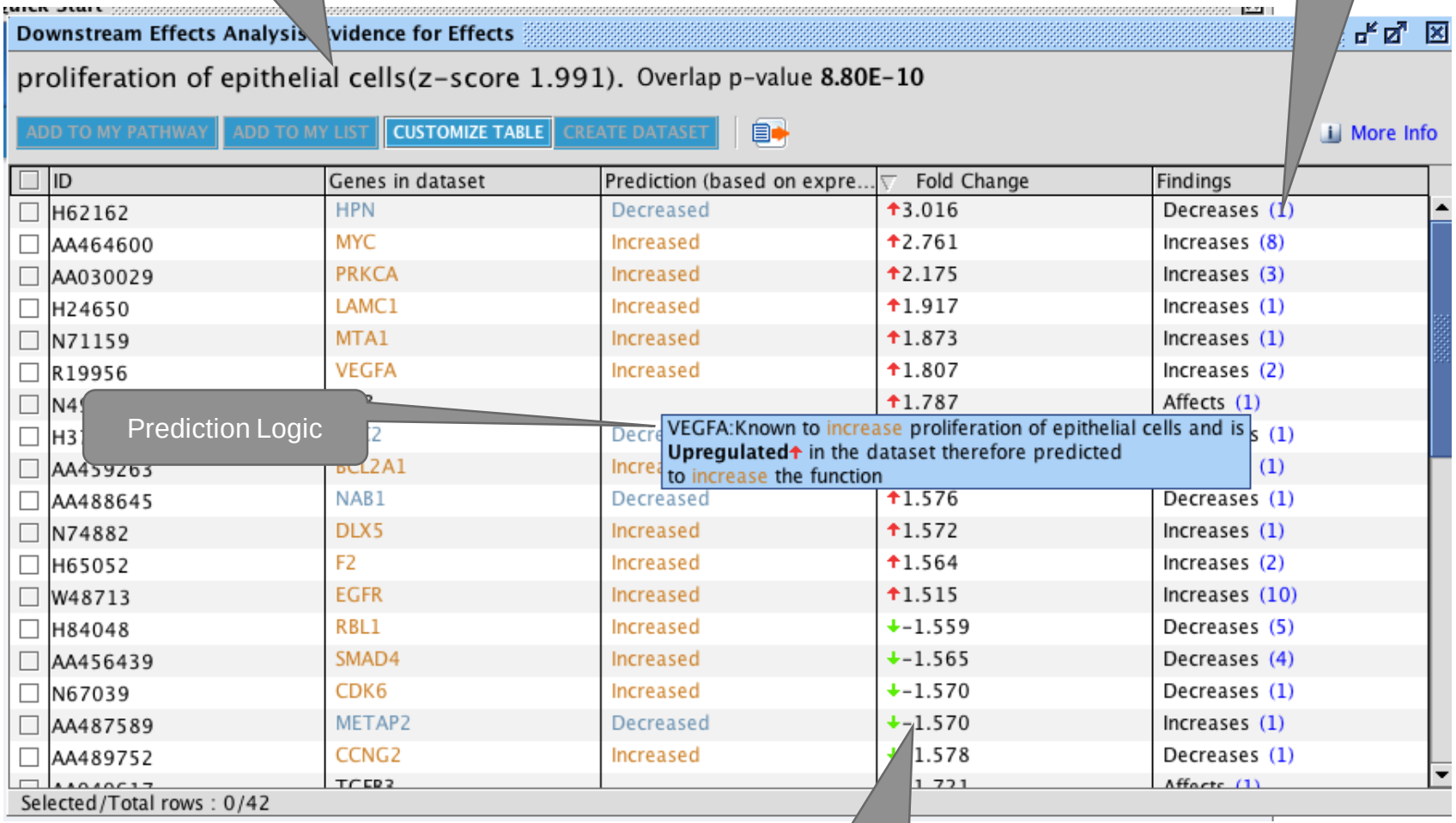


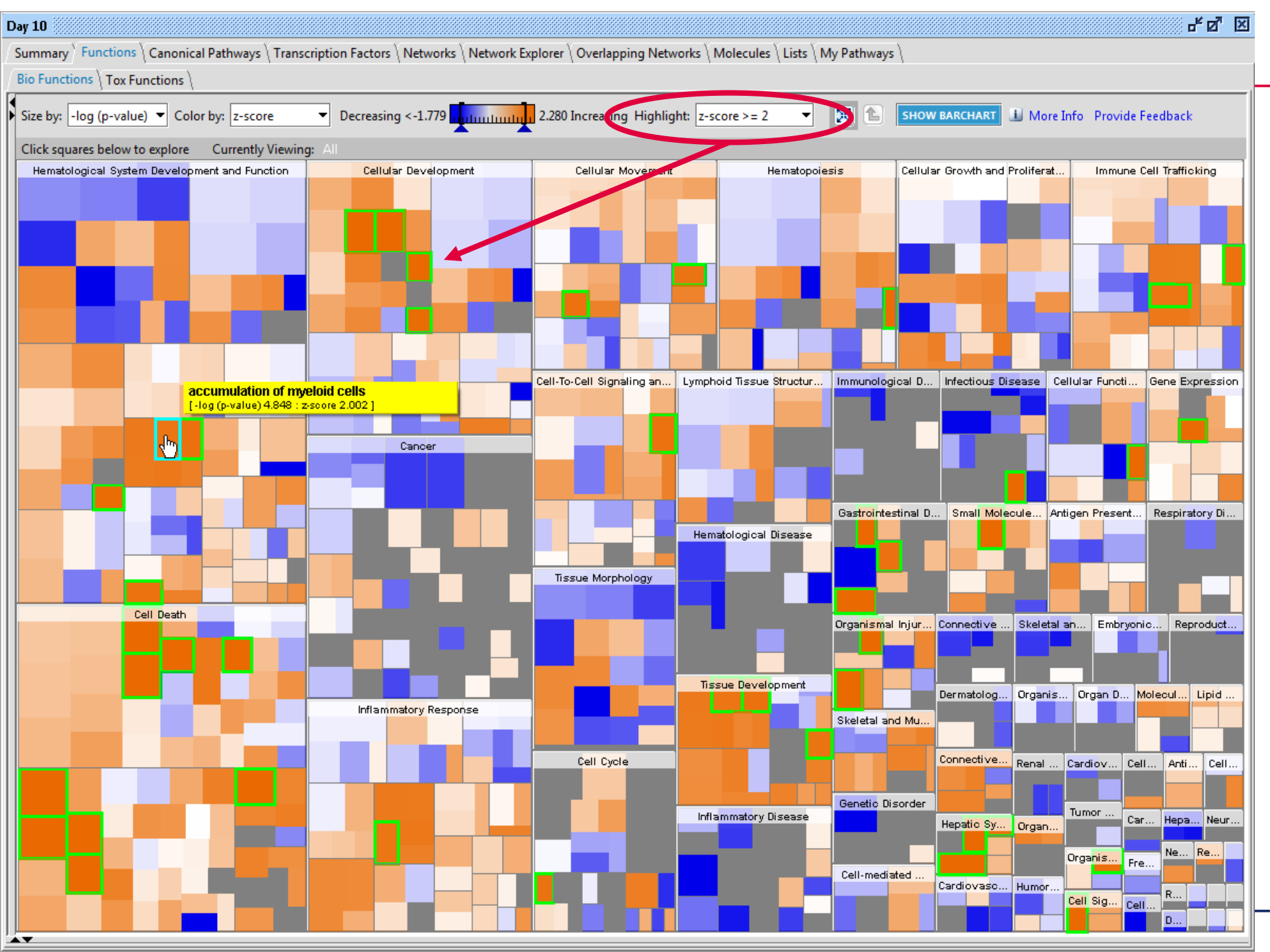
- Each box represents a biological process or disease
- The size of the box represents gene enrichment
- The color of the box indicates the predicted increase or decrease

Downstream Effect on Bio Function




Ontology Levels
Bio Functions
Tox Functions
Size by: -log (p-value)
Color by: z-score
Decreasing <-1.760
1.938 Increasing
Highlight: None
SHOW BARCHART
Click squares below to explore
Currently Viewing: All > Cellular Growth and Proliferation > proliferation
proliferation of cells
proliferation of epithelial cells
proliferation of muscle cells
proliferation of colon can...
proliferation of bone marrow...
proliferation of tu...
proliferation of cancer cells
proliferation of blood
proliferation of connective...
proliferation of smooth ...
proliferation of breast cancer...
proliferation of breas
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 expre...
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)
1.787 Affects (1)
Prediction Logic
VEGFA:Known to increase proliferation of epithelial cells and is Upregulated in the dataset therefore predicted to increase the function
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)
Selected/Total rows : 0/42
Expression Value in Your Dataset


Day 10
Summary
Functions
Canonical Pathways
Transcription Factors
Networks
Network Explorer
Overlapping Networks
Molecules
Lists
My Pathways
Bio Functions
Tox Functions
Size by: -log (p-value)
Color by: z-score
Decreasing <-1.779
2.280 Increasing
Highlight: z-score >= 2
SHOW BARCHART
More Info
Provide Feedback
Click squares below to explore
Currently Viewing: All
Hematological System Development and Function
Cellular Development
Cellular Movement
Hematopoiesis
Cellular Growth and Proliferat...
Immune Cell Trafficking
accumulation of myeloid cells
[-log (p-value) 4.848 : z-score 2.002]
Cell-To-Cell Signaling an...
Lymphoid Tissue Structur...
Immunological D...
Infectious Disease
Cellular Functi...
Gene Expression
Cancer
Gastrointestinal D...
Small Molecule...
Antigen Present...
Respiratory Di...
Hematological Disease
Tissue Morphology
Cell Death
Organismal Injur...
Connective ...
Skeletal an...
Embryonic...
Reproduct...
Tissue Development
Dermatolog...
Organis...
Organ D...
Molecul...
Lipid ...
Inflammatory Response
Skeletal and Mu...
Cell Cycle
Connective...
Renal ...
Cardiov...
Cell...
Anti...
Cell...
Genetic Disorder
Tumor ...
Inflammatory Disease
Hepatic Sy...
Organ...
Car...
Hepa...
Neur...
Organis...
Fre...
Ne...
Re...
Cell-mediated ...
Cardiovasc...
Humor...
Cell Sig...
Cell...
R...
D...

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 :

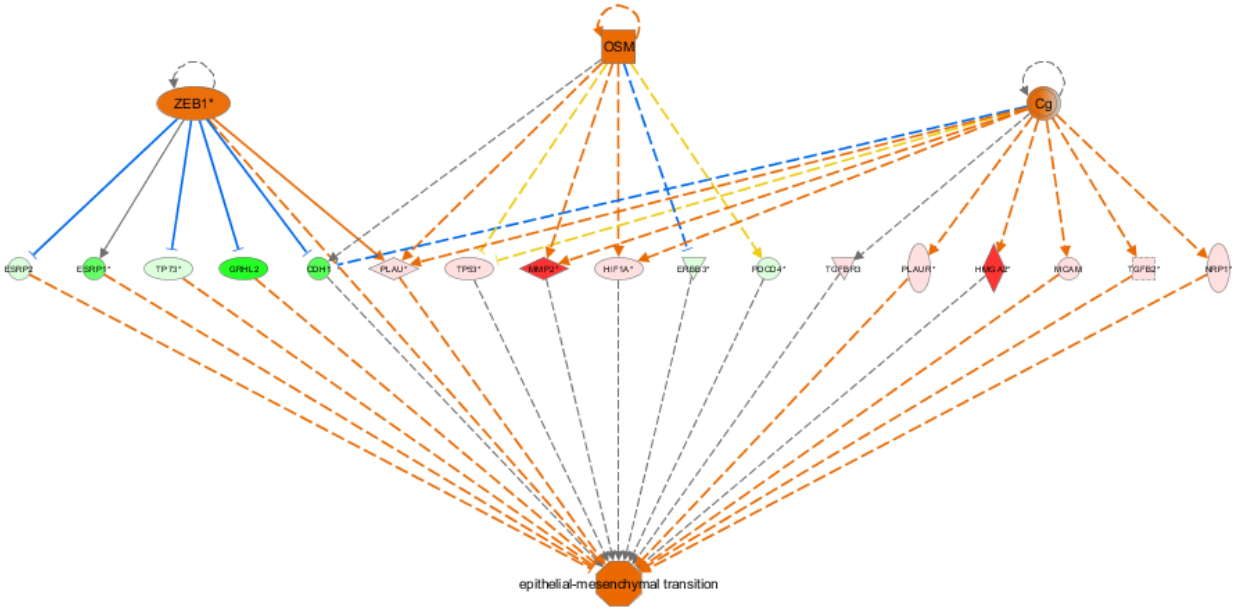
- sent to a pathway for building and/or overlay analysis
- saved as a new Data Set and sent to Core Analysis for additional categorization and segmentation

Analyzing Results

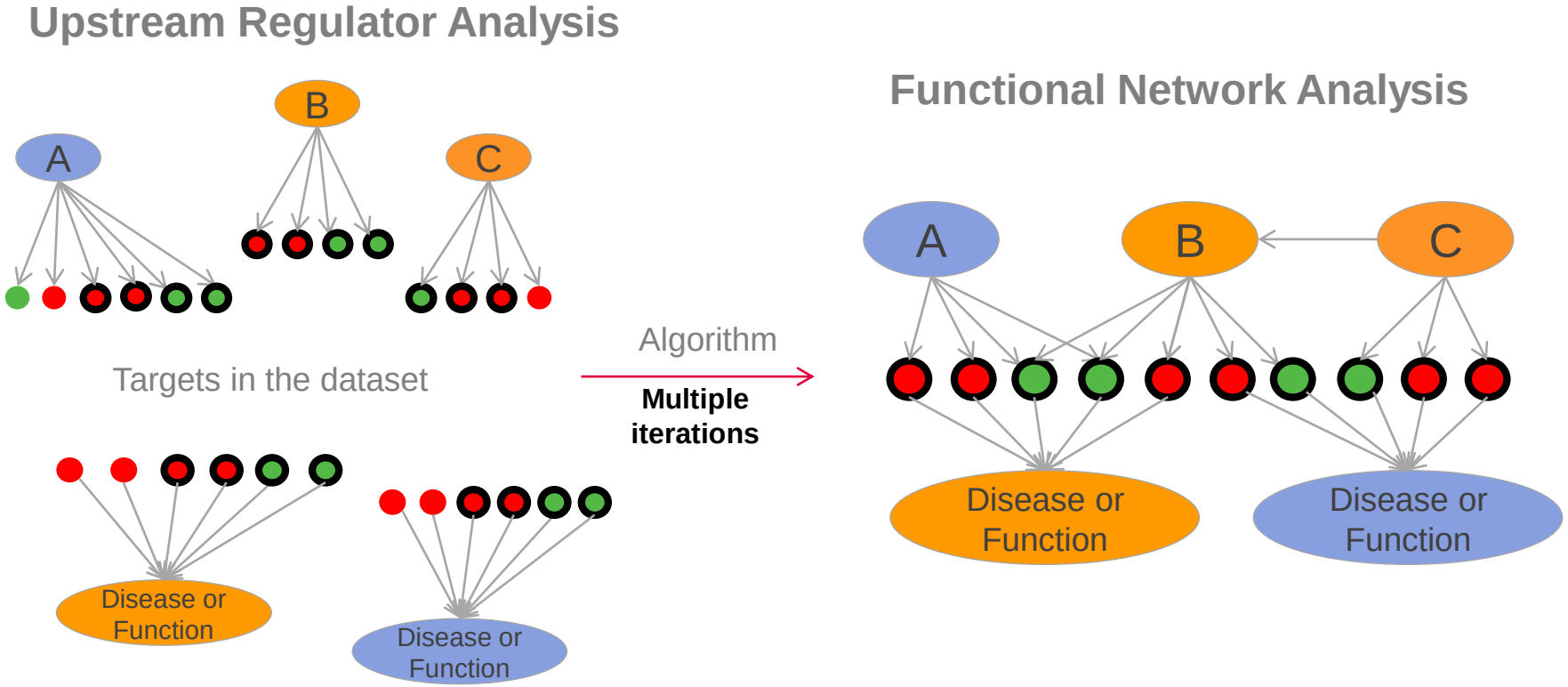
Regulator Effects

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



Concept of Functional Network Analysis



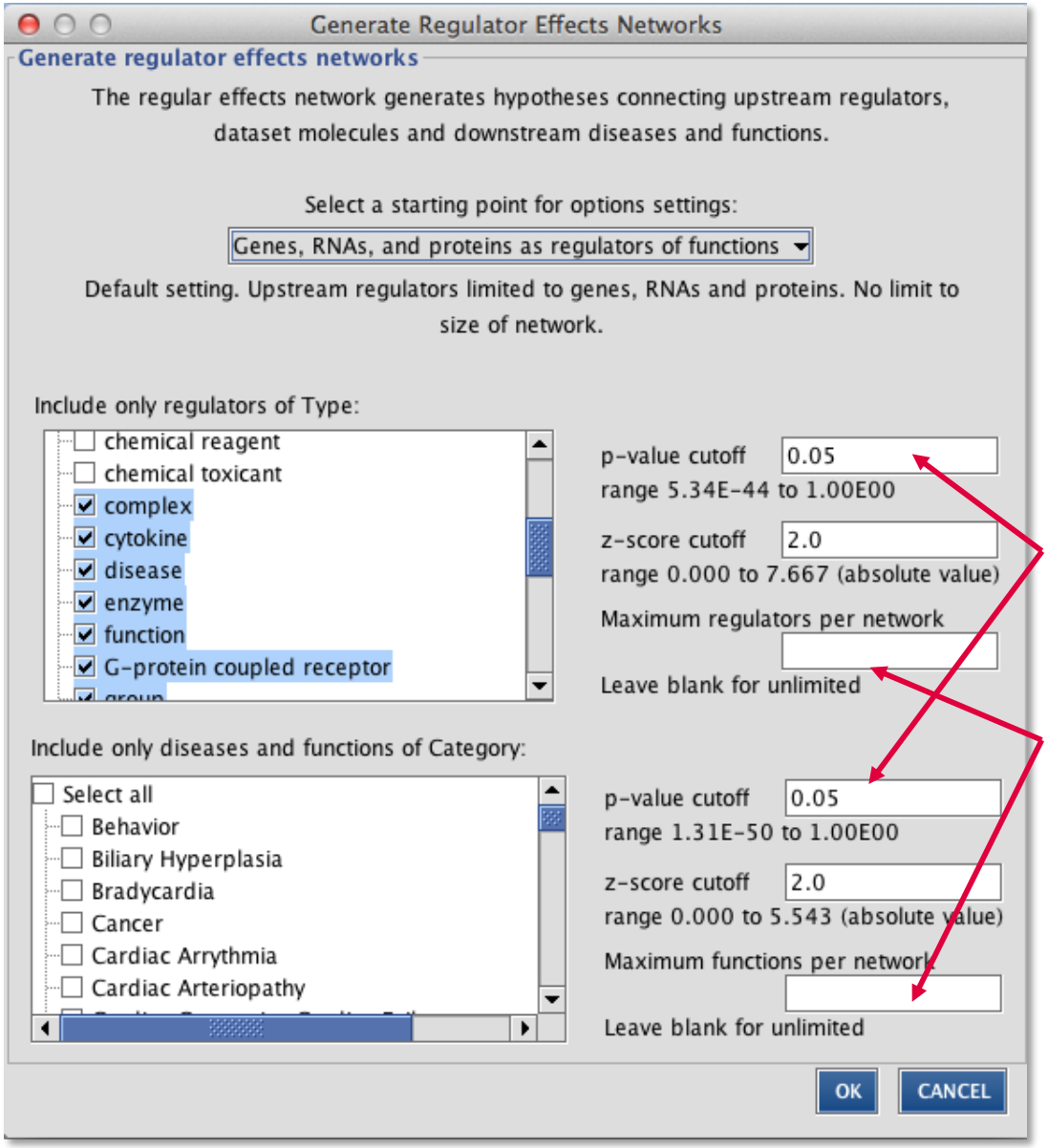
Customize Regulator Effects to answer your question

Default:

Genes, RNAs, proteins as regulators

vs.

Any type of disease or function

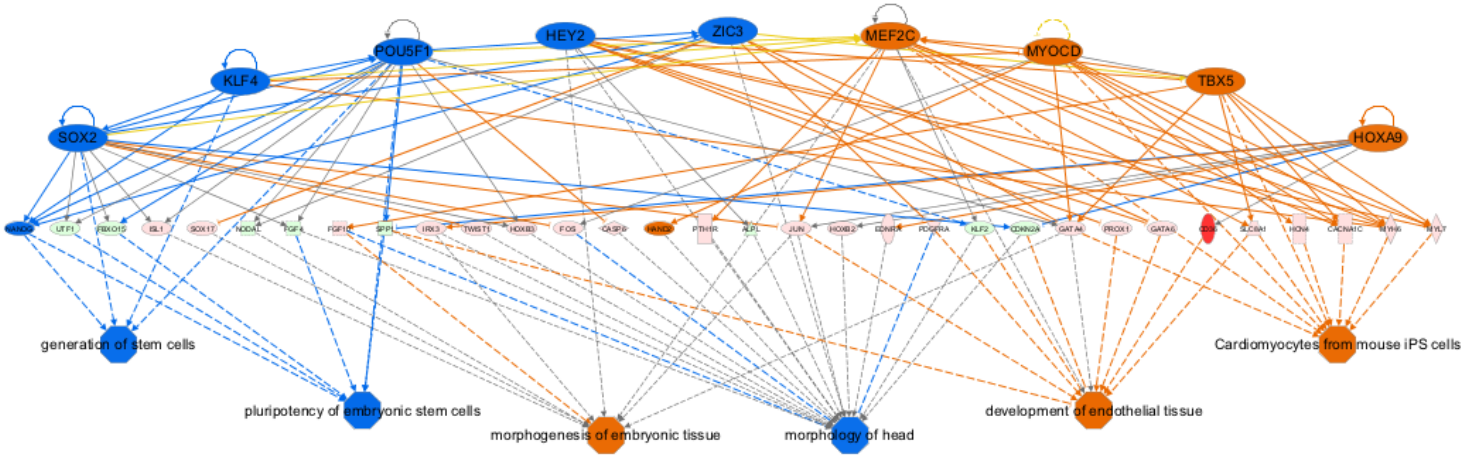


Recommend decreasing p-value cutoffs to 1.0E-3 or lower in most cases.

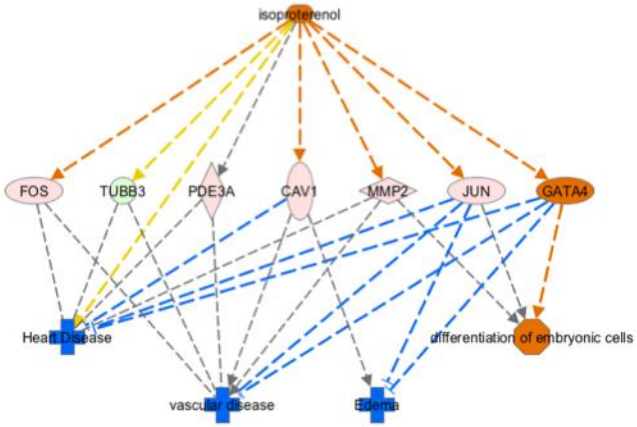
Recommend setting network size to 1 regulator and 1 function in addition to viewing with no value

Alternate presets to answer different questions

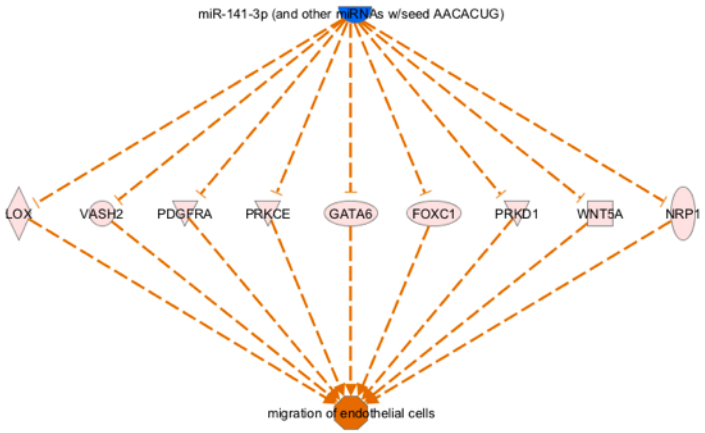
Transcription factors and functions



Compound as regulator of functions



Minimal regulator to function networks



Analyzing Results

Networks

Networks in IPA

- 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

- 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

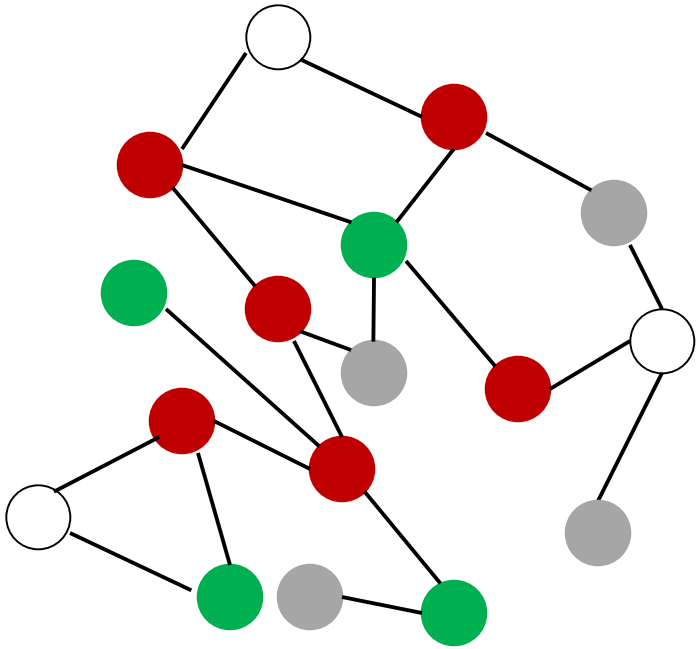
Focus molecules are “seeds”

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 Val

1700 Seaport Blvd., 3rd Floor

Redwood City

CA 94063, USA

INGENUITY® PATHWAY ANALYSIS

INGENUITY® iREPORT

INGENUITY® VARIANT ANALYSIS

Help for IPA

For Help and Technical Support contact our Customer Support team by email to 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

To view our future scientific seminars, and to watch the series archive visit 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

IPA training videos: Search & explore

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/yEJjqlUM4So

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
- Comparison analyses http://youtu.be/JCanWpyfvQE
- Analysis results http://youtu.be/rrppI9OGPUY
- Statistical calculation http://www.youtube.com/watch?v=0oxCQ9dOQlE
- Canonical pathways http://youtu.be/6iZdD9OjIl0
- 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
- Biomarker filter and comparison analysis http://youtu.be/XQFUy0s6wCU
- MicroRNA target filter http://www.youtube.com/embed/06xoKQL9-KA

IPA tutorials

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

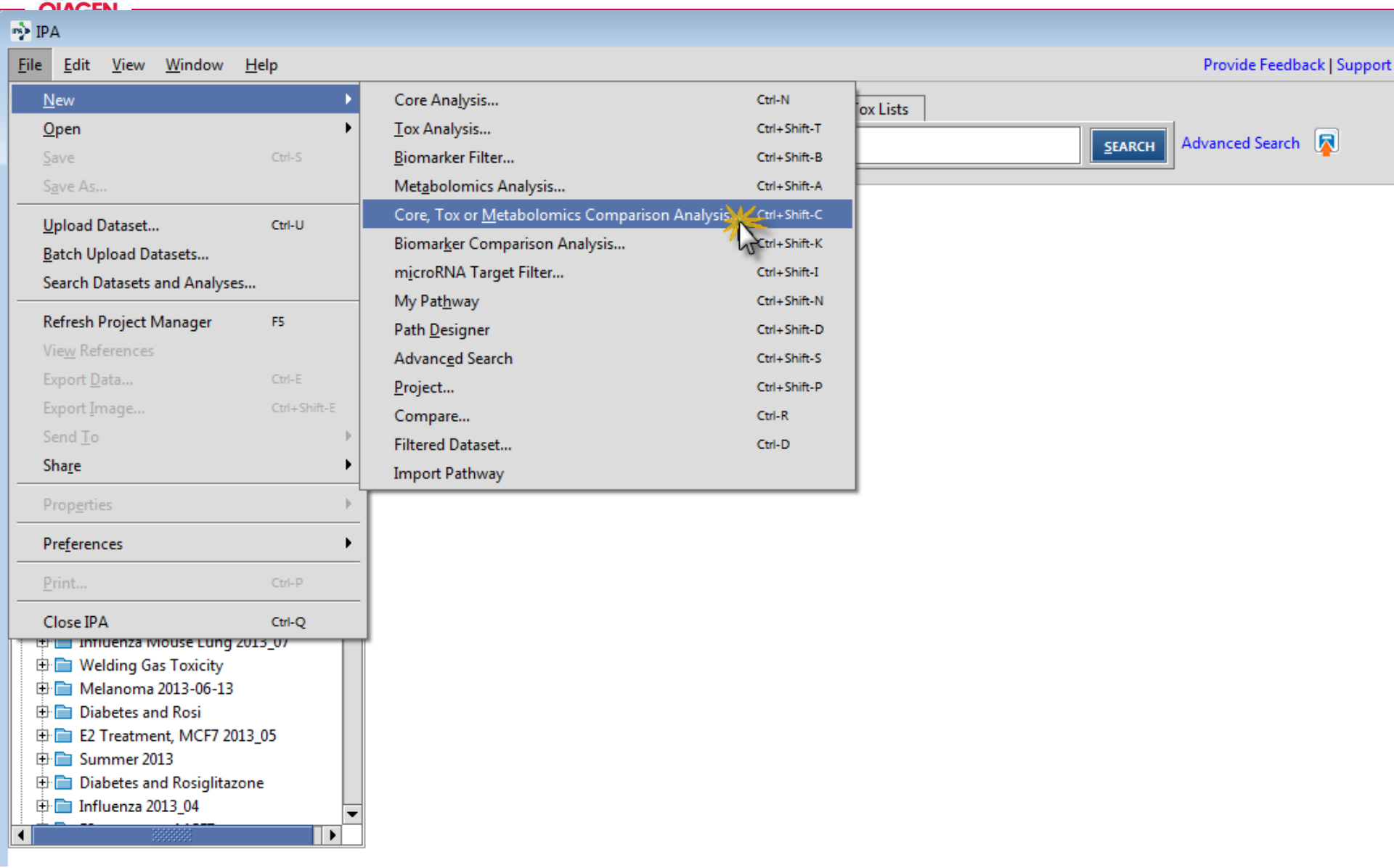
- Search for genes tutorial
- Analysis results tutorial
- Upload and analyze example eata tutorial
- Upload and analyze your own expression data tutorial
- Visualize connections among genes tutorial
- Learn about specialized features
- Human isoforms view tutorial
- Transcription factor analysis tutorial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

Comparing Core Analyses

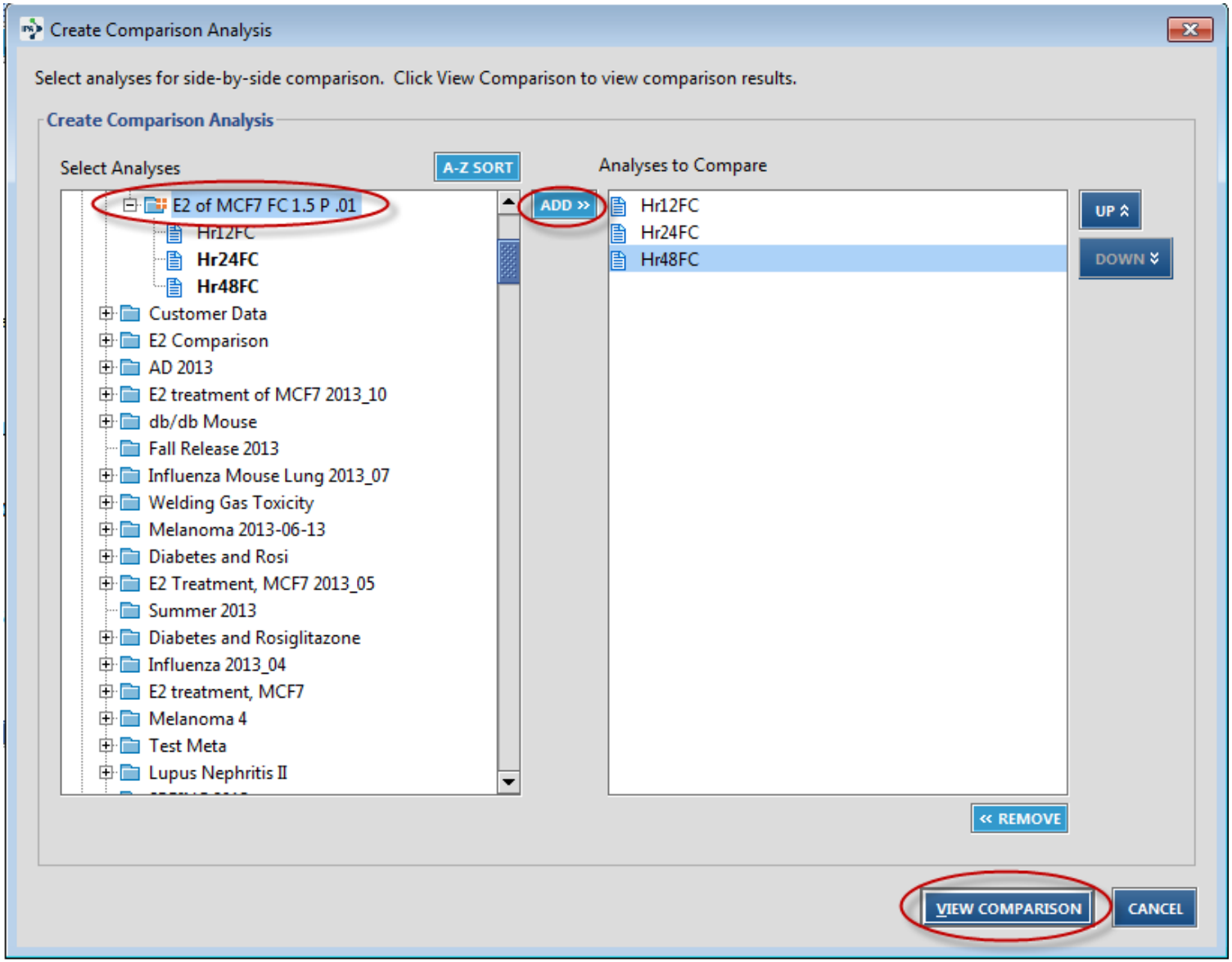
Comparison Analysis

- 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



Selecting Analyses to Compare

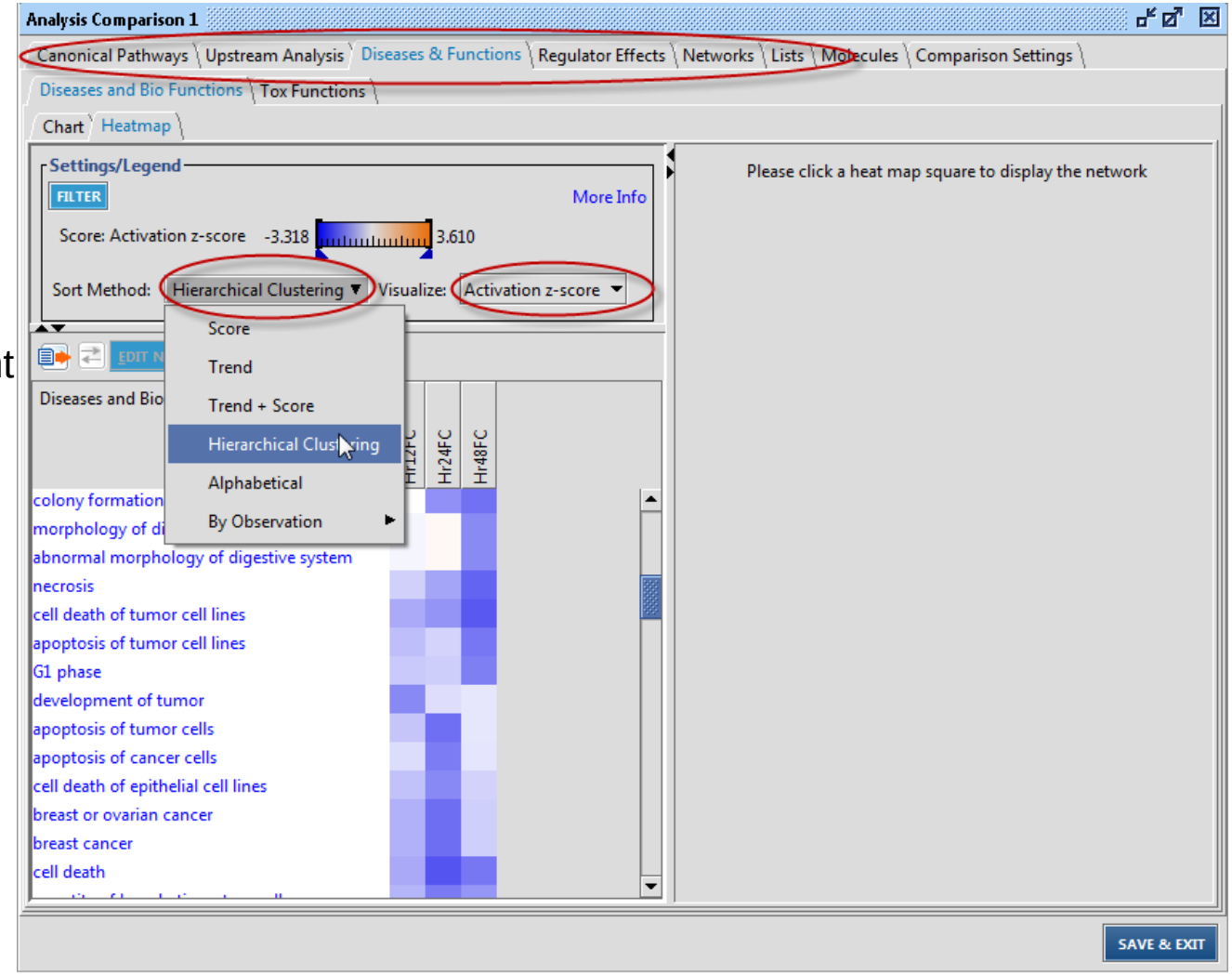


Analyzing Comparison Results

Tabs at top navigate to the analysis-type of interest

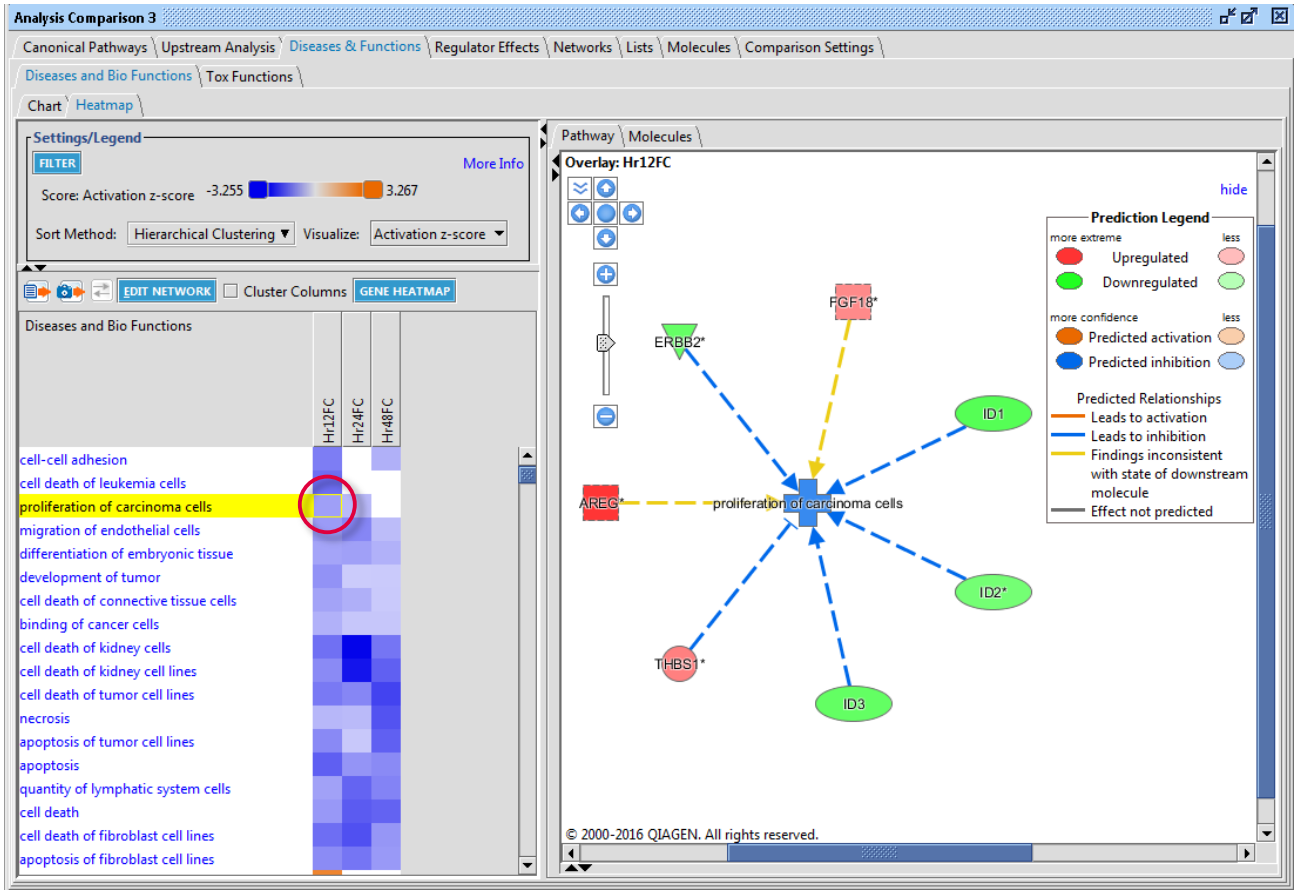
Heatmap can be generated using different calculation methods

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



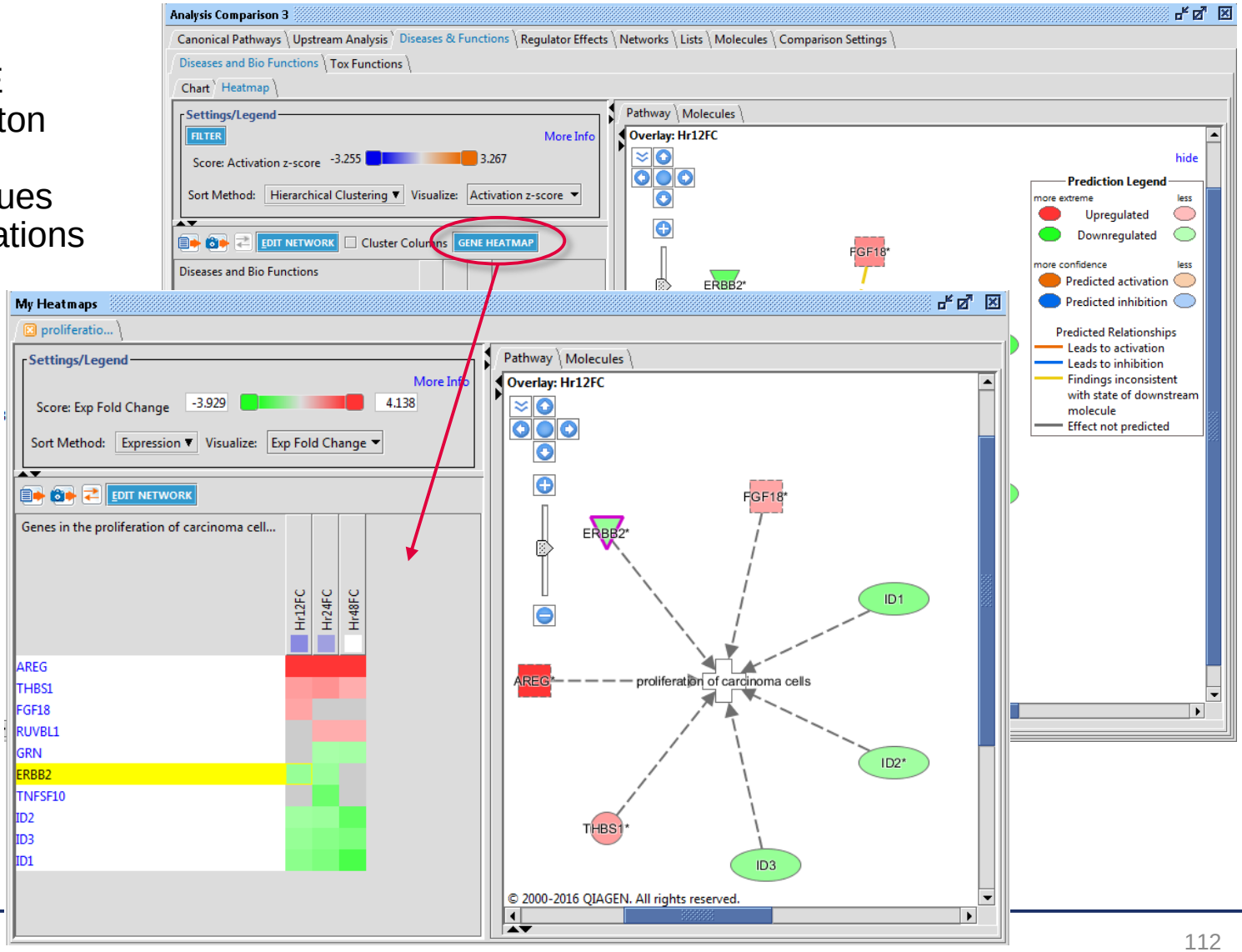
Analyzing Comparison Results

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

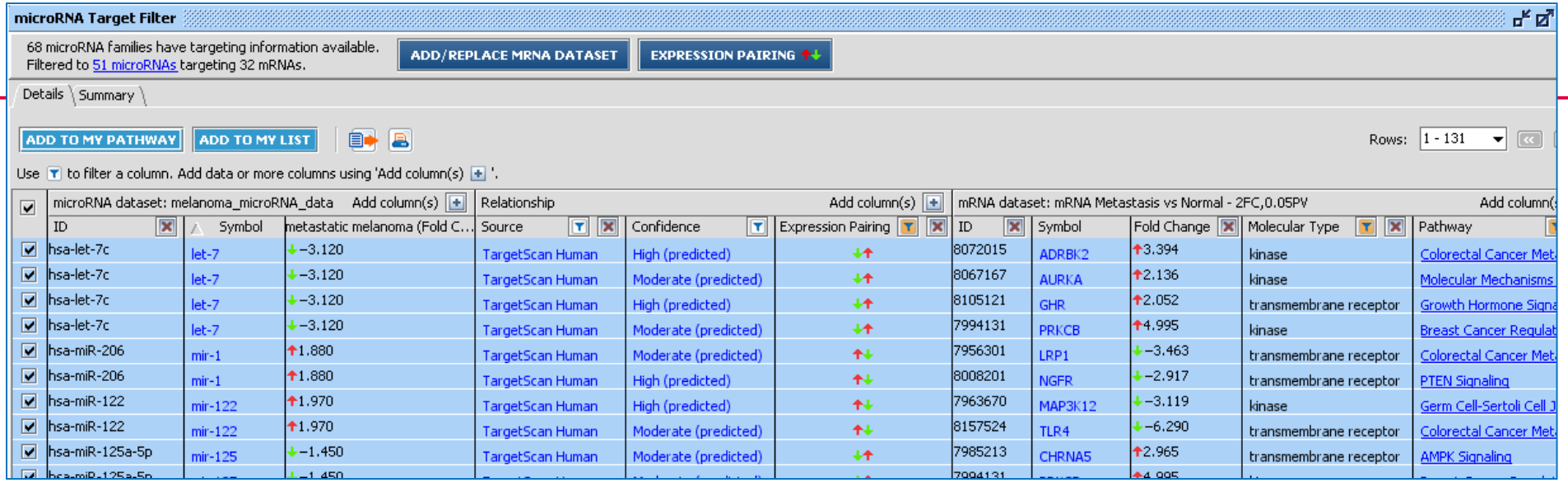
Workflow for miRNA Target Filter

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.



88 data points	13,690 targets	1,090 targets	333 targets	39 targets	32 targets

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

microRNA Content Details:

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

miRNA Naming and Mapping

- 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)

miRNA Naming and Mapping

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.

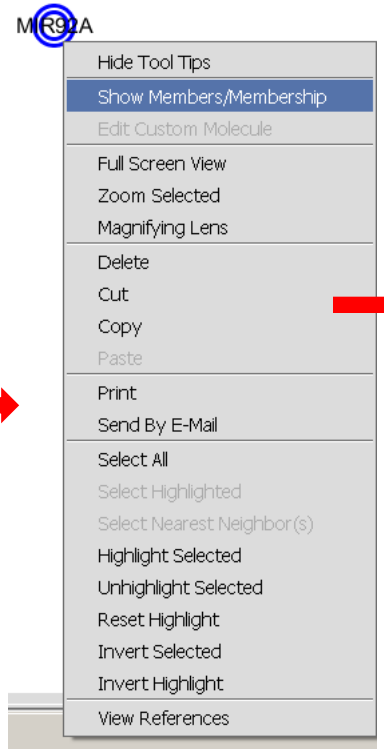
Working with miRNA Groups

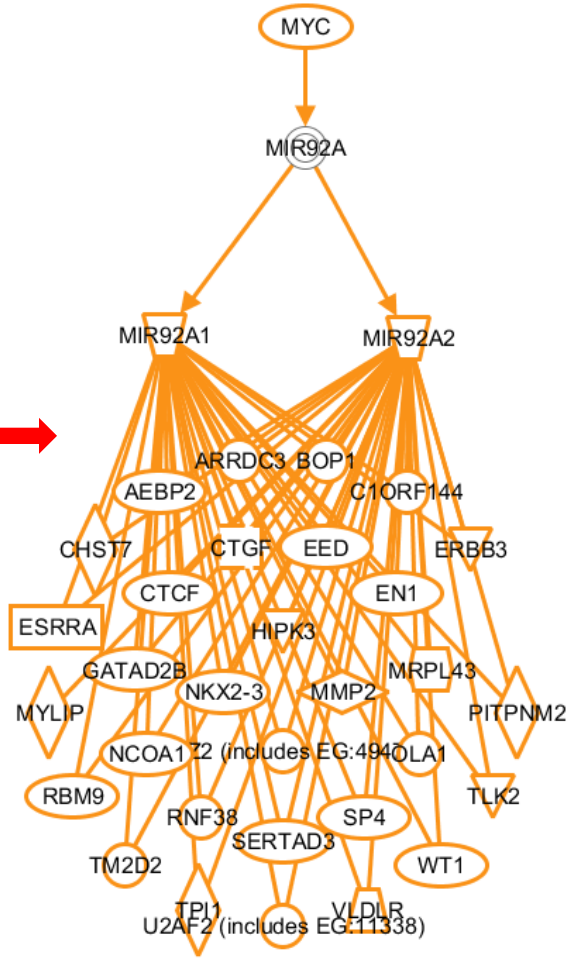
- Mature miRNAs may arise from multiple precursors:
  - A given mature form may arise from multiple distinct miRNA precursors.
  - A given precursor may arise from multiple distinct loci.

- 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

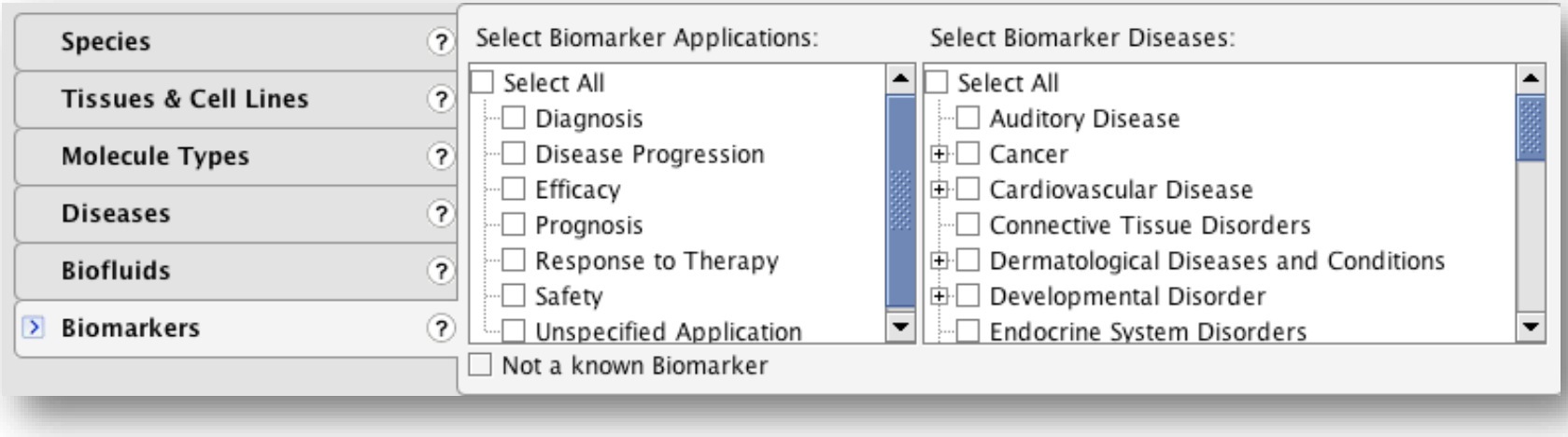
- It does not calculate functions, Canonical Pathways, or networks

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

- Calculates unique and common molecules

Biomarker Filters

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



Clinical Usage (Biomarkers):

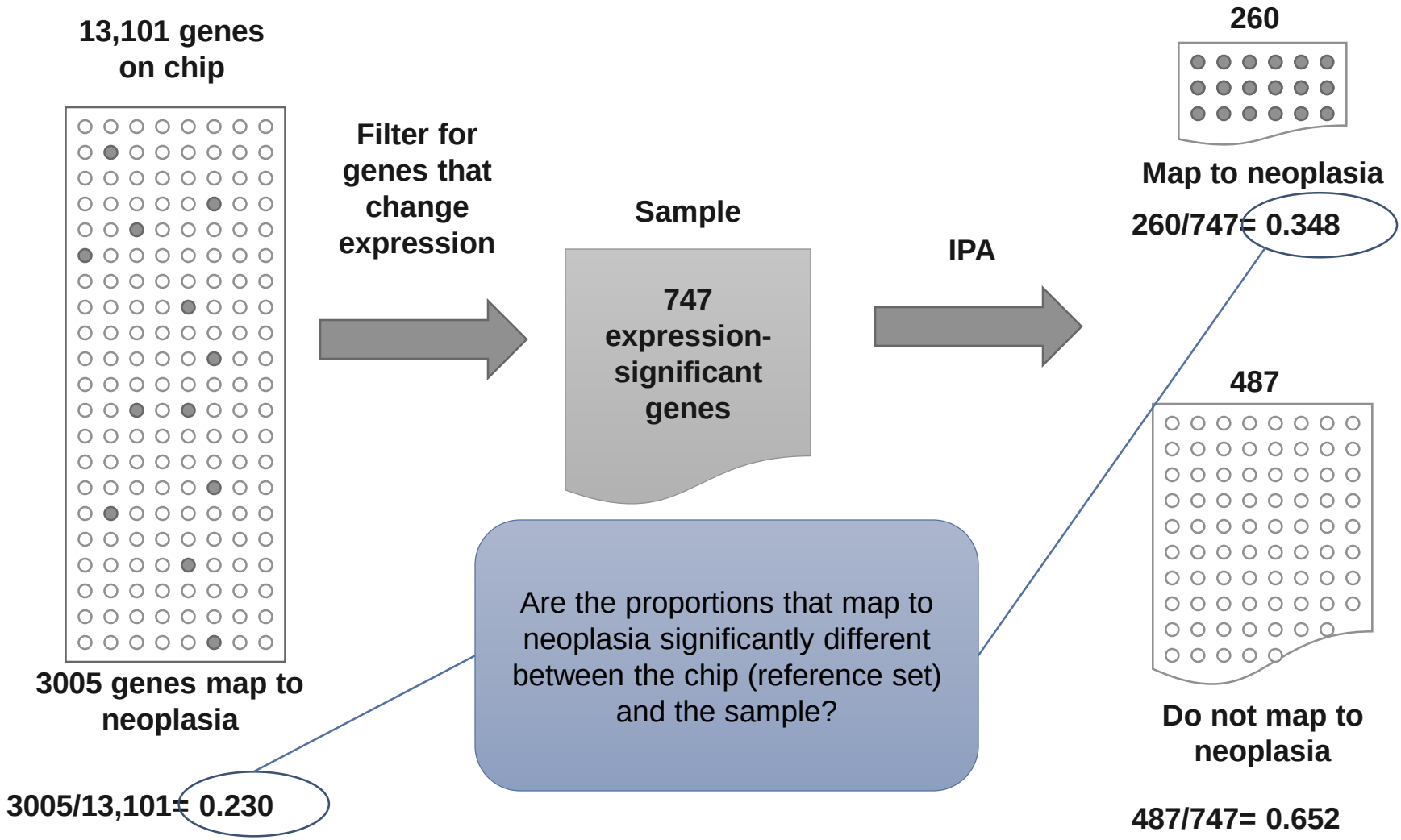
Identify biomarkers by their specific application, including markers for Disease Diagnosis and Prognosis, Disease Progression, markers of Drug Efficacy and Safety, and Patient Response to Therapy

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 measureable 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.

Multiple Testing Correction

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

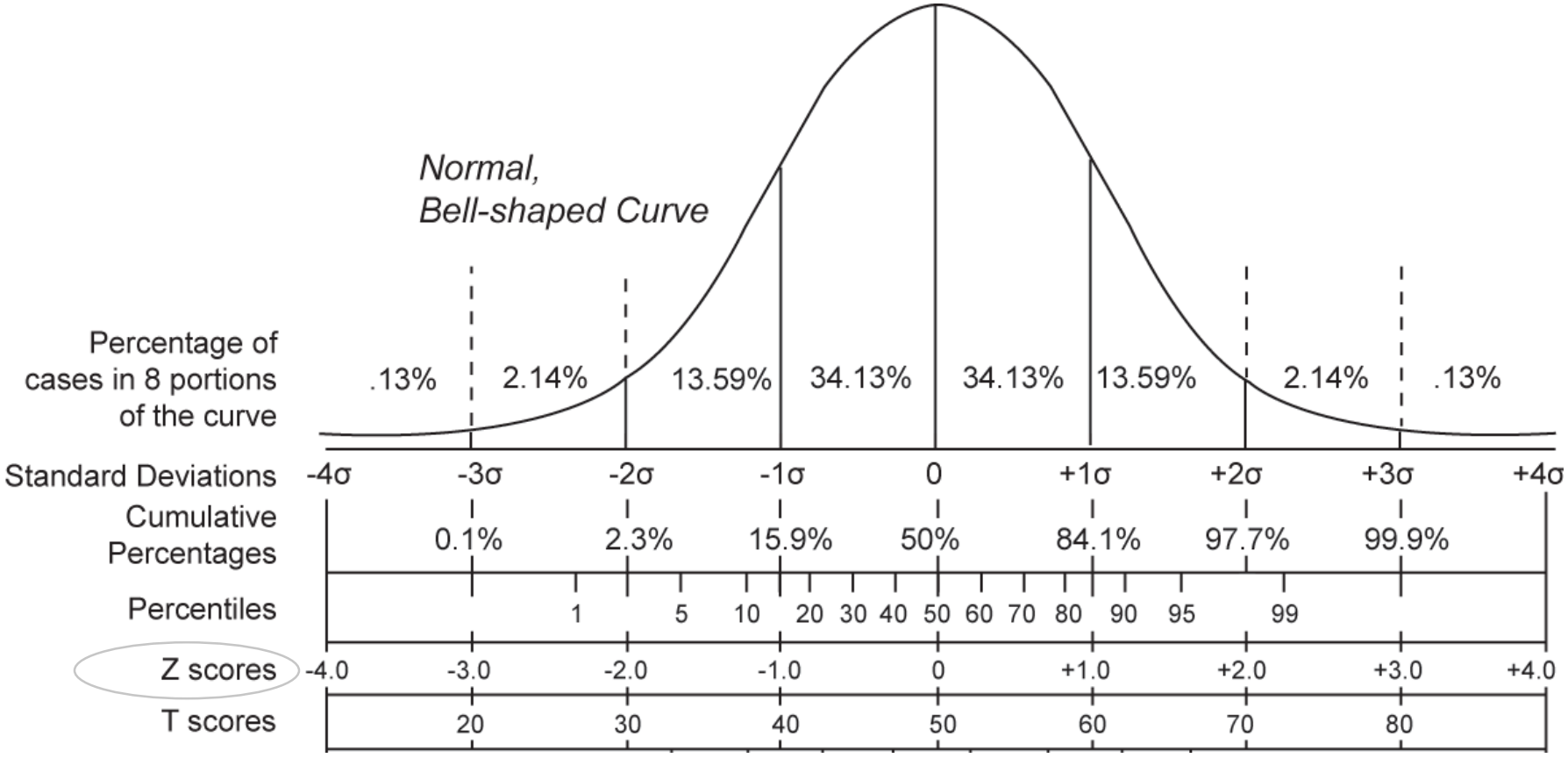
“What are all significant functions within this dataset?”

- Use Benjamini-Hochberg multiple testing correction

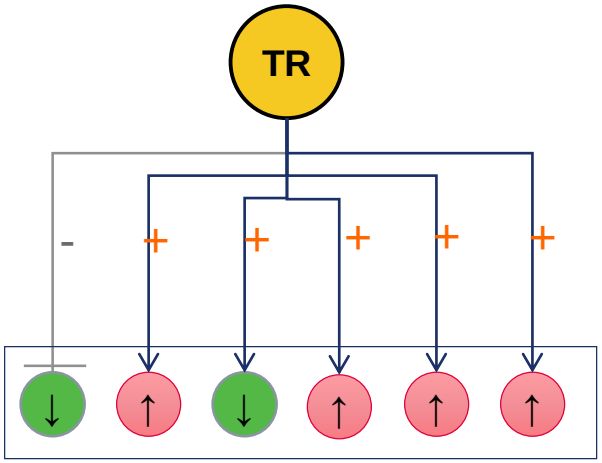
z-scores and Normal Distribution

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



1 1 -1 1 1 1

Every TR is analyzed

Literature-based effect TR has on downstream genes

Differential Gene Expression (Uploaded Data)

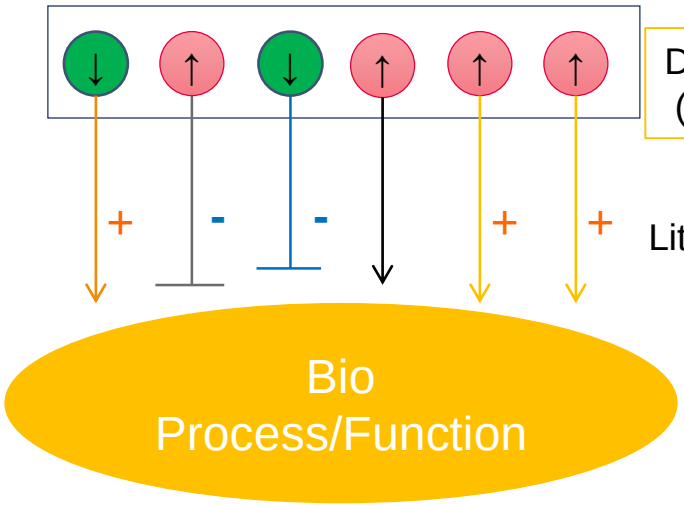
Predicted activation state of TR:
1: activated (correlated), -1: inhibited (anti-correlated)

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

- 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* weighted by relationship, relationship bias, data bias

Downstream Effect z-score



Differential Gene Expression
(Uploaded Data)

Literature-based effect genes have on process or function

-1 -1 1 0 1 1

Predicted Effect-
1: Increasing (correlated), -1: inhibited (anti-correlated)

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

- "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