INGENUITY®
SYSTEMS

IPA® Variant Analysis™ iReport™

Ingenuity Pathways Analysis (IPA) Of Large Datasets

Darryl Gietzen, Ph.D.

IPA Analysis Overview

An IPA (Core) Analysis maps your data to the Ingenuity Knowledge Base (KB) and provides the following outputs:

- Biological **functions** and diseases that are over-represented in your data, and the **predicted directional effects** on these functions and diseases
- Signaling and metabolic **canonical pathways** enriched in your data.
- **Predicted upstream transcription regulators** that might explain the changes observed in your data.
- Molecular **networks** (algorithmically generated pathways describing potential molecular interactions in your experimental system)

Data Upload to IPA

Supported Identifiers for Data Upload

Vendor IDs	Gene	Protein	RNA-Seq	MicroRNA	SNP	Chemical
Affymetrix	Entrez Gene (LocusLink)*	GenPept	Ensembl	miRBase (mature)	Affy SNP IDs	CAS Registry Number
Agilent	GenBank	International Protein Index (IPI)	RefSeq	miRBase (stemloop)	dbSNP	HMDB
ABI	Gene Symbol-human (HUGO/ HGNC, EG)	UniProt/ Swiss-Prot Accession	UCSC (hg18)			KEGG
Codelink	Gene Symbol-mouse (EG)		UCSC (hg19)			PubChem CID
Illumina	Gene Symbol-rat (EG)					
Ingenuity	GI Number					
	UniGene					

*Primary mapping to Entrez Gene

Species Support

Species Supported in IPA
The following species are supported with full content in IPA.

- Human
- Mouse
- Rat

For the following species, we support DbSNP, Entrez Gene, Genbank, Genpept, NCBI GI number, Unigene, and Swissprot/Uniprot identifiers as well as some array-specific data. The identifiers for these species are mapped according to their HomoloGene to the ortholog information in the Ingenuity Knowledge Base. The supporting content for these specific species is not yet linked to their mappings (i.e. the content will be specific to human, mouse, and rat). Species marked with a * indicate that we only support Affymetrix identifiers of this species

- Arabidopsis thaliana
- Bos taurus (bovine)
- Caenorhabditis elegans
- Gallus gallus (chicken)
- Pan troglodytes (chimpanzee)
- Danio rerio (zebrafish)
- Canis lupus familiaris (canine)
- Drosophila melanogaster
- Macaca mulatta (Rhesus Monkey)*
- Saccharomyces cerevisiae
- Schizosaccharomyces pombe

Key Terminology

Observation

- An experimental condition or sample

Expression Value

- Numerical value indicating level of expression or significance of a specific identifier

Uploading Multiple Observations

- IPA calculates significance based on the number of genes/molecules that map to a biological function, pathway, or network.
- Differences between observations will be determined by differences in the number of genes that map to these categories.
- Values associated with IDs need to be either filtered prior to upload or filtered via a threshold during analysis set up.

Multiple Observation: No Filtering

PROBE_ID	Observation 1	Observation 2	Observation 3	Observation 4
1007_s_at	1.102060201	-1.950174692	-2.284966033	-1.490350557
200012_x_at	-1.056204495	-1.278105602	-1.126543599	-2.976190286
200061_s_at	-1.169087572	-1.202210924	-1.331014726	-2.310772708
200062_s_at	1.04380303	-1.251690101	-1.400551146	-2.511834387
200095_x_at	-1.039891293	-1.393501117	-1.155986718	-2.106450544
200633_at	-1.18377169	-1.502370828	-1.137044297	-3.547585559
200650_s_at	1.105123325	-1.503295614	-1.263102761	-3.071208931
200680_x_at	-1.121915583	-1.19019341	-1.261911465	-2.902124686
200716_x_at	1.055138928	-1.395614383	-1.199132766	-2.592501325
Significant Genes	10	10	10	10

All observations have 10 significant genes and will look identical in IPA

Multiple Observations: Pre-filtering

PROBE_ID	Observation 1	Observation 2	Observation 3	Observation 4
1007_s_at		-1.950174692	-2.284966033	
200012_x_at				-2.976190286
200061_s_at				-2.310772708
200062_s_at				-2.511834387
200095_x_at				-2.106450544
200633_at		-1.502370828		-3.547585559
200650_s_at		-1.503295614		-3.071208931
200680_x_at				-2.902124686
200716_x_at				-2.592501325
Significant Genes	0	3	1	9

Pre-filtering data for fold change less than 1.5 will differentiate the four observations because there are differing numbers of probes/genes for each observation.

Multiple Observations: Filtering During Analysis Set-up

PROBE_ID	Observation 1	Observation 2	Observation 3	Observation 4
1007_s_at	1.102060201	-1.950174692	-2.284966033	-1.490350557
200012_x_at	-1.056204495	-1.278105602	-1.126543599	-2.976190286
200061_s_at	-1.169087572	-1.202210924	-1.331014726	-2.310772708
200062_s_at	1.04380303	-1.251690101	-1.400551146	-2.511834387
200095_x_at	-1.039891293	-1.393501117	-1.155986718	-2.106450544
200633_at	-1.18377169	-1.502370828	-1.137044297	-3.547585559
200650_s_at	1.105123325	-1.503295614	-1.263102761	-3.071208931
200680_x_at	-1.121915583	-1.19019341	-1.261911465	-2.902124686
200716_x_at	1.055138928	-1.395614383	-1.199132766	-2.592501325
Significant Genes	0	3	1	9

Setting a threshold of 1.5 during analysis set up will result in IPA ignoring values less than 1.5 fold up or down, differentiating the samples.

Best Practices of Data Upload

- Calculate metrics outside of IPA (e.g. fold-change, p-value).
- Create an Excel spreadsheet or tab delimited file
  - One column must have identifiers, preferably the left-most column
  - Can have up to 20 observations
  - 0 to 3 expression value-types per observation
  - Only 1 header row allowed
  - Place data in top Excel worksheet
  - Place metric you would like to visualize, usually fold-change (recommend log2[ratio]), in the first column after ID.
  - Optionally upload the max intensity, or RPKM, between case and control to understand magnitude of signal
- If you can group related observations into a single spreadsheet, do so.
- Specify array platform (chip) if possible
  - Its is OK for “Not specified/applicable”
- Pre-filter you data at the lowest threshold that you have confidence in.
  - For example, probe measurement p-value of .05 or other criteria.

Why don't all of the molecules in my dataset map to the knowledge base?

- 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.

Data Upload Format Examples

	A	B
1	Accession ID	Fold Change
2	AA434409	1.0099
3	AA406422	1.1391
4	NR3C1	-1.0707
5	NOTCH4	-1.1334
6	NFKB1	1.2358
7	DPYD	1.0293
8	SP100	1.6847
9	C4BPA	1.1657
10	LOX	-1.1096
11	ABCC5	1.1666
12	GABRE	1.0746
13	H69576	1.2516

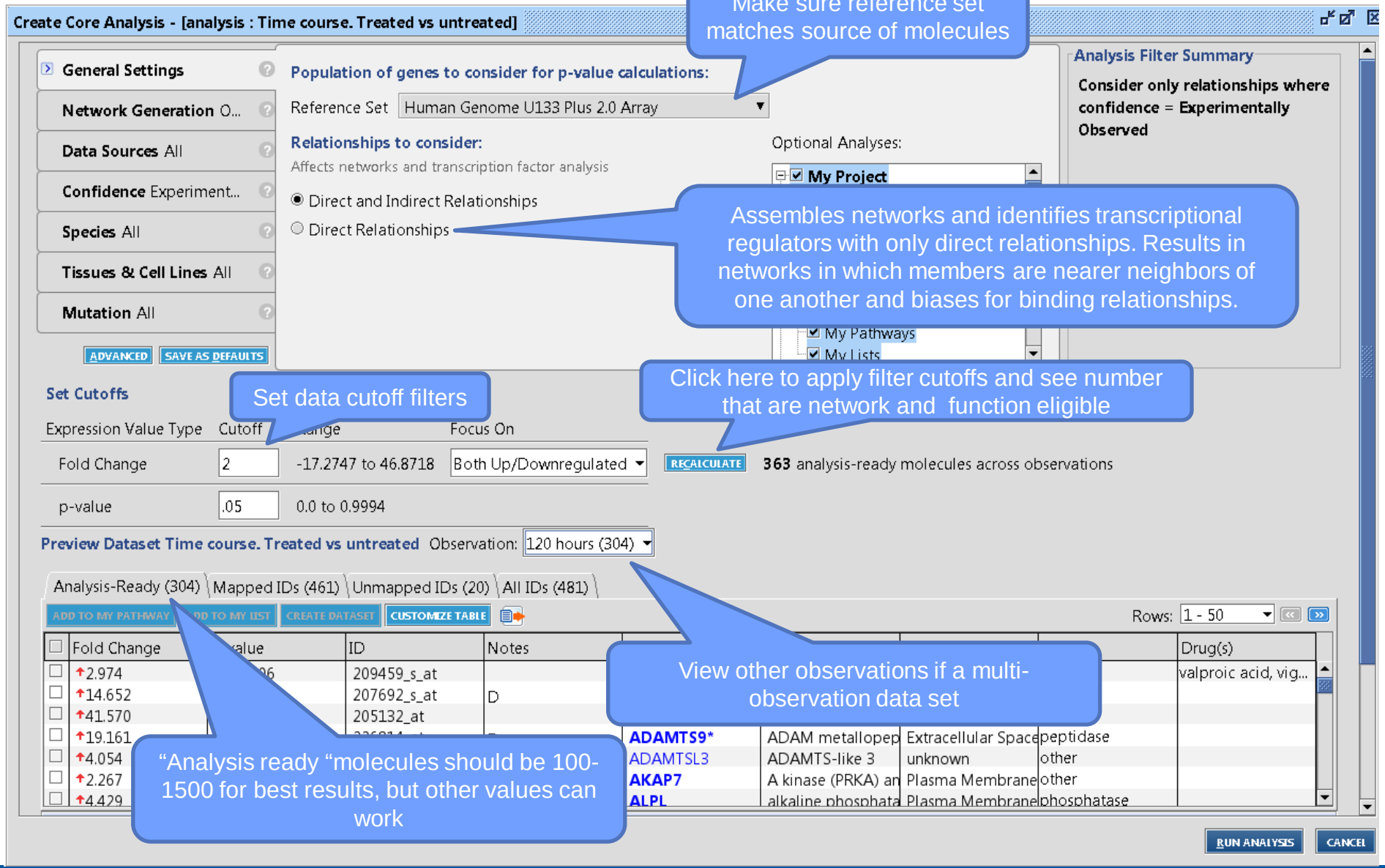
	A	B	C
1	Accession ID	Log(Ratio)	Fold Change
2	AA434409	0.0143	1.0099
3	AA406422	0.1879	1.1391
4	NR3C1	-0.0986	-1.0707
5	NOTCH4	-0.1806	-1.1334
6	NFKB1	0.3055	1.2358
7	DPYD	0.0417	1.0293
8	SP100	0.7525	1.6847
9	C4BPA	0.2211	1.1657
10	LOX	-0.1501	-1.1096
11	ABCC5	0.2222	1.1666
12	GABRE	0.1038	1.0746
13	H69576	0.3237	1.2516

	A	B	C	D	E
1	Accession ID	A Log(Ratio)	A Fold Change	B Log(Ratio)	B Fold Change
2	AA434409	0.0143	1.0099	-0.2831	-1.2168
3	AA406422	0.1879	1.1391	-0.6167	-1.5334
4	NR3C1	-0.0986	-1.0707	0.0763	1.0543
5	NOTCH4	-0.1806	-1.1334	0.0297	1.0208
6	NFKB1	0.3055	1.2358	-0.2092	-1.1561
7	DPYD	0.0417	1.0293	0.1218	1.0881
8	SP100	0.7525	1.6847	-0.5269	-1.4408
9	C4BPA	0.2211	1.1657	-0.1748	-1.1288
10	LOX	-0.1501	-1.1096	0.0778	1.0554
11	ABCC5	0.2222	1.1666	-0.0800	-1.0570
12	GABRE	0.1038	1.0746	-0.5624	-1.4767
13	H69576	0.3237	1.2516	0.2257	1.1694

- Gene list with 1 or more identifiers
  - Tip: IPA will auto-detect the identifier type if it is in the first column of your spreadsheet
- Gene list with 1 or more identifiers and a single measurement of one observation
  - Tip: IPA can import fold-change, log(ratio), ratio, p-value, intensity/RPKM
- Gene list with 1 or more identifiers and a multiple measurements of one observation
  - Tip: Up to 3 different measurements can be mapped
  - Tip: Measurements can be in any order but ‘easiest’ if log(ratio) is the first measurement
- Gene list with 1 or more identifiers and a multiple measurements of multiple observations
  - Tip: File can contain up to 20 different observations
- Data can be in text file, or Excel file
  - Tip: If using an Excel file ONLY the first worksheet is read and used

Creating an IPA Core Analysis

Creating an IPA Core Analysis



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.

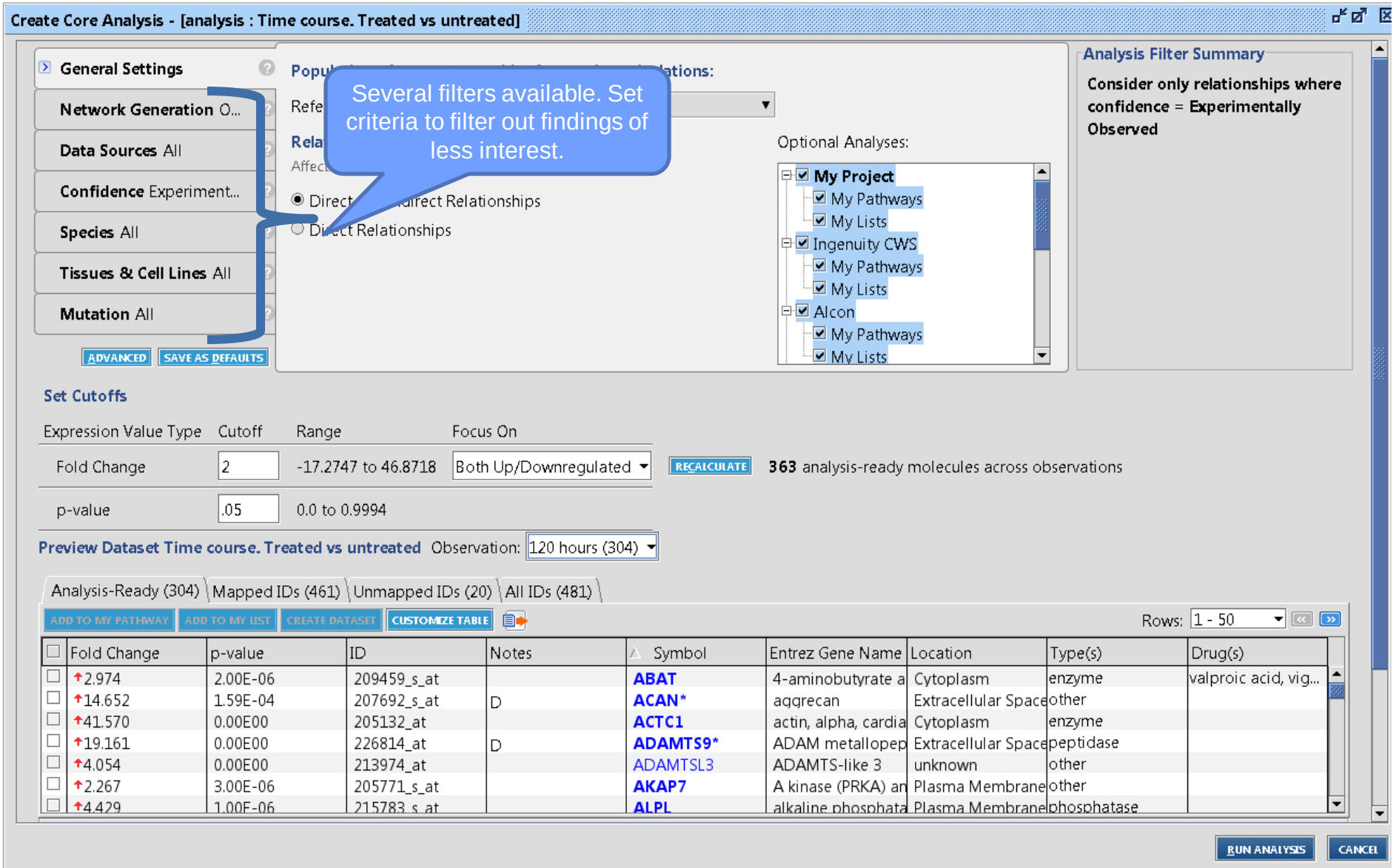
Set data cutoff filters

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

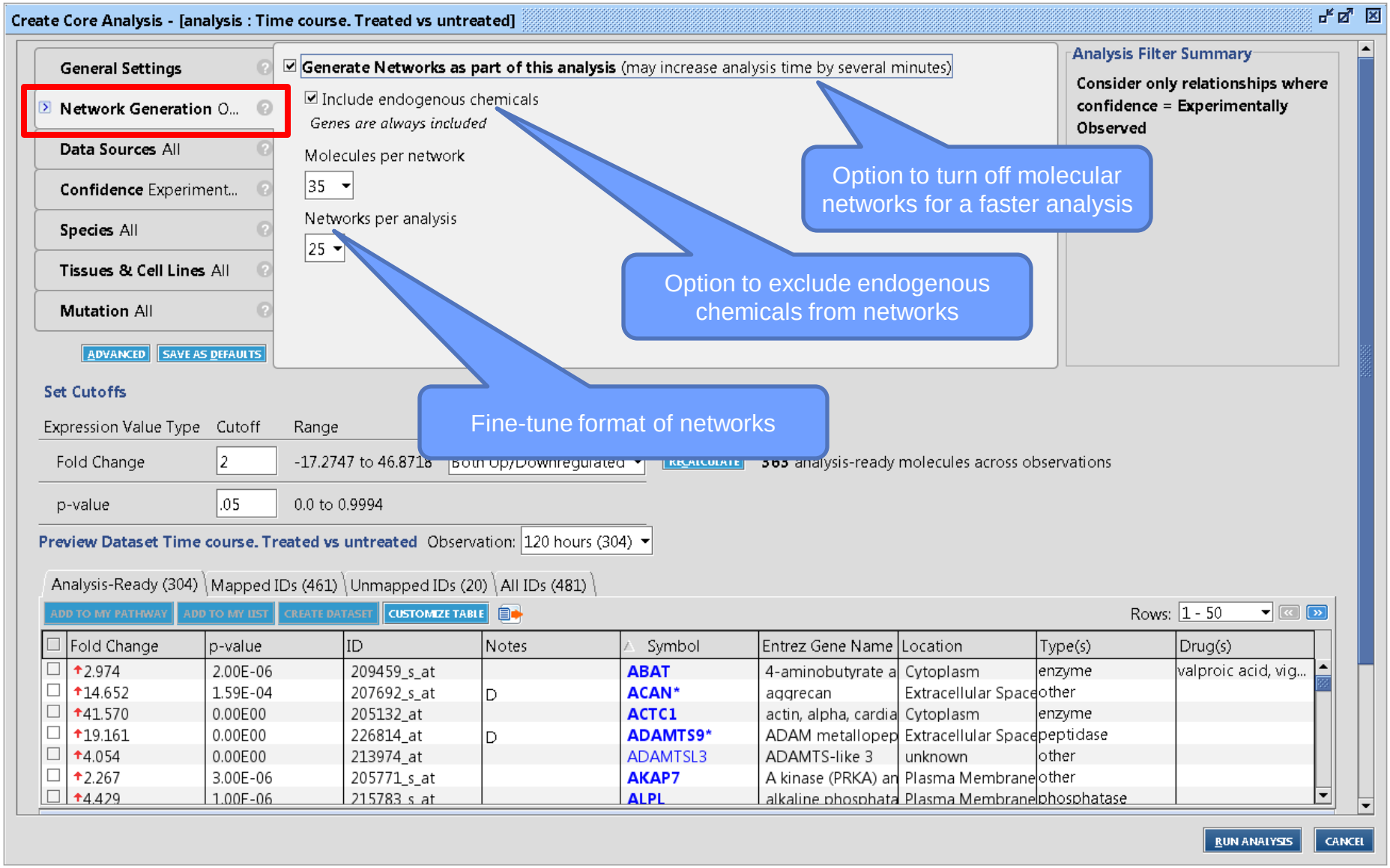
View other observations if a multi-observation data set

"Analysis ready "molecules should be 100-1500 for best results, but other values can work

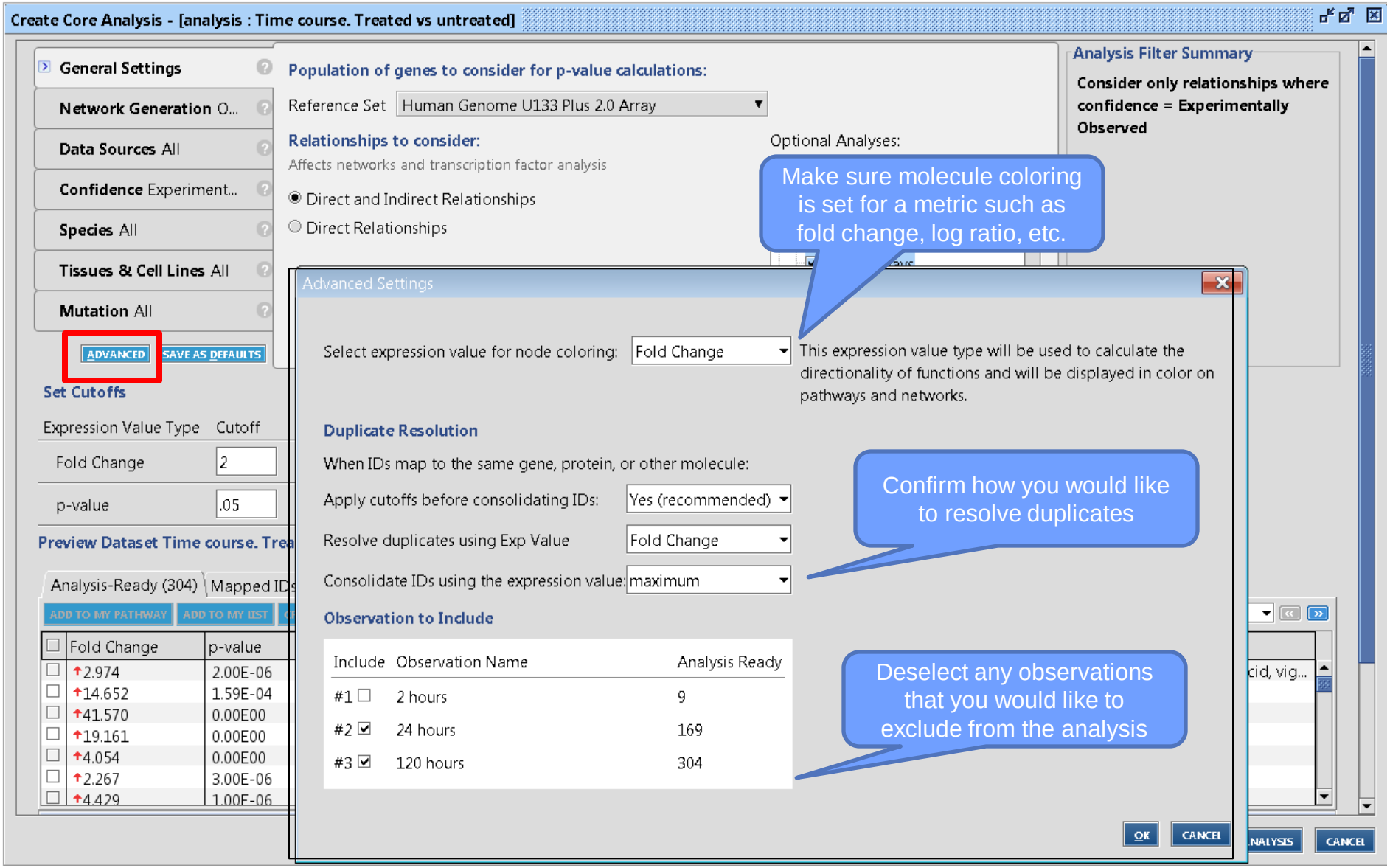
Creating an IPA Core Analysis: Using Filters



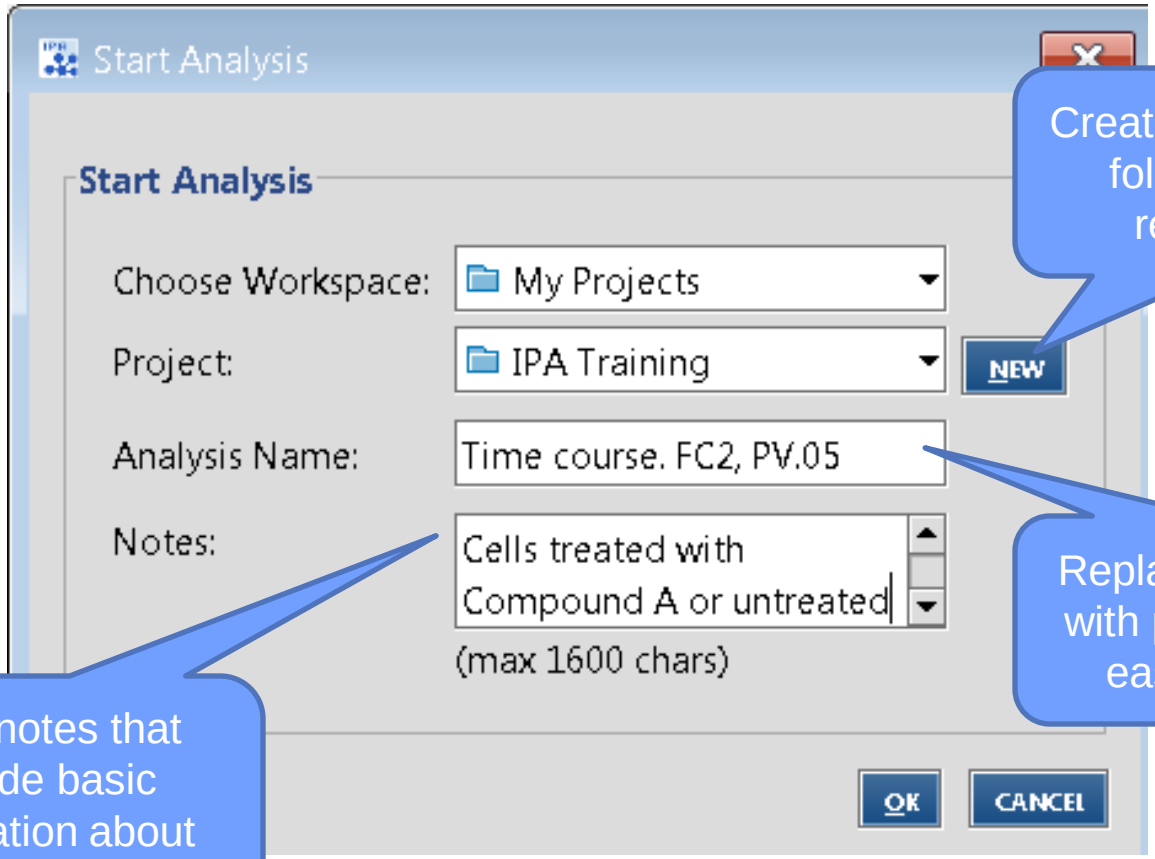
Creating an IPA Core Analysis- Network



Creating an IPA Core Analysis- Advanced Settings



Saving an IPA Core Analysis

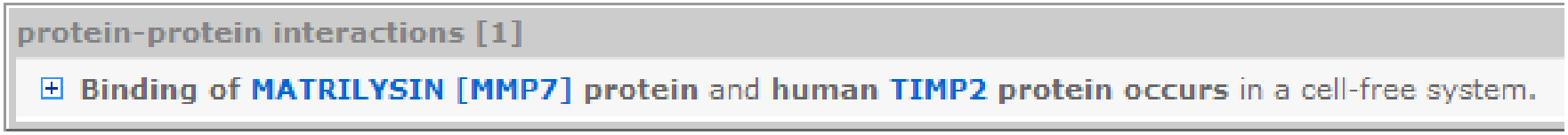


Using Core Analysis Pre-filters

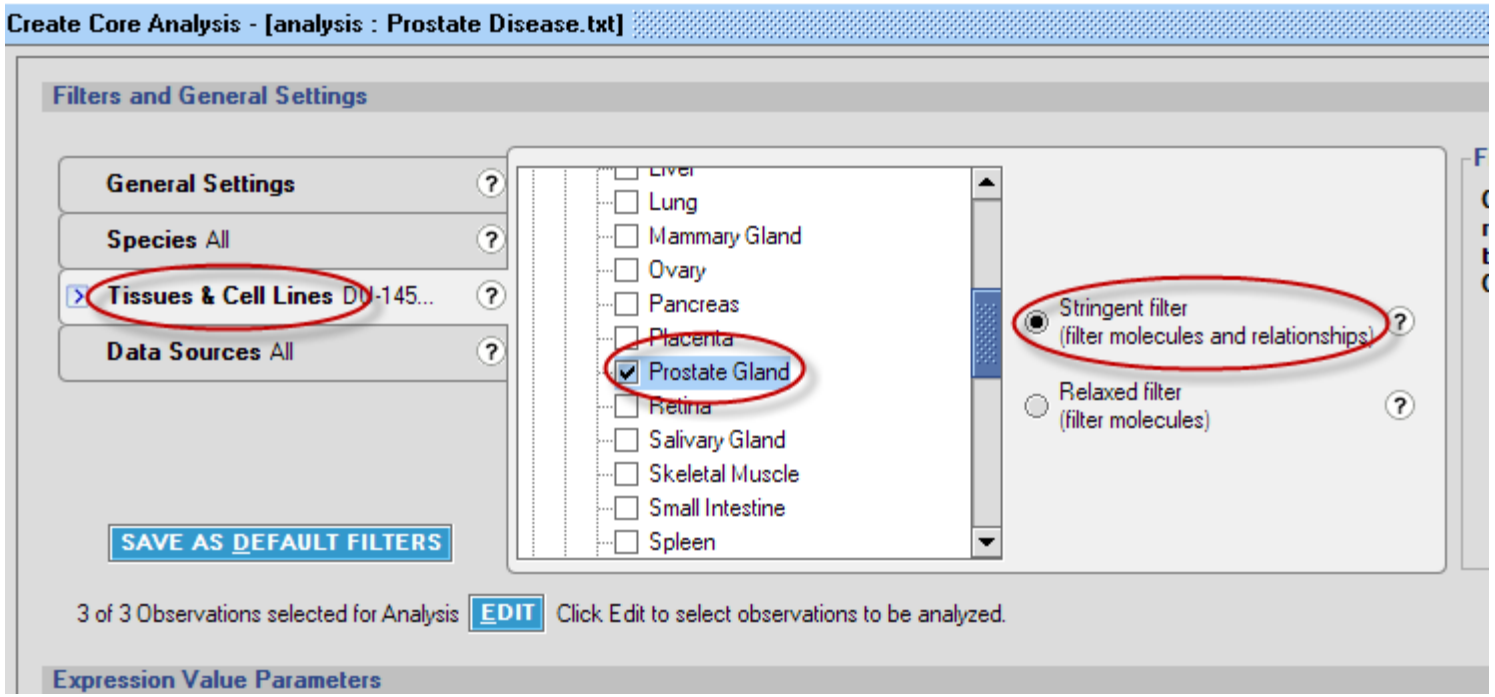
- Set criteria to filter out findings of less interest.
  - Species
  - Tissue
  - Data source
- Filter stringency
  - A “Stringent” setting requires that each of a pair of molecules and the relationship that connects them meet the filter criteria
  - A “Relaxed” filter requires that the gene or protein 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 DU-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

Analyzing Results

Ingenuity Definitions

Networks

• Generated de novo based upon input genes, proteins, or chemicals

Canonical Pathways (Signaling and Metabolic)

- Are generated prior to data input, based on the literature
- Do NOT change upon data input
- More directional (proceed “from A to Z”)

My Pathways and Path Designer Pathways

- Custom pathways
- You can add information/edit custom pathways

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.

Core Analysis Statistics

IPA calculates the following statistics for the outputs of a core analysis:

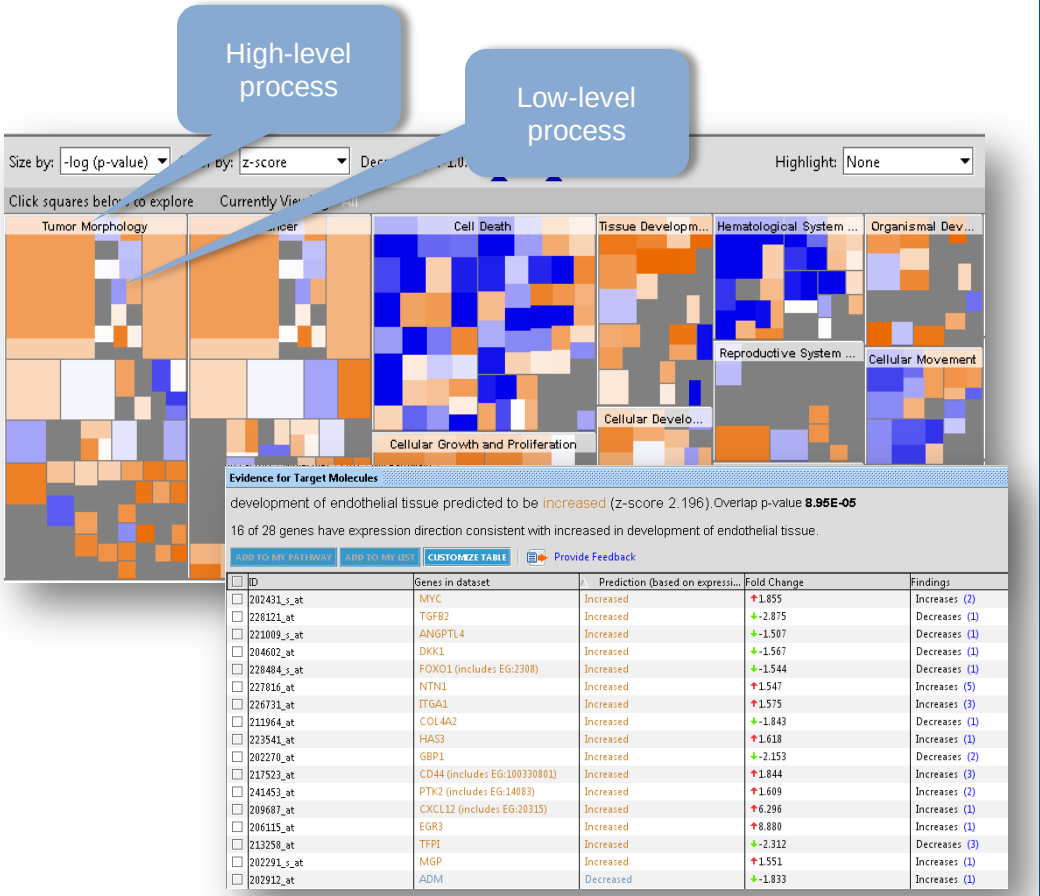
Analysis Type	P-value (FET)*	Z-score*
Functions Analysis	√	
Downstream Effects Analysis	√	√
Canonical Pathway Analysis	√	
Transcription Factor Analysis	√	√
Network Analysis	√	

*Different values will enter into the calculations for p-values and Z-scores, depending on the analysis type.

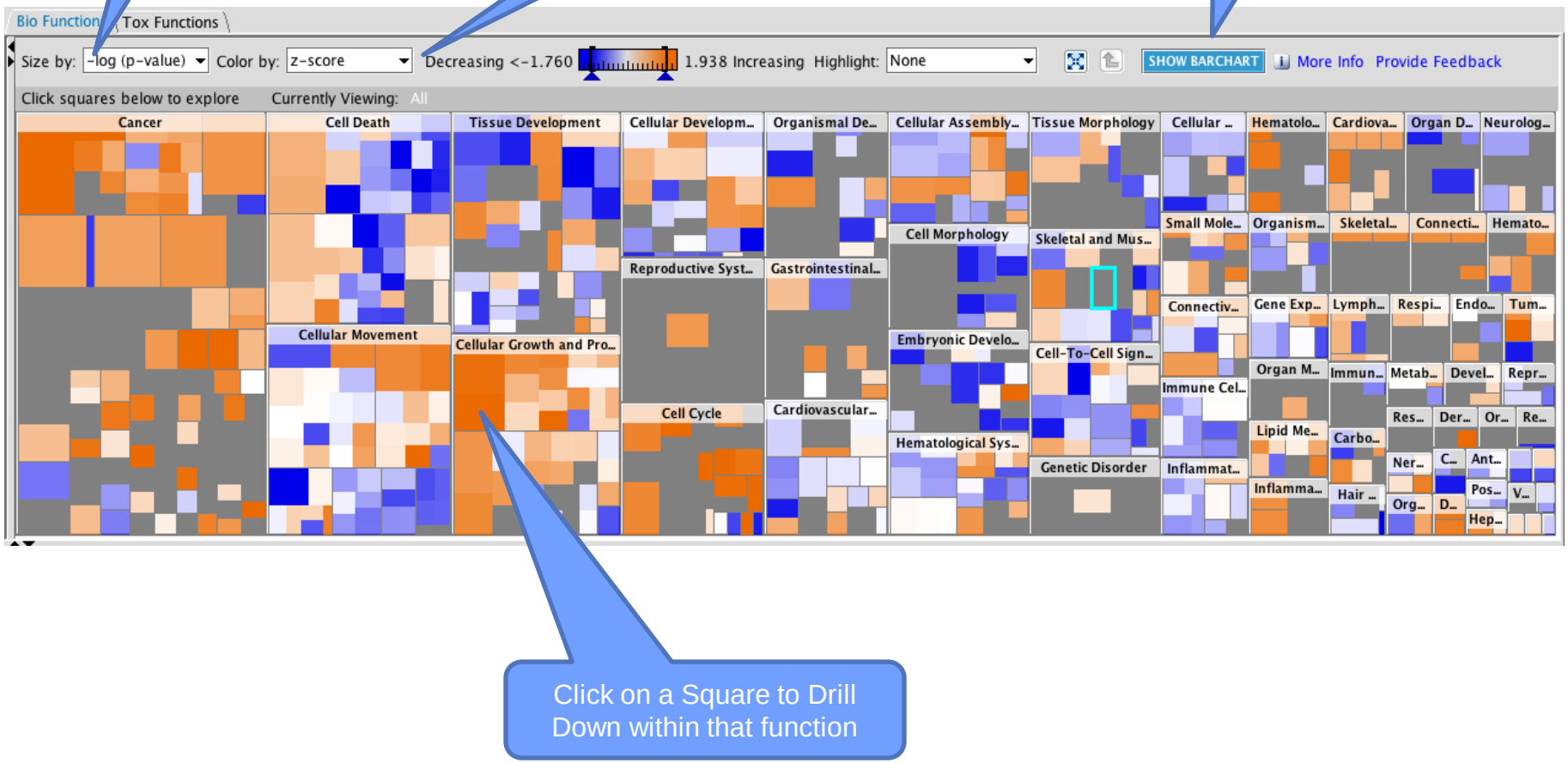
Downstream effects analysis (functional analysis)

Downstream effects analysis - a whole new way to visualize biological trends in your experiment

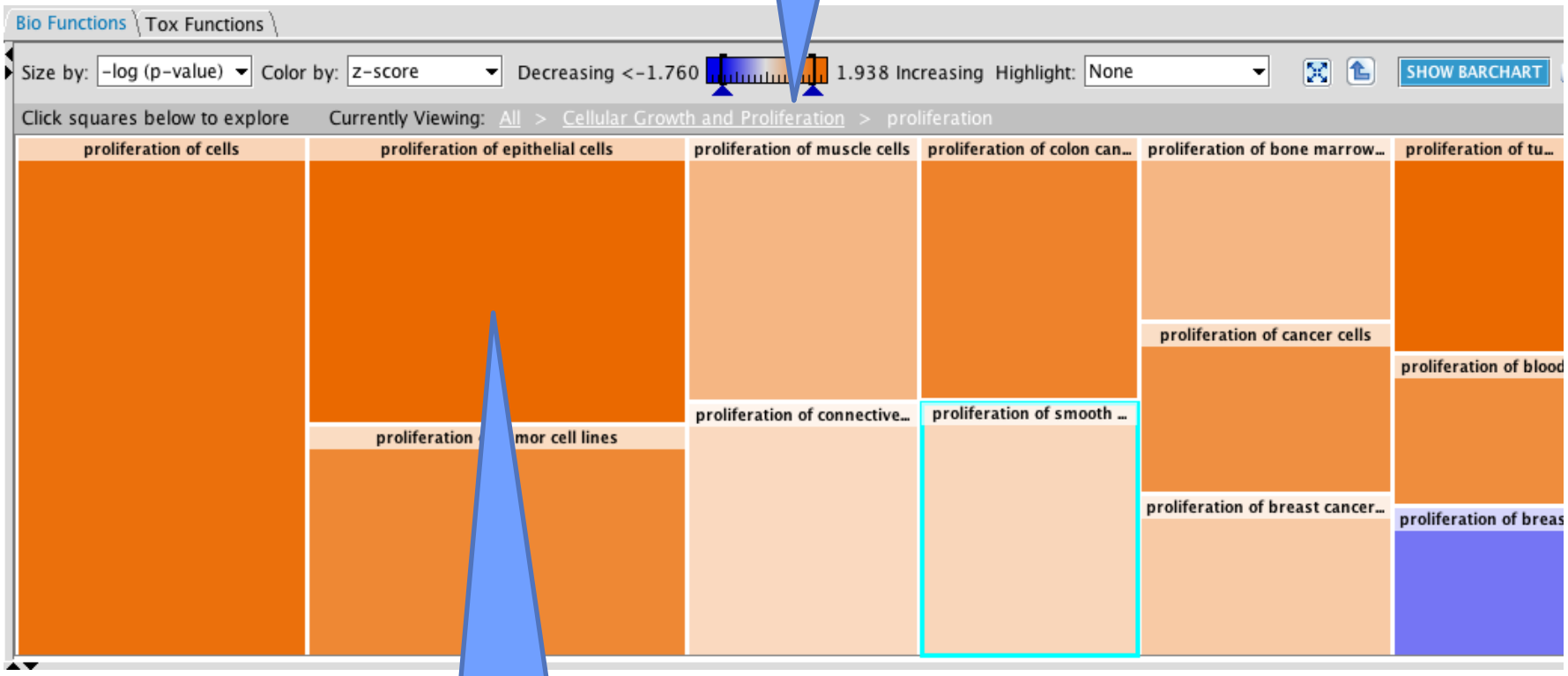
- Uses Z-score to predict increases or decreases in downstream biological processes
- Use a hierarchical overview to quickly get a top-to-bottom view of the biology affected in your experiment
- Zoom in on areas of interest and quickly see the specific genes and references that support a prediction



Downstream Effect on Bio Function



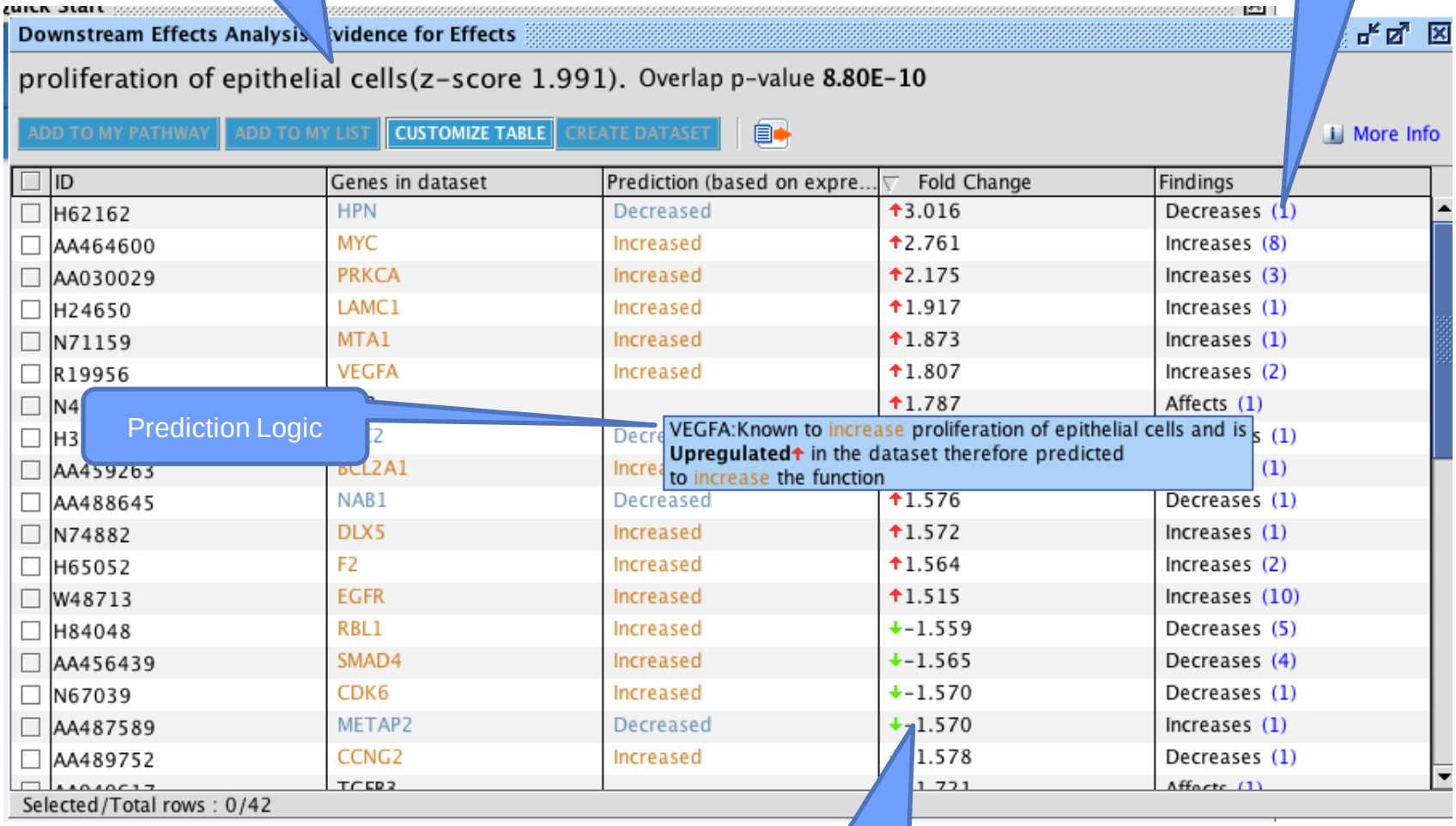
Ontology Levels



Click to See the Specific Genes and Findings

Functional Category and Statistical Result

Access Findings



Prediction Logic

Expression Value in Your Dataset

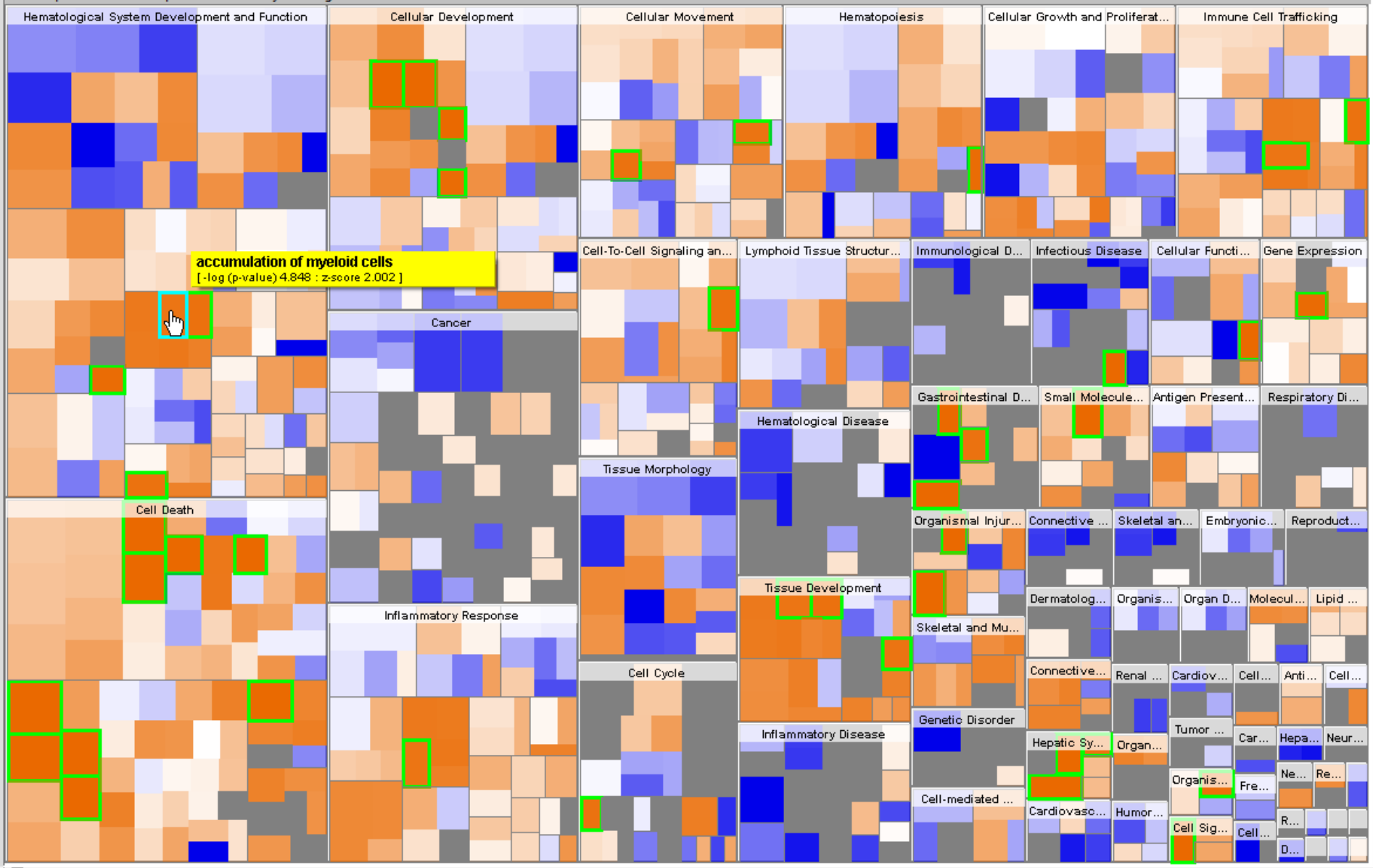
Day 10

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



Downstream Effects Analysis: Evidence for Effects

accumulation of myeloid cells predicted to be increased (z-score 2.002). Overlap p-value **1.42E-05**

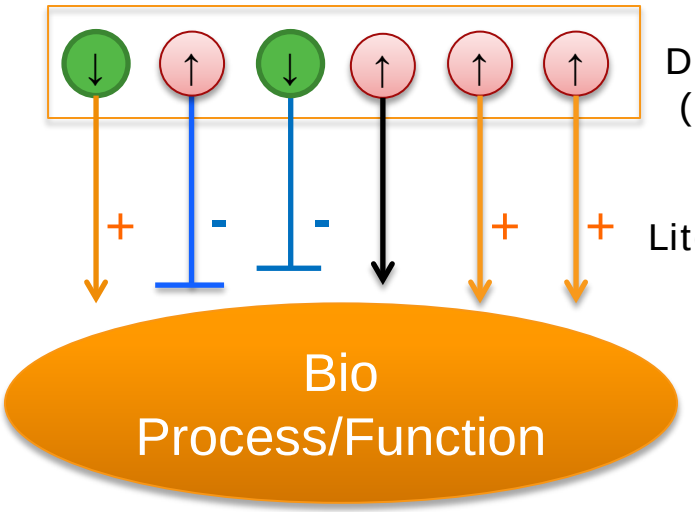
12 of 15 genes have expression direction consistent with increased in accumulation of myeloid cells.

ADD TO MY PATHWAY | ADD TO MY LIST | CUSTOMIZE TABLE | Provide Feedback | More Info

ID	Genes in dataset	Prediction (based on expressi...	Fold Change	Findings
1368742_at	C5AR1	Increased	↑2.000	Increases (1)
1373579_at	RARA	Increased	↑1.800	Increases (1)
1398256_at	IL1B	Increased	↑5.800	Increases (1)
1386893_at	GRN	Increased	↑2.400	Increases (2)
1370083_at	CCR1	Increased	↑4.100	Increases (1)
1368321_at	EGR1	Increased	↑5.100	Increases (2)
1368940_at	P2RY2	Increased	↑2.900	Increases (4)
1374468_at	MYD88	Increased	↑3.800	Increases (3)
1368494_at	S100A8	Increased	↑5.000	Increases (1)
1391384_at	TNF	Increased	↑1.600	Increases (8)
1398275_at	MMP9	Increased	↑5.400	Increases (3)
1370224_at	STAT3	Increased	↑2.700	Increases (2)
1367715_at	TNFRSF1A	Decreased	↑4.800	Decreases (2)
1368885_at	ENTPD1	Decreased	↑6.300	Decreases (1)
1371249_at	XBP1 (includes EG:140614)	Decreased	↑1.500	Decreases (1)

Selected/Total rows : 0/15

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

Functional Analysis (FA) Workflow

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

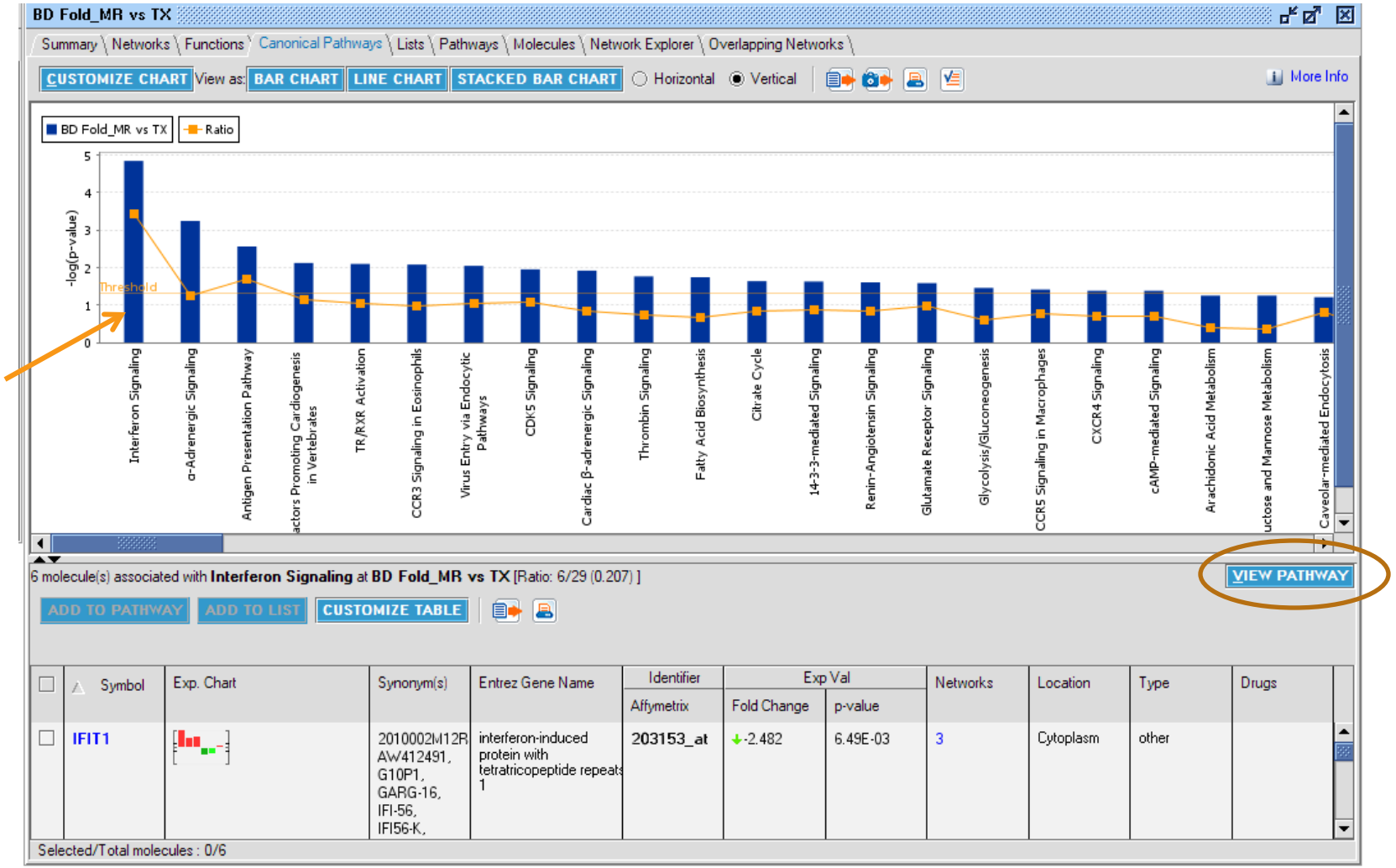
- Genes related to a particular function can be :
  - sent to a pathway for building and/or overlay analysis
  - saved as a new Data Set and sent to Core Analysis for additional categorization and segmentation

Canonical pathway analysis

Canonical Pathway (CP) Workflow

- Scan CP names for pathways of particular interest
- Survey many top or interesting CPs by viewing
  - Click on CP bar in chart and click “VIEW PATHWAY”
- If your data has values, are overlaid values consistent with molecular relationships between molecules?
  - If CP indicates that gene A increases expression of gene B, do the data indicate that genes A and B are co-regulated?

Canonical Pathway tab



Pathway Navigation

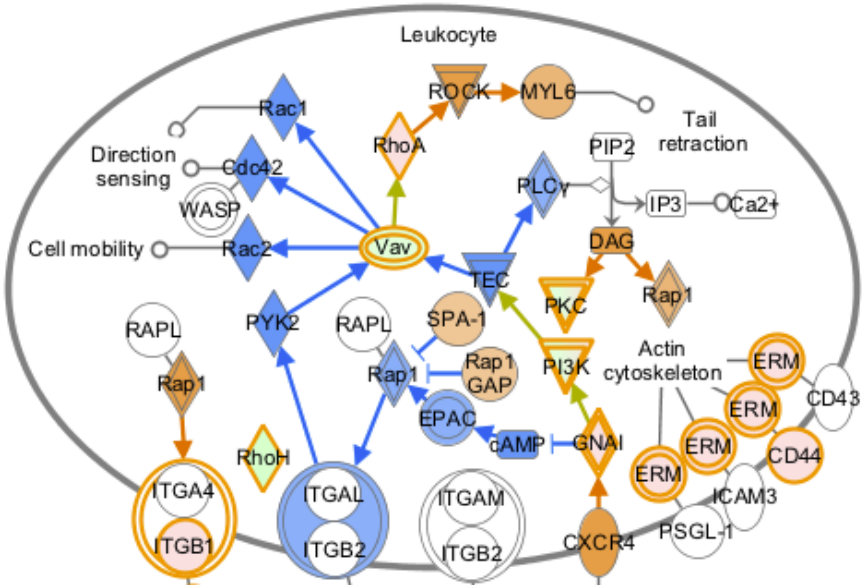
- Scroll-wheel on mouse controls zoom, or use toolbar zoom buttons.
- Left-click selects (turns blue)
- Left-click-drag on nodes moves the node
- Right-click hold-and-drag moves your view
- Right-click brings up menu for controlling
  - tool tip (mouse-over node pop-up)
  - copy/past
  - Highlight
  - 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

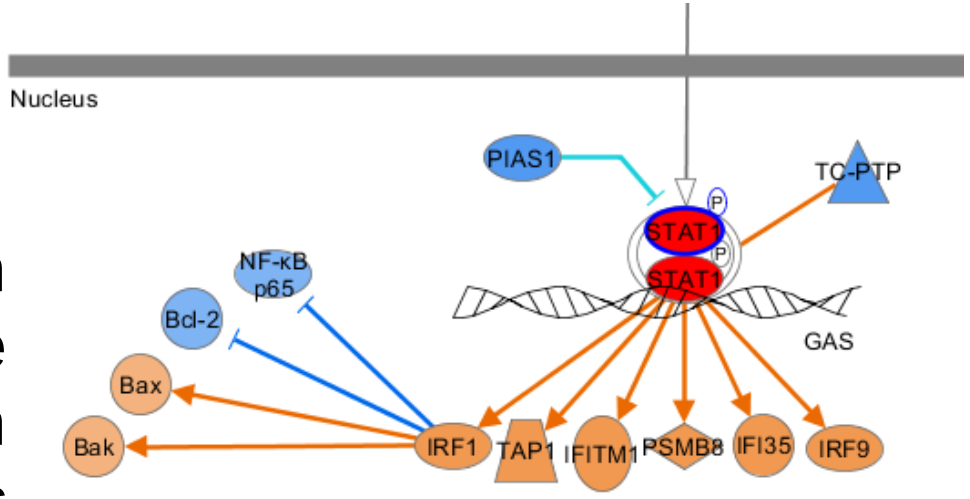
Pathway Navigation, continued

- 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

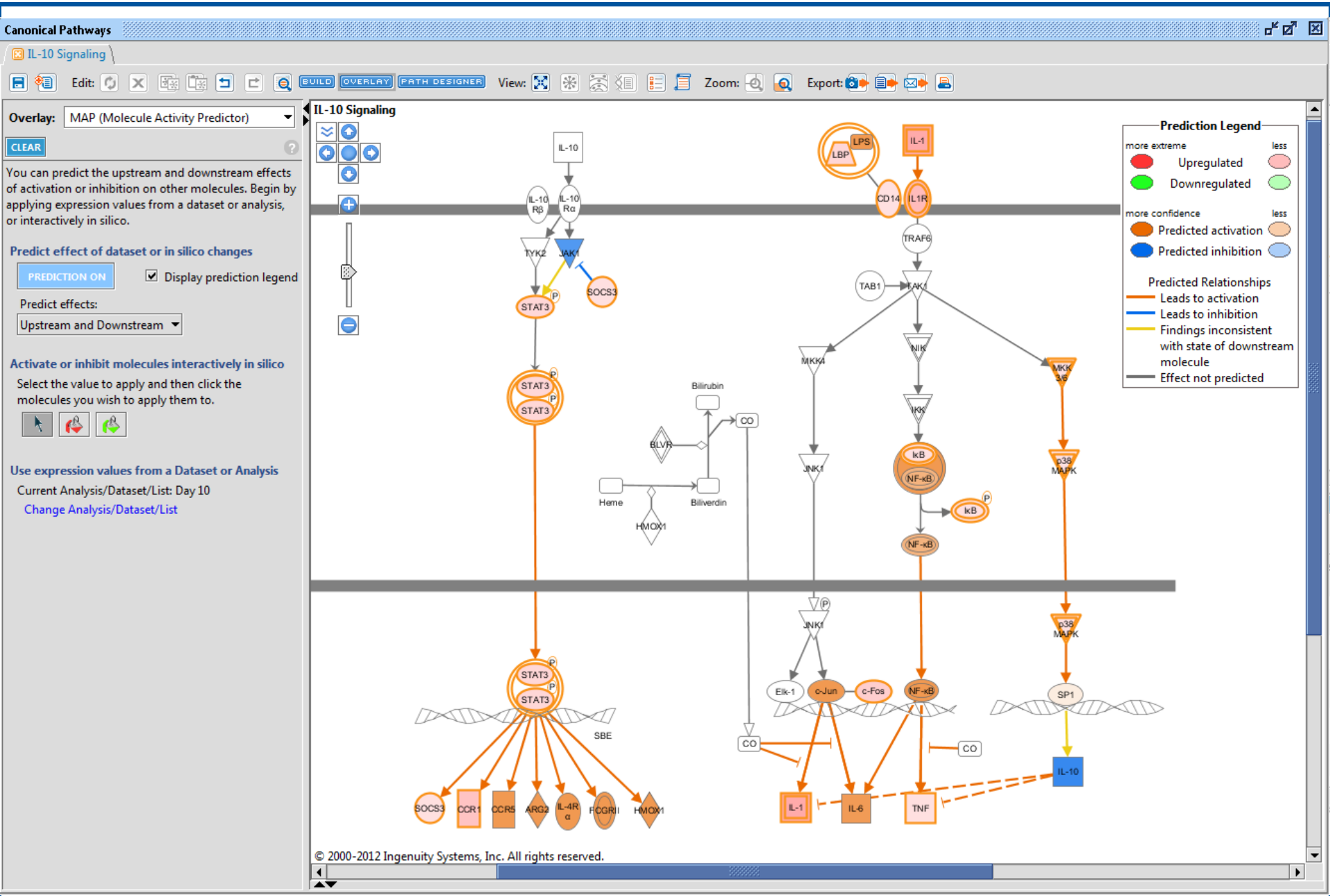
Directional Effects Applied to Pathways & Networks

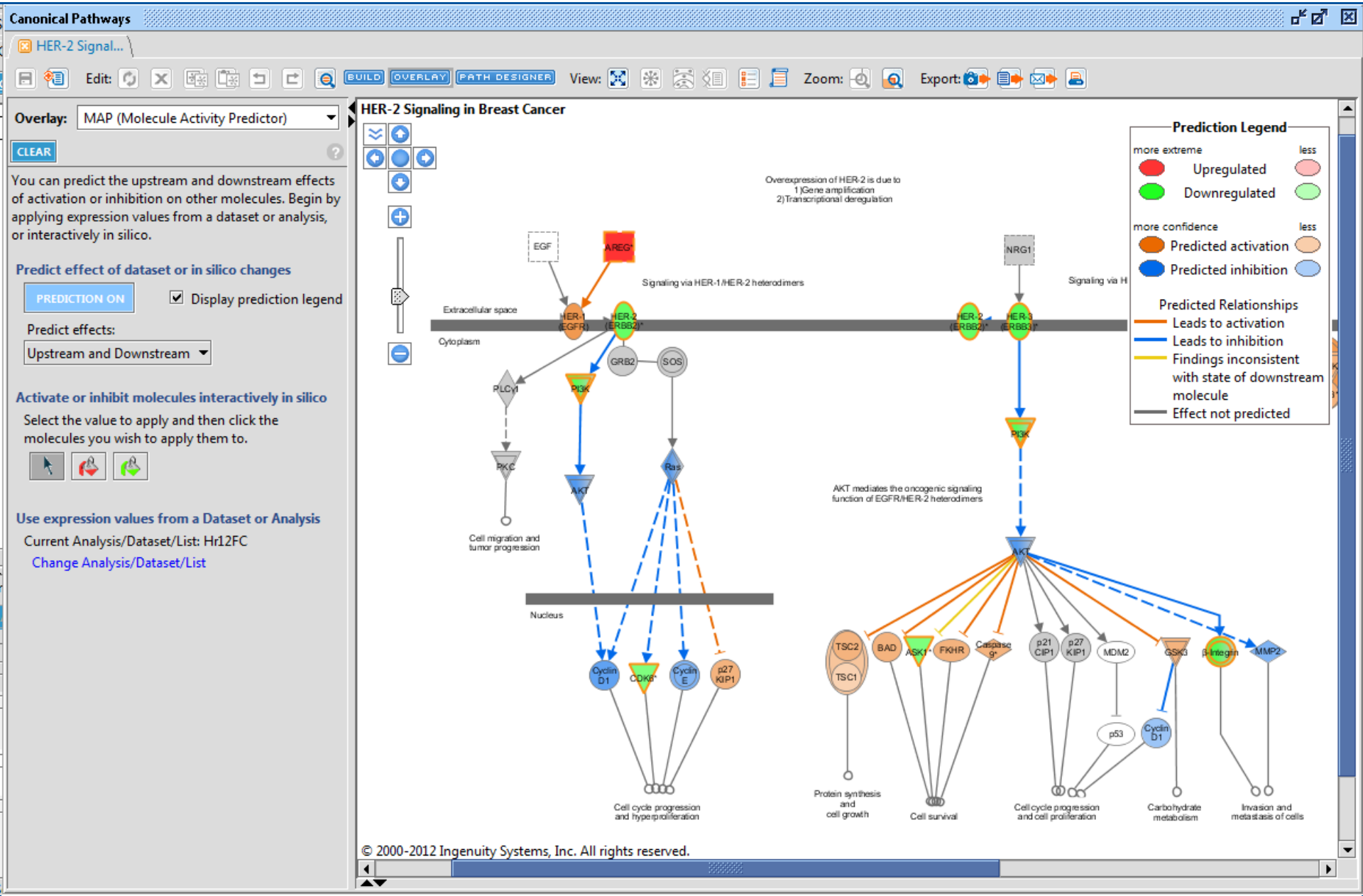


Use observed expression changes (↑↓) to suggest functional effects (↑↓) on neighbouring molecules



Set altered activation states to observe predicted effects on canonical pathways


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
IL-10 Signaling
Prediction Legend
more extreme
less
Upregulated
Downregulated
more confidence
less
Predicted activation
Predicted inhibition
Predicted Relationships
Leads to activation
Leads to inhibition
Findings inconsistent with state of downstream molecule
Effect not predicted
IL-10
IL-10 Rβ
IL-10 Rα
TYK2
JAK1
SOCS3
STAT3
LPS
LBP
IL-1
CD14
IL1R
TRAF6
TAB1
TAK1
MKK4
NIK
MKK 3/6
IKK
kB
NF-kB
p38 MAPK
JNK1
Bilirubin
CO
BLVR
Heme
Biliverdin
HMOX1
SBE
SOCS3
CCR1
CCR5
ARG2
IL-4R α
FCGRII
Elk-1
c-Jun
c-Fos
SP1
IL-1
IL-6
TNF
IL-10
© 2000-2012 Ingenuity Systems, Inc. All rights reserved.


Canonical Pathways
HER-2 Signal...
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: Hr12FC
Change Analysis/Dataset/List
HER-2 Signaling in Breast Cancer
Overexpression of HER-2 is due to
1)Gene amplification
2)Transcriptional deregulation
Prediction Legend
more extreme
less
Upregulated
Downregulated
more confidence
less
Predicted activation
Predicted inhibition
Predicted Relationships
Leads to activation
Leads to inhibition
Findings inconsistent with state of downstream molecule
Effect not predicted
EGF
AREG*
NRG1
Signaling via HER-1/HER-2 heterodimers
Signaling via H
Extracellular space
Cytoplasm
HER-1 (EGFR)
HER-2 (ERBB2)*
HER-2 (ERBB2)*
HER-3 (ERBB3)*
GRB2
SOS
PLCγ1
PI3K
PKC
Ras
AKT
Cell migration and tumor progression
Nucleus
Cyclin D1
CDK6*
Cyclin E
p27 KIP1
Cell cycle progression and hyperproliferation
PI3K
AKT mediates the oncogenic signaling function of EGFR/HER-2 heterodimers
AKT
TSC2
TSC1
BAD
ASK1*
FKHR
Caspase 9*
p21 CIP1
p27 KIP1
MDM2
GSK3
β-Integrin
MMP2
p53
Cyclin D1
Protein synthesis and cell growth
Cell survival
Cell cycle progression and cell proliferation
Carbohydrate metabolism
Invasion and metastasis of cells
© 2000-2012 Ingenuity Systems, Inc. All rights reserved.

Upstream Regulator Analysis/ Transcription Factor Analysis

What is Upstream Regulator Analysis?

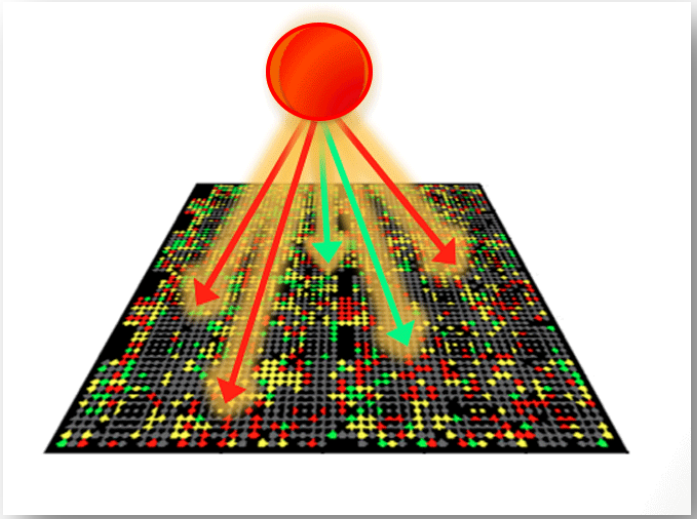
- The upstream regulator analysis is based **on prior knowledge** of expected effects between **transcriptional regulators (TR)** and their **target genes** stored in the Ingenuity® Knowledge Base.

- The analysis examines *how many known targets* of each transcription regulator are present in the user's dataset, and also *compares their direction of change* (i.e. expression in the experimental sample(s) relative to control) to what is expected from the literature in order to predict likely relevant transcriptional regulators.

- IPA's definition of upstream transcriptional regulator **is quite broad** – any molecule that can affect the expression of other molecules, which means that upstream regulators can be almost **any type of molecule**, from transcription factor, to microRNA, kinase, compound or drug.

Upstream Regulator Analysis: How does it work?

- Use experimentally observed relationships (vs. Predicted event) between Upstream Regulators and genes to predict potential regulator and activation



- Predict activation or inhibition of regulator to explain the changes in gene expression in your dataset

- Calculates two complementary statistical measures:
  - Activation z-score
  - Overlap p-value

How does it work?

Can we predict the activation state (activated/inhibited) of a potential regulator from expression data?

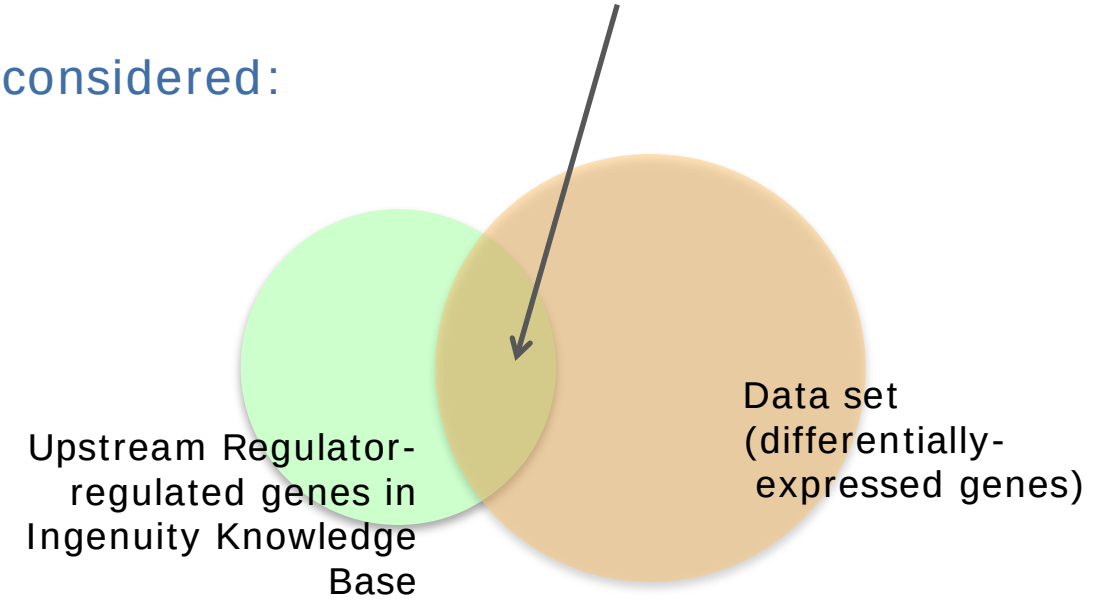
Approach: Two complementary statistical measures:

- Activation z-score
- Overlap p-value

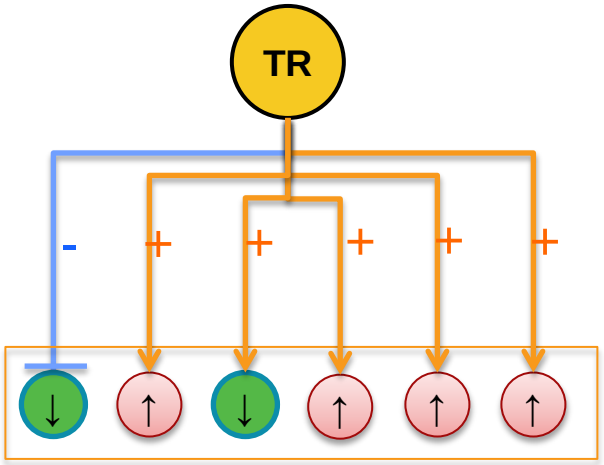
Evaluate the perturbed genes in the dataset that are known targets of a particular regulator

TR → target edge types considered:

- Expression
- Transcription
- Protein-DNA binding



Activation z-score



1 1 -1 1 1 1

Every TR is analyzed

Literature-based effect TR has on downstream genes

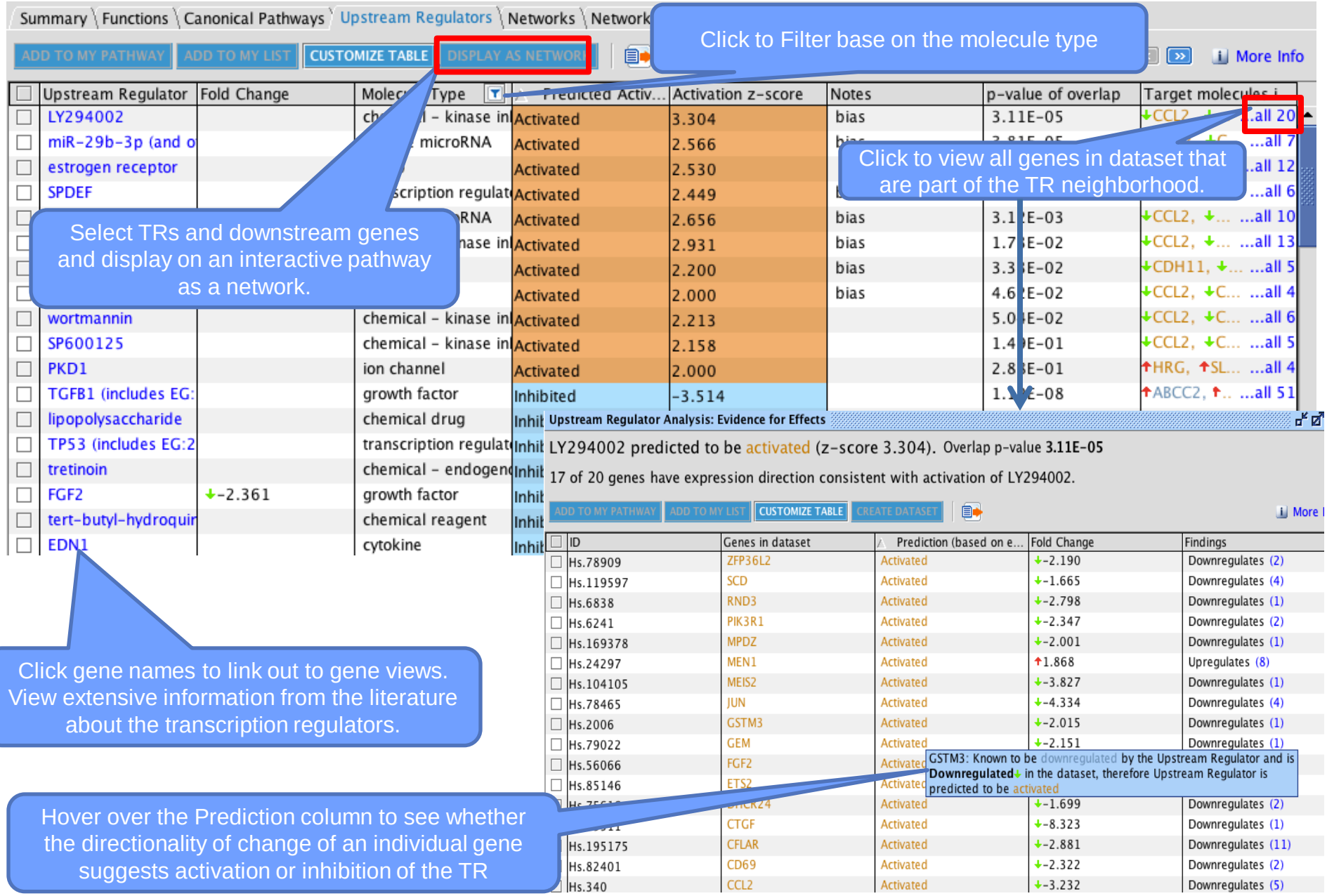
Differential Gene Expression (Uploaded Data)

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

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

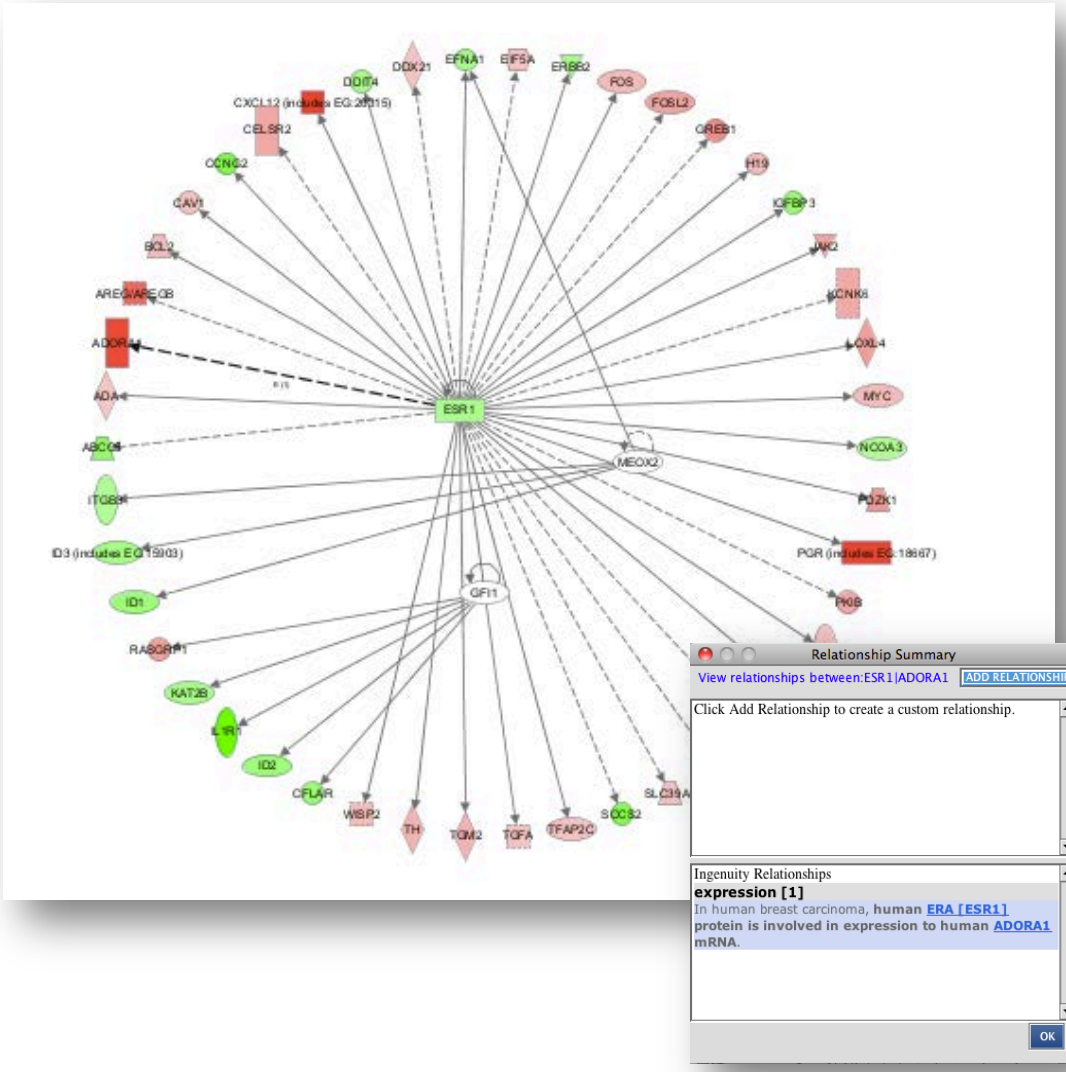
- z-score is statistical measure of correlation between relationship direction and gene expression.
- z-score > 2 or < -2 is considered significant

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


Summary | Functions | Canonical Pathways | Upstream Regulators | Networks | Network
ADD TO MY PATHWAY
ADD TO MY LIST
CUSTOMIZE TABLE
DISPLAY AS NETWORK
More Info
Click to Filter base on the molecule type
Upstream Regulator | Fold Change | Molecule Type | Predicted Activ... | Activation z-score | Notes | p-value of overlap | Target molecules i...
LY294002 | kinase in | Activated | 3.304 | bias | 3.11E-05 | ...all 20
miR-29b-3p (and o | microRNA | Activated | 2.566
estrogen receptor | Activated | 2.530 | ...all 12
SPDEF | scription regulat | Activated | 2.449 | ...all 6
RNA | Activated | 2.656 | bias | ...all 10
nase in | Activated | 2.931 | bias | ...all 13
Activated | 2.200 | bias | ...all 5
Activated | 2.000 | bias | ...all 4
wortmannin | chemical - kinase in | Activated | 2.213 | ...all 6
SP600125 | chemical - kinase in | Activated | 2.158 | ...all 5
PKD1 | ion channel | Activated | 2.000 | ...all 4
TGFB1 (includes EG: | growth factor | Inhibited | -3.514 | ...all 51
lipopolysaccharide | chemical drug
TP53 (includes EG:2 | transcription regulat
tretinoin | chemical - endogen
FGF2 | -2.361 | growth factor
tert-butyl-hydroquir | chemical reagent
EDN1 | cytokine
Select TRs and downstream genes and display on an interactive pathway as a network.
Click to view all genes in dataset that are part of the TR neighborhood.
Upstream Regulator Analysis: Evidence for Effects
LY294002 predicted to be activated (z-score 3.304). Overlap p-value 3.11E-05
17 of 20 genes have expression direction consistent with activation of LY294002.
ADD TO MY PATHWAY
ADD TO MY LIST
CUSTOMIZE TABLE
CREATE DATASET
ID | Genes in dataset | Prediction (based on e... | Fold Change | Findings
Hs.78909 | ZFP36L2 | Activated | -2.190 | Downregulates (2)
Hs.119597 | SCD | Activated | -1.665 | Downregulates (4)
Hs.6838 | RND3 | Activated | -2.798 | Downregulates (1)
Hs.6241 | PIK3R1 | Activated | -2.347 | Downregulates (2)
Hs.169378 | MPDZ | Activated | -2.001 | Downregulates (1)
Hs.24297 | MEN1 | Activated | 1.868 | Upregulates (8)
Hs.104105 | MEIS2 | Activated | -3.827 | Downregulates (1)
Hs.78465 | JUN | Activated | -4.334 | Downregulates (4)
Hs.2006 | GSTM3 | Activated | -2.015 | Downregulates (1)
Hs.79022 | GEM | Activated | -2.151 | Downregulates (1)
Hs.56066 | FGF2 | Activated
Hs.85146 | ETS2 | Activated
Activated | -1.699 | Downregulates (2)
CTGF | Activated | -8.323 | Downregulates (1)
Hs.195175 | CFLAR | Activated | -2.881 | Downregulates (11)
Hs.82401 | CD69 | Activated | -2.322 | Downregulates (2)
Hs.340 | CCL2 | Activated | -3.232 | Downregulates (5)
GSTM3: Known to be downregulated by the Upstream Regulator and is Downregulated in the dataset, therefore Upstream Regulator is predicted to be activated
Click gene names to link out to gene views. View extensive information from the literature about the transcription regulators.
Hover over the Prediction column to see whether the directionality of change of an individual gene suggests activation or inhibition of the TR

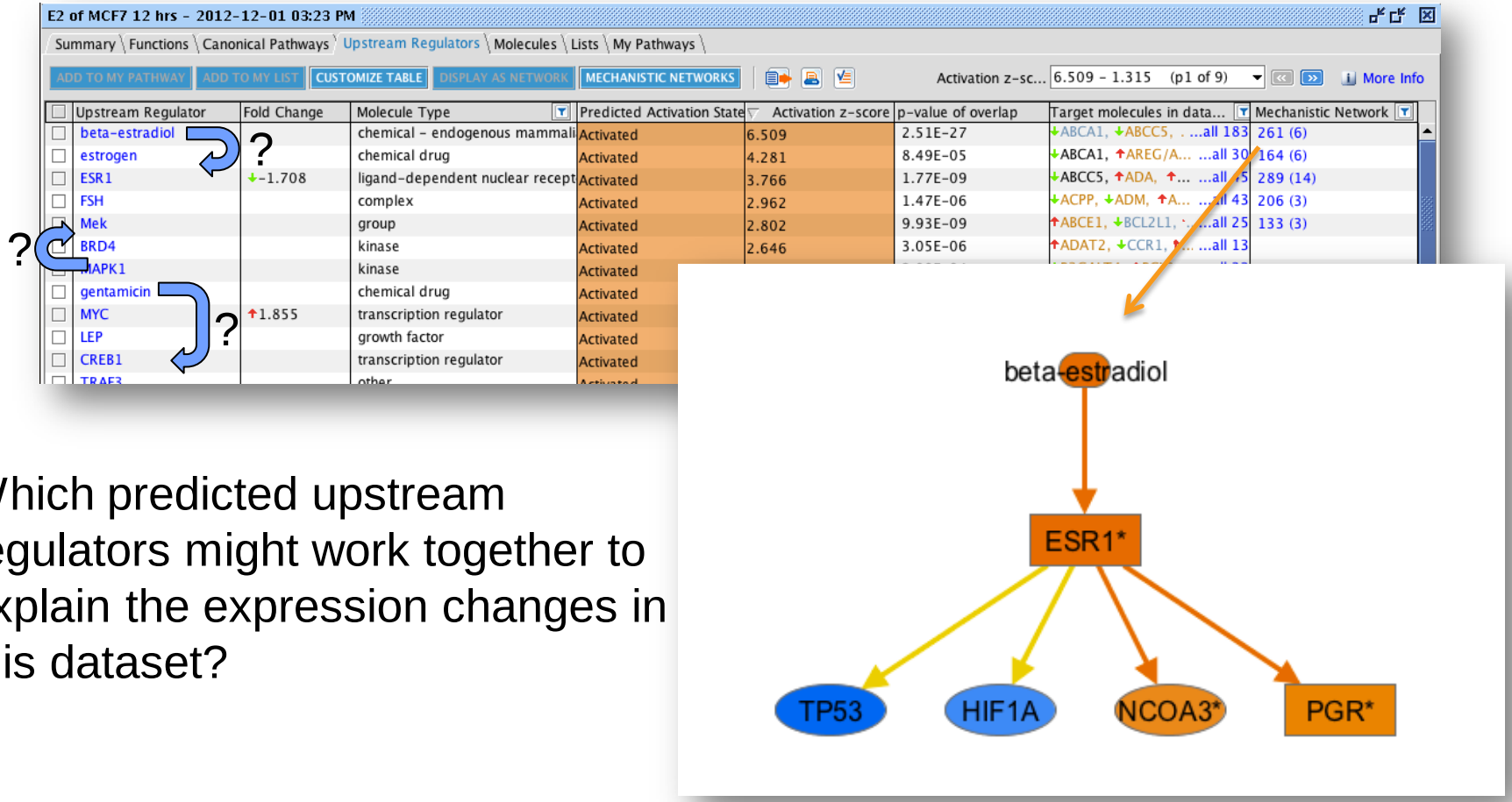
Identify cross-talk between Upstream Regulator in focused regulatory networks

- Automatically generate a directed TR-target network
- Add relationships to the regulatory network, e.g. upstream signaling molecules, or to disease, biological process associations
- See published evidence for the regulatory interactions



Mechanistic Networks

Upstream Regulator Analysis Results:

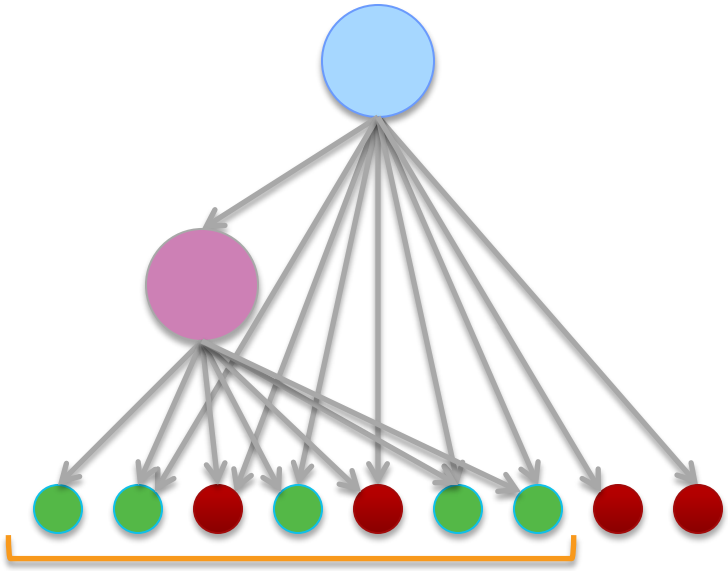


Which predicted upstream regulators might work together to explain the expression changes in this dataset?

Mechanistic Network Algorithm

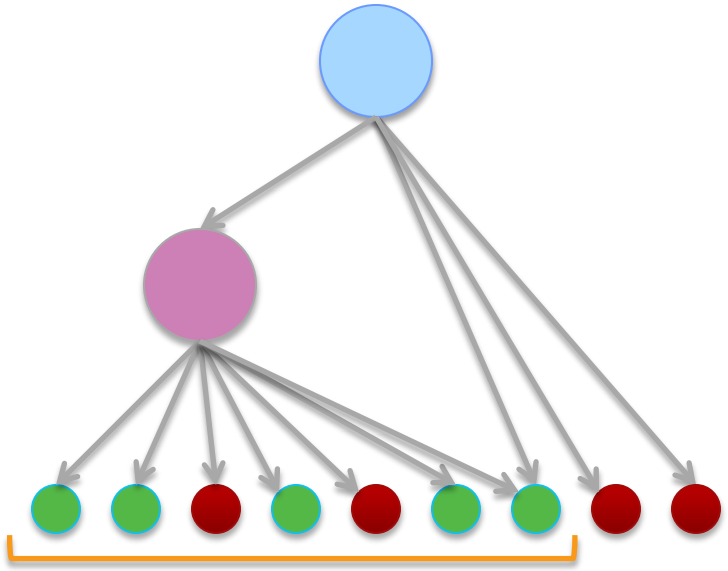
- Algorithm seeks large overlaps between an upstream regulator's targets and a more downstream regulator's targets

Upstream molecule **likely** to operate thru this more downstream regulator



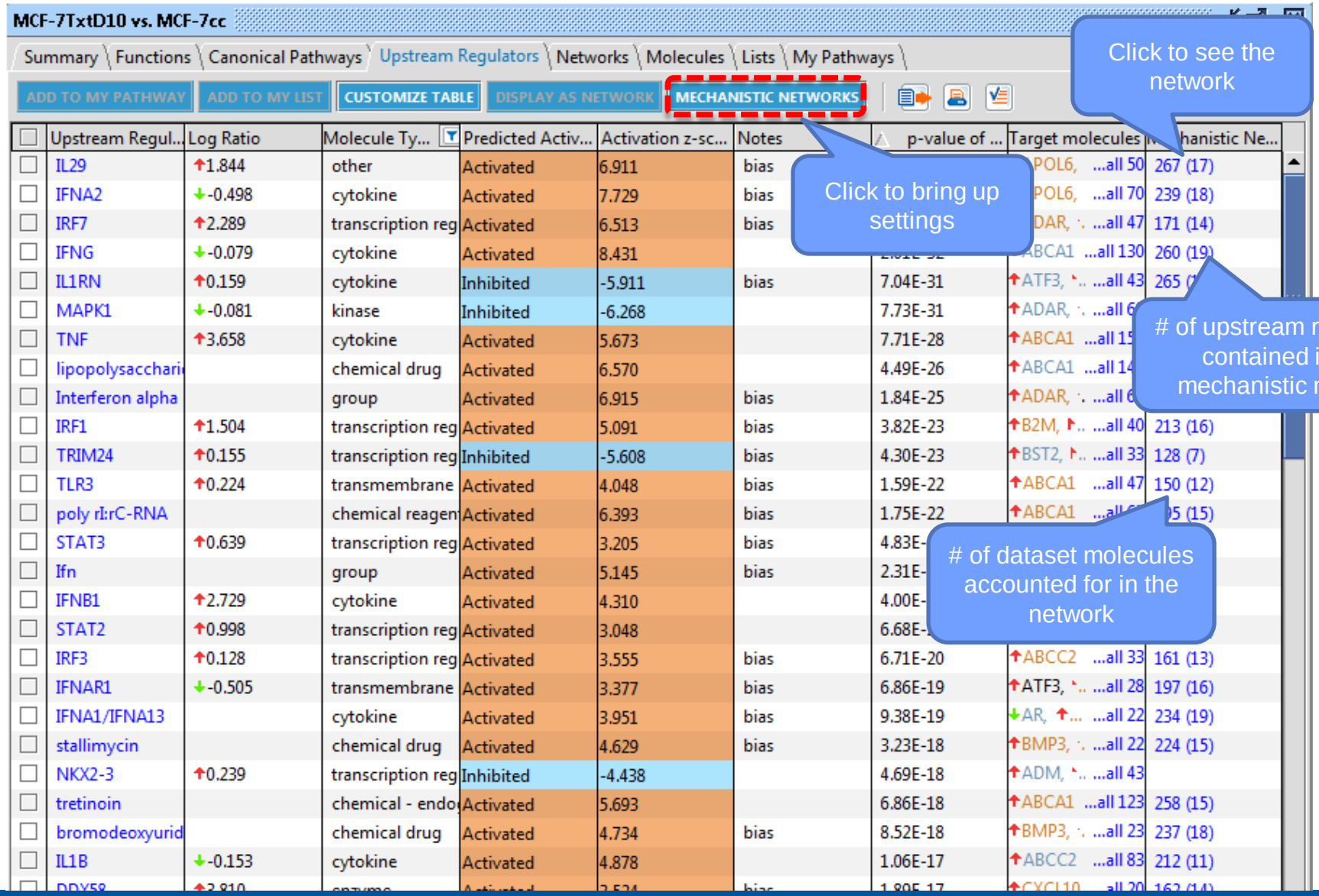
Shares **6 of 7** of the more downstream regulator's targets

Upstream molecule **less likely** to operate thru this more downstream regulator



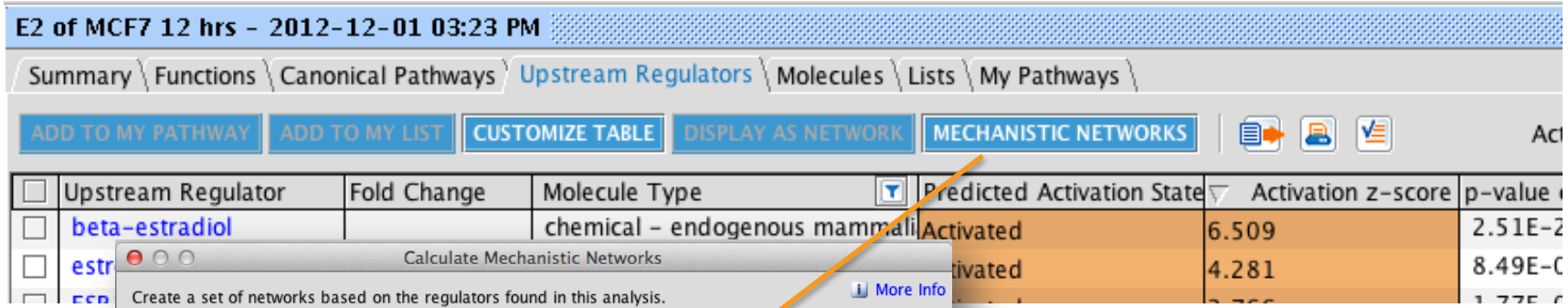
Shares **1 of 7** of the more downstream regulator's targets

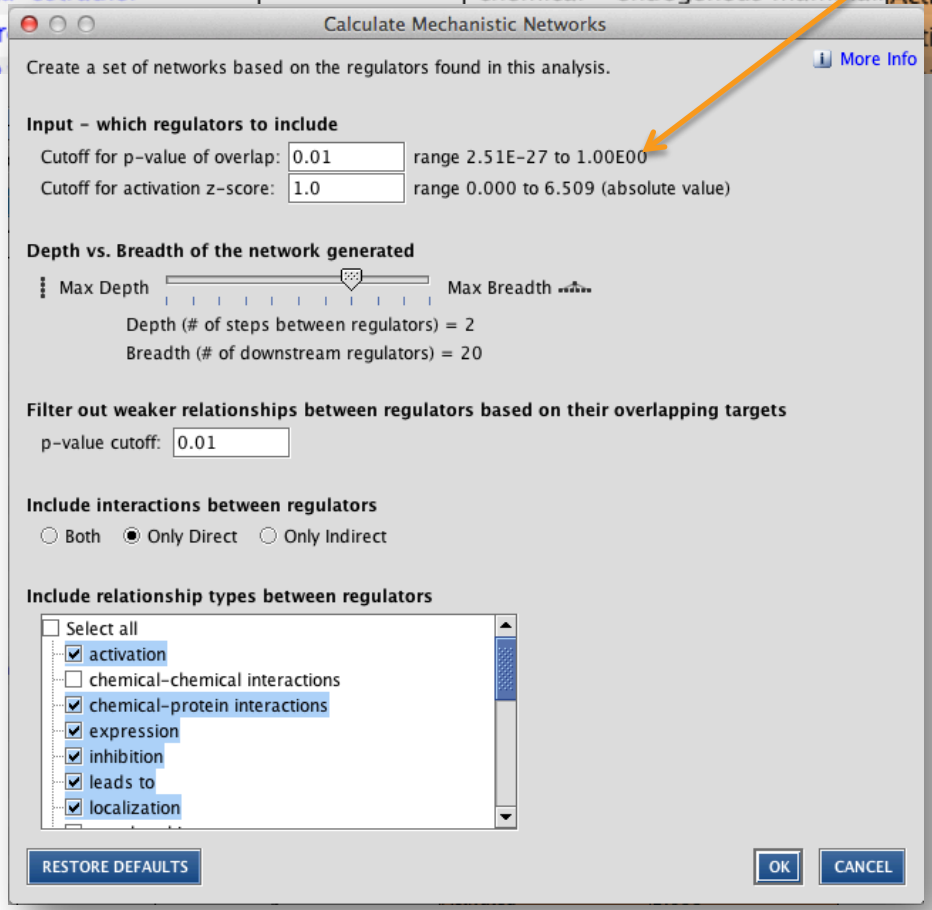
IPA Upstream Regulator Analysis results table with Mechanistic Networks



Upstream Regul...	Log Ratio	Molecule Ty...	Predicted Activ...	Activation z-sc...	Notes	p-value of ...	Target molecules	...hanistic Ne...
IL29	↑1.844	other	Activated	6.911	bias		...POL6, ...all 50	267 (17)
IFNA2	↓-0.498	cytokine	Activated	7.729	bias		...POL6, ...all 70	239 (18)
IRF7	↑2.289	transcription reg	Activated	6.513	bias		...DAR, ...all 47	171 (14)
IFNG	↓-0.079	cytokine	Activated	8.431			...ABCA1 ...all 130	260 (19)
IL1RN	↑0.159	cytokine	Inhibited	-5.911	bias	7.04E-31	↑ATF3, ...all 43	265 (...
MAPK1	↓-0.081	kinase	Inhibited	-6.268		7.73E-31	↑ADAR, ...all 6...	
TNF	↑3.658	cytokine	Activated	5.673		7.71E-28	↑ABCA1 ...all 15...	
lipopolysacchari		chemical drug	Activated	6.570		4.49E-26	↑ABCA1 ...all 14...	
Interferon alpha		group	Activated	6.915	bias	1.84E-25	↑ADAR, ...all 6...	
IRF1	↑1.504	transcription reg	Activated	5.091	bias	3.82E-23	↑B2M, ...all 40	213 (16)
TRIM24	↑0.155	transcription reg	Inhibited	-5.608	bias	4.30E-23	↑BST2, ...all 33	128 (7)
TLR3	↑0.224	transmembrane	Activated	4.048	bias	1.59E-22	↑ABCA1 ...all 47	150 (12)
poly rI:rC-RNA		chemical reagen	Activated	6.393	bias	1.75E-22	↑ABCA1 ...all 6...	...5 (15)
STAT3	↑0.639	transcription reg	Activated	3.205	bias	4.83E-		
Ifn		group	Activated	5.145	bias	2.31E-		
IFNB1	↑2.729	cytokine	Activated	4.310		4.00E-		
STAT2	↑0.998	transcription reg	Activated	3.048		6.68E-		
IRF3	↑0.128	transcription reg	Activated	3.555	bias	6.71E-20	↑ABCC2 ...all 33	161 (13)
IFNAR1	↓-0.505	transmembrane	Activated	3.377	bias	6.86E-19	↑ATF3, ...all 28	197 (16)
IFNA1/IFNA13		cytokine	Activated	3.951	bias	9.38E-19	↓AR, ↑... ...all 22	234 (19)
stallimycin		chemical drug	Activated	4.629	bias	3.23E-18	↑BMP3, ...all 22	224 (15)
NKX2-3	↑0.239	transcription reg	Inhibited	-4.438		4.69E-18	↑ADM, ...all 43	
tretinoin		chemical - endo	Activated	5.693		6.86E-18	↑ABCA1 ...all 123	258 (15)
bromodeoxyurid		chemical drug	Activated	4.734	bias	8.52E-18	↑BMP3, ...all 23	237 (18)
IL1B	↓-0.153	cytokine	Activated	4.878		1.06E-17	↑ABCC2 ...all 83	212 (11)
DDX58	↑3.810	enzyme	Activated	2.524	bias	1.80E-17	↑CXCL10 ...all 20	162 (14)

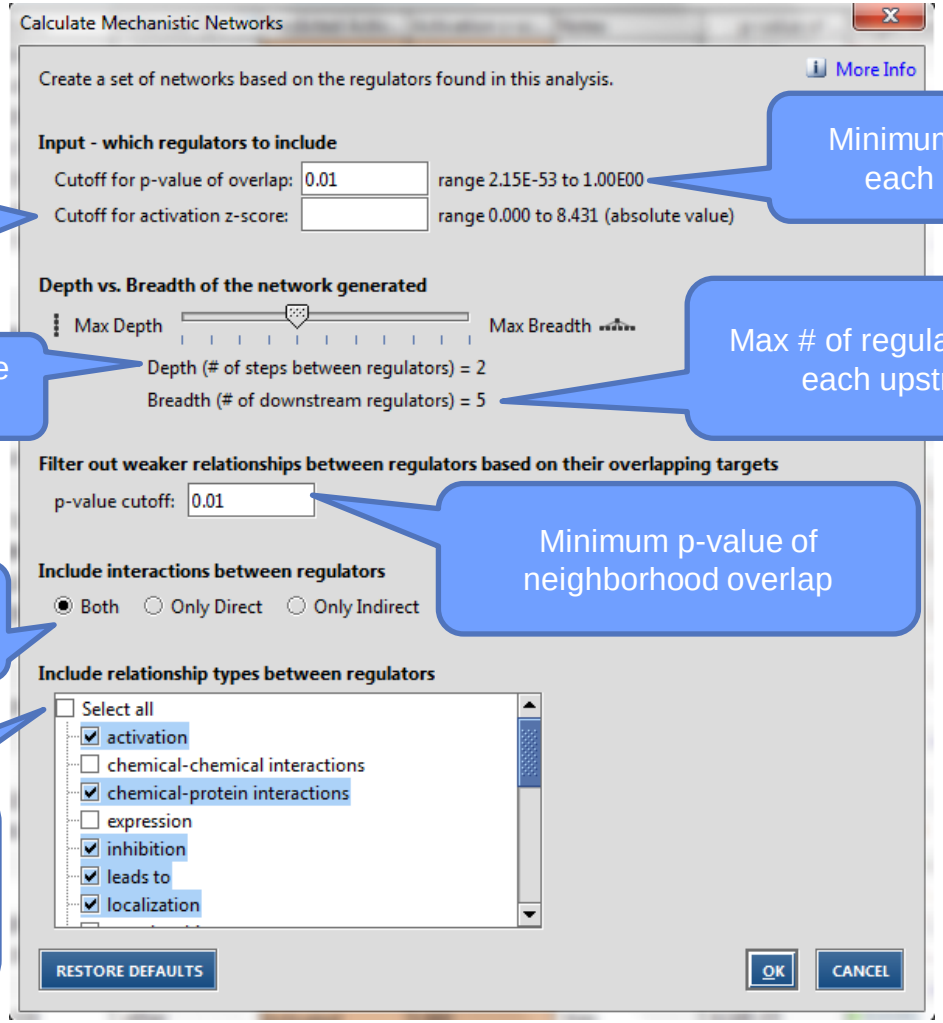
You can change the settings and re-run





New results replace the original ones

Mechanistic Network Settings



Minimum p-value of each regulator

Minimum absolute z-score of each regulator

Max # of layers below the top regulator

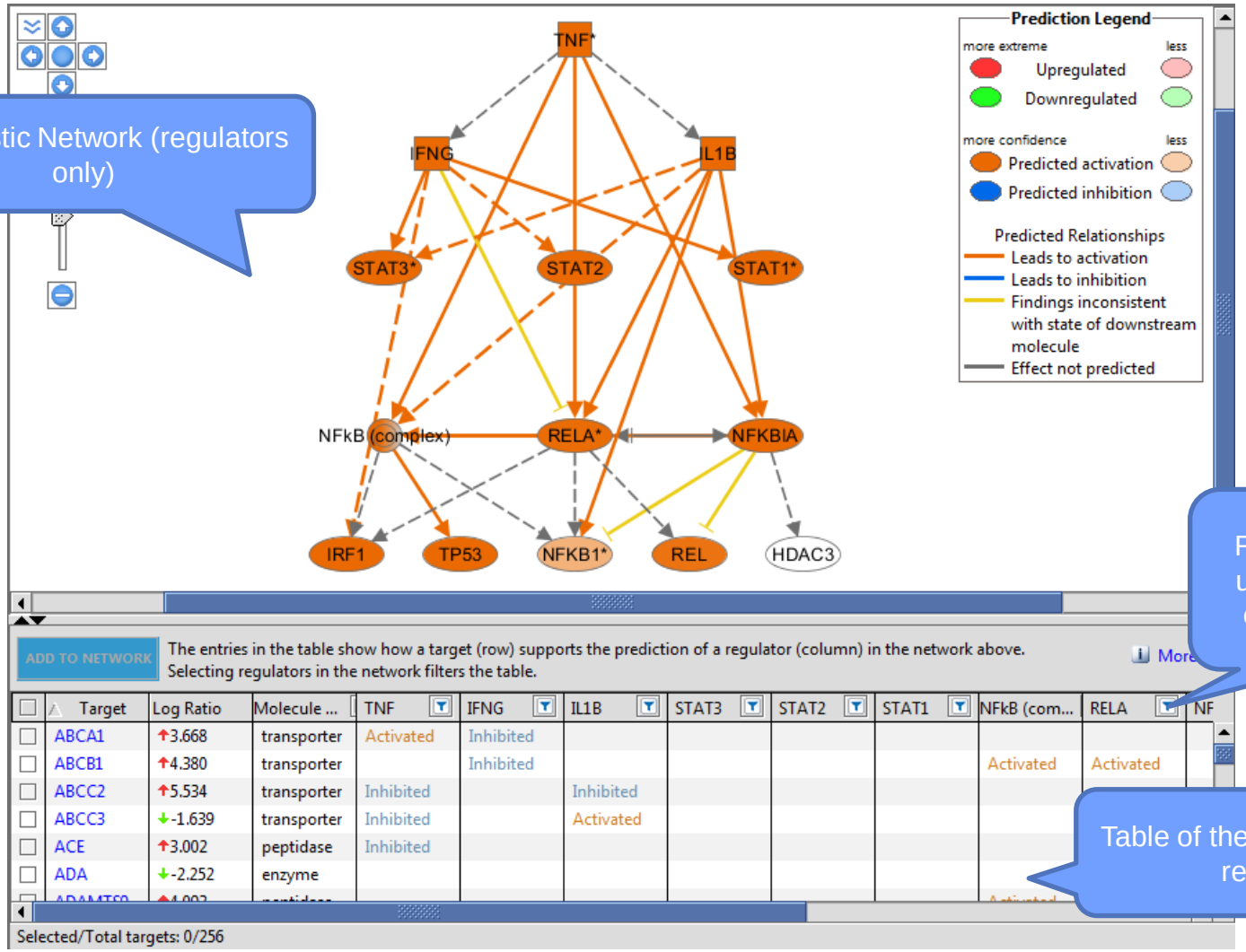
Max # of regulators downstream of each upstream regulator

Minimum p-value of neighborhood overlap

Interaction types

Types of relationships that can connect regulators

Example: TNF mechanistic network



Target	Log Ratio	Molecule ...	TNF	IFNG	IL1B	STAT3	STAT2	STAT1	NFkB (com...	RELA	NF
ABCA1	↑3.668	transporter	Activated	Inhibited							
ABCB1	↑4.380	transporter		Inhibited					Activated	Activated	
ABCC2	↑5.534	transporter	Inhibited		Inhibited						
ABCC3	↓-1.639	transporter	Inhibited		Activated						
ACE	↑3.002	peptidase	Inhibited								
ADA	↓-2.252	enzyme									

Network analysis
(Different from Mechanistic Networks)

Networks in IPA

- Purpose:
  - To show as many interactions between user-specified molecules in a given dataset and how they might work together at the molecular level
- Why are Ingenuity networks biologically interesting?
  - 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.

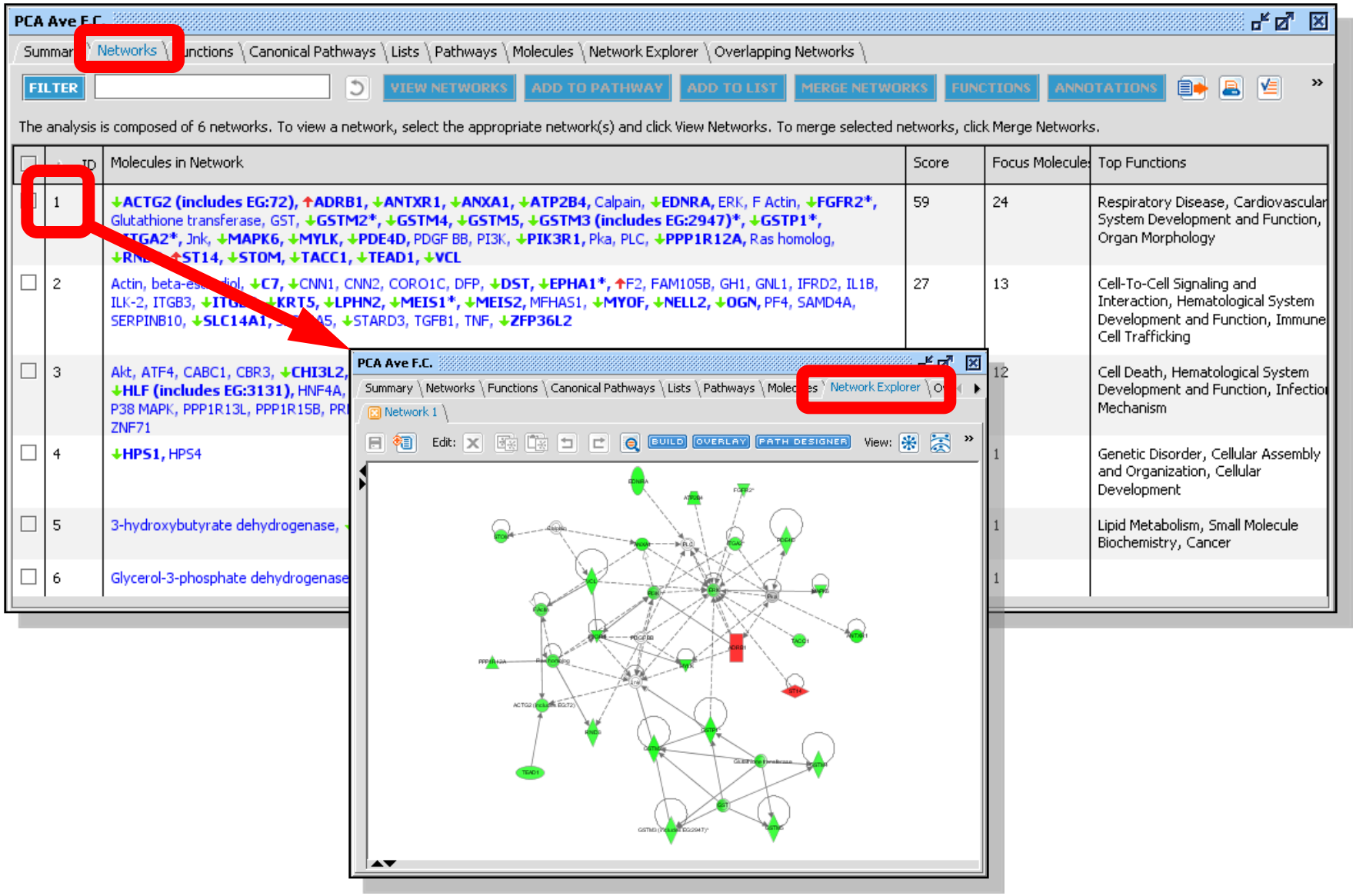
Networks, cont.

- Keep in mind:
  - Networks may contain smaller networks of connectivity related to specific functions, i.e., don't be afraid to subset a network.
  - Larger cellular networks may span IPA assembled networks, i.e., don't be afraid to merge networks.
  - Networks should be treated as "starter pathways" that you then modify based on your biological understanding and questions you want answered using pathway building (Build button) and analysis (overlay button) tools.

Key Terminology Review

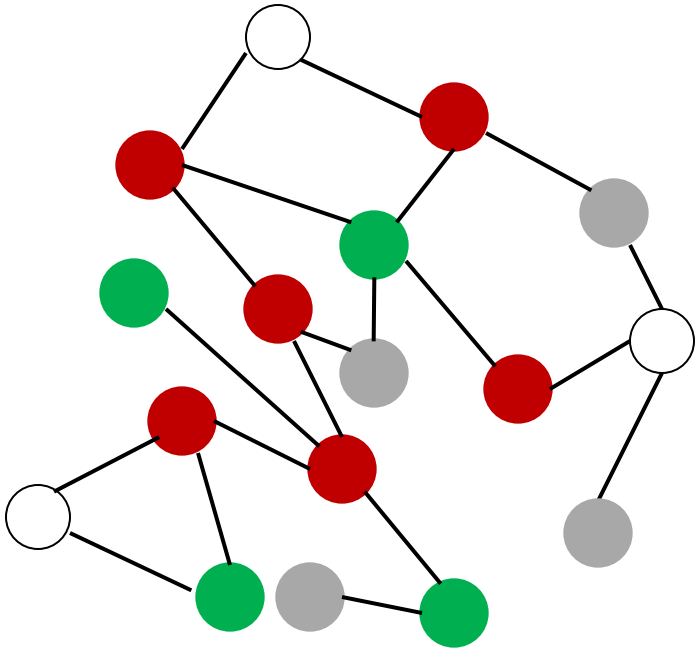
- Focus Molecule:
  - Molecules that are from uploaded list, pass filters are applied, and are available for generating networks
- Networks:
  - Generated de novo based upon input data
  - Do not have directionality
  - Contain molecules involved in a variety of Canonical Pathways
- Canonical Pathways (Signaling and Metabolic)
  - Are generated prior to data input, based on the literature
  - Do NOT change upon data input
  - Do have directionality (proceed “from A to Z”)
- My Pathways and Path Designer Pathways
  - Custom built pathways manually created based on user input

Viewing Networks



How Networks Are Generated

1. Focus molecules are “seeds”
2. Focus molecules with the most interactions to other focus molecules are then connected together to form a network
3. Non-focus molecules from the dataset are then added
4. Molecules from the Ingenuity’s KB are added
5. Resulting Networks are scored and then sorted based on the score



35 molecules per network for visualization purposes

Example Network