

Maximizing the Biological Interpretation of Gene, Transcript & Protein Expression Data with IPA (Ingenuity Pathway Analysis)

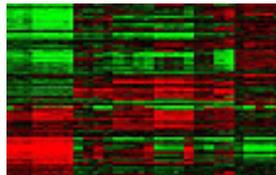
Yvonne Trulin
Account Manager
ytrulin@ingenuity.com

Darryl Gietzen, Ph.D.
Sr. Application Scientist
dgietzen@ingenuity.com

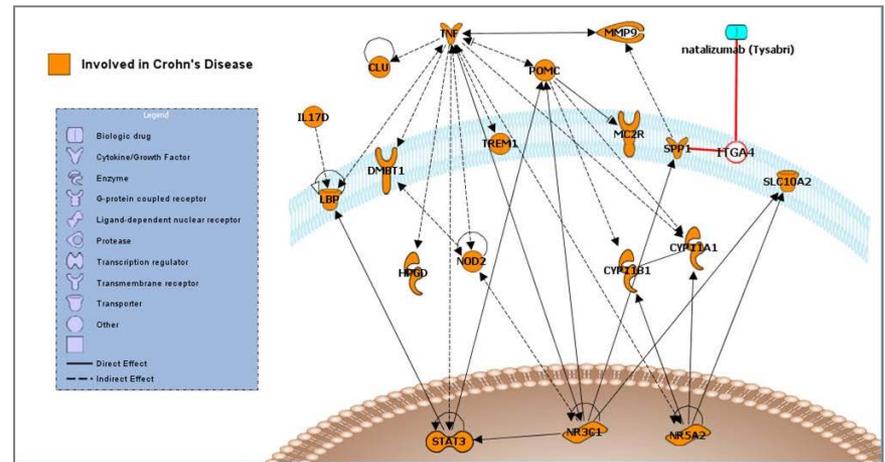
Biological Interpretation Based on Assimilating Published Experimental Findings

	A	B	C
1	AOC3		
2	CCL19		
3	CD58		
4	CXCL12		
5	ITK		
6	MYD88		
7	PLAU		
8	RAP1GAP		
9	RASGRP1		
10	SPP1		
11	TGFB1		
12	THBS1		
13	TLR4		
14	VCAM1		

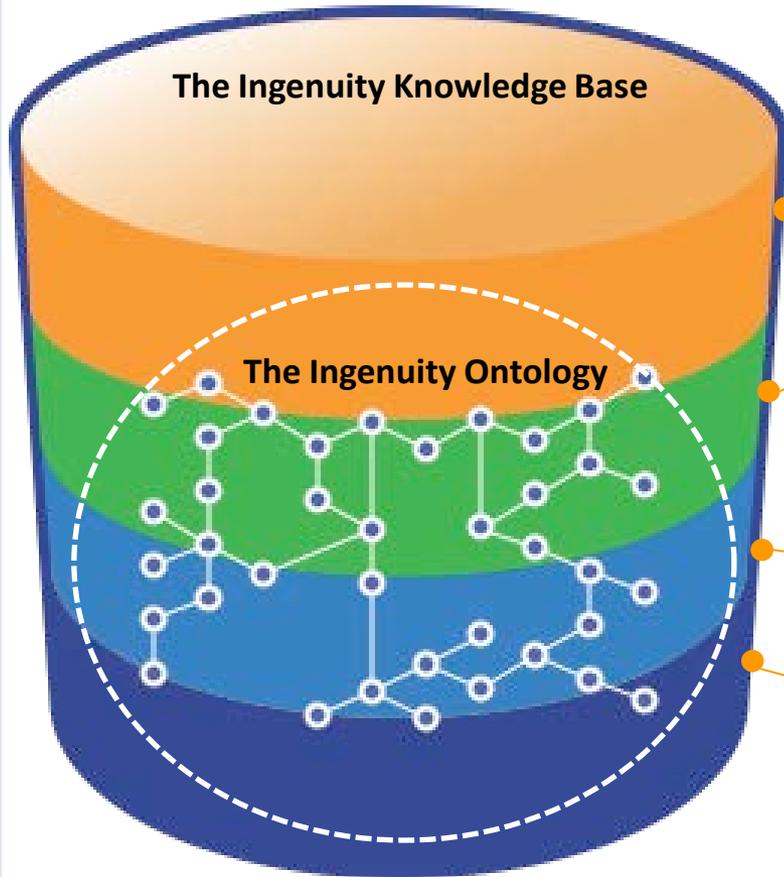
→
IPA



→
IPA



Ingenuity Content



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

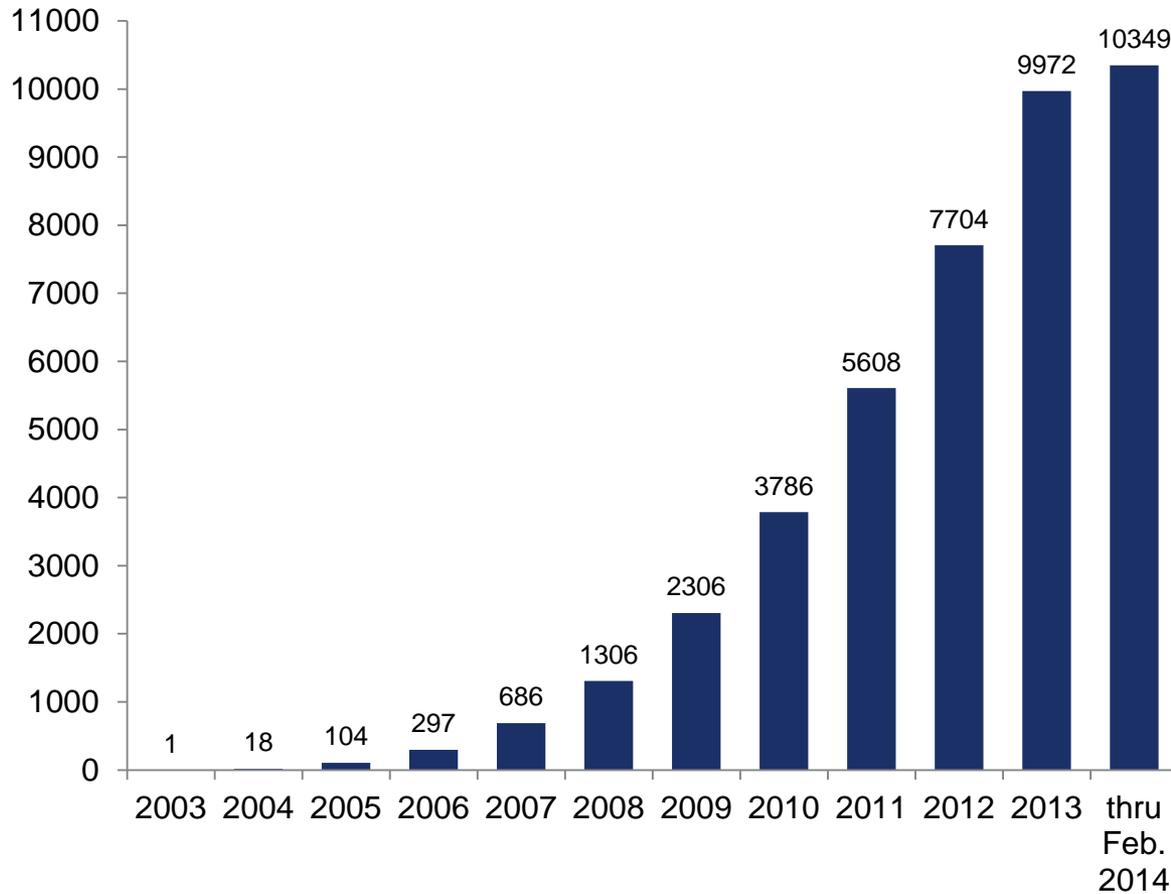
Ingenuity® Expert Knowledge – Content we model such as pathways, toxicity lists, etc.

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



10,351 publications and growing!

Peer-reviewed publications citing QIAGEN's Ingenuity products



Expression, Proteomics, SNP, Copy Number, RNAi, miRNA,

Oncology, Cardiovascular Disease, Neuroscience, Metabolic Disease, Inflammation/Immunology, Infectious Disease

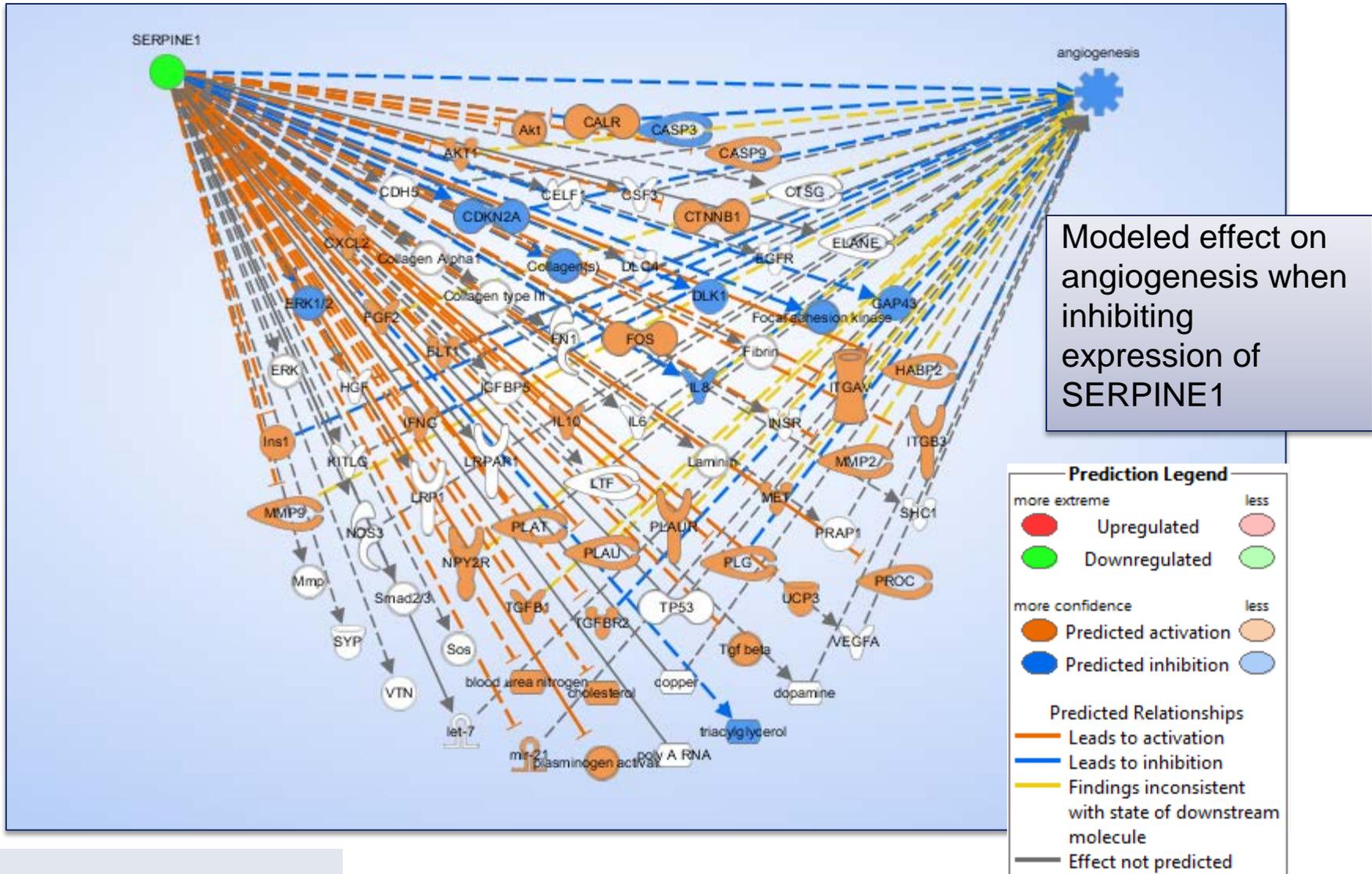
Basic, Translational, Drug Discovery & Development Research

How can IPA help you?

- Deep pathway understanding of a single gene/protein
 - Drug/therapeutic target discovery

- Biological understanding of large data sets
 - Differential gene expression, array and RNAseq (transcriptomics)
 - Differential protein expression (proteomics)
 - Protein phosphorylation
 - Methylation
 - Metabolomics
 - miRNA expression
 - Chip-on-chip / chip seq
 - Gene List
 - siRNA screening

SERPINE1 downstream effect modeling



How can IPA help you?

- Deep pathway understanding of a single gene/protein
 - Drug/therapeutic target discovery

- Biological understanding of large data sets
 - Differential gene expression, array and RNAseq (transcriptomics)
 - Differential protein expression (proteomics)
 - Protein phosphorylation
 - Methylation
 - Metabolomics
 - miRNA expression
 - Chip-on-chip / chip seq
 - Gene List
 - siRNA screening

Gene/Protein Expression 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

Data Upload Format Examples

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
1	130786	0.14
2	015380	-0.99
3	138932	-0.02
4	014576	-0.02
5	138933	0.02
6	000014	-4.79
7	026971	-0.67
8	144670	-5.96
9	001080438	-1.97
10	017436	-1.09
11	016161	2.02
12	015665	-0.27

+significance stat

	B	C
	Log2Ratio	p-value
36	0.14	8.68E-01
40	-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
71	-0.67	6.17E-01
9	-5.96	1.30E-01
10	-1.97	3.47E-01
11	-1.09	5.02E-01
12	2.02	5.97E-02
13	-0.27	5.68E-01

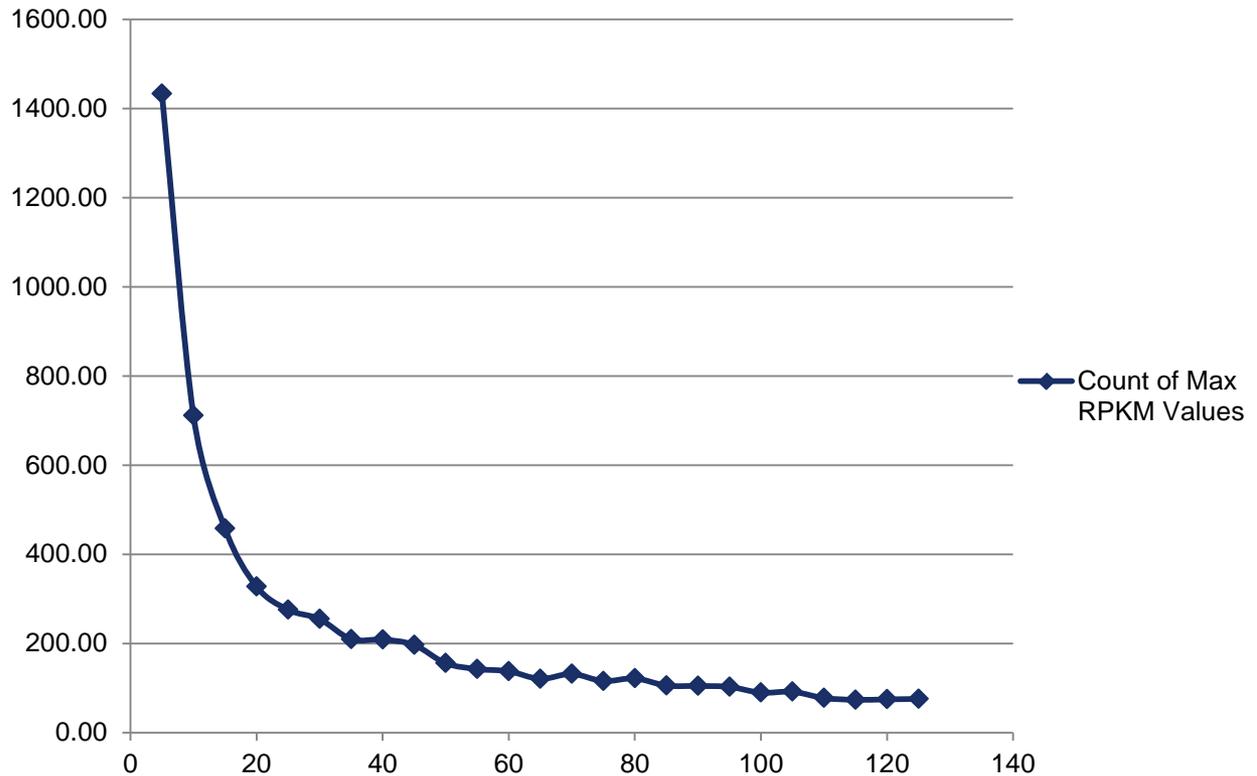
+RPKM

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

	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	-5.96	1.30E-01	610.64
10	-1.97	3.47E-01	3.91
11	-1.09	5.02E-01	6186.83
12	2.02	5.97E-02	149.85
13	-0.27	5.68E-01	13330.34

Typical Distribution of RPKM Values in RNAseq Data

- 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



Uploading Multiple Observations

- 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 (up to 20 observations)
 - 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_023078	1.12	1.03E-02	22828.15	0.17	7.18E-01	3330.36

Canonical Pathways | IL-10 Signaling

Overlay: **MAP (Molecule Activity Predictor)**

Prediction Legend

more extreme	less
● Upregulated	● less
● Downregulated	● less
more confidence	less
● Predicted activation	● less
● Predicted inhibition	● less

Predicted Relationships

- Leads to activation
- Leads to inhibition
- Findings inconsistent with state of downstream molecule
- Effect not predicted

IL-10 Signaling Pathway Diagram:

Left Panel:

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.

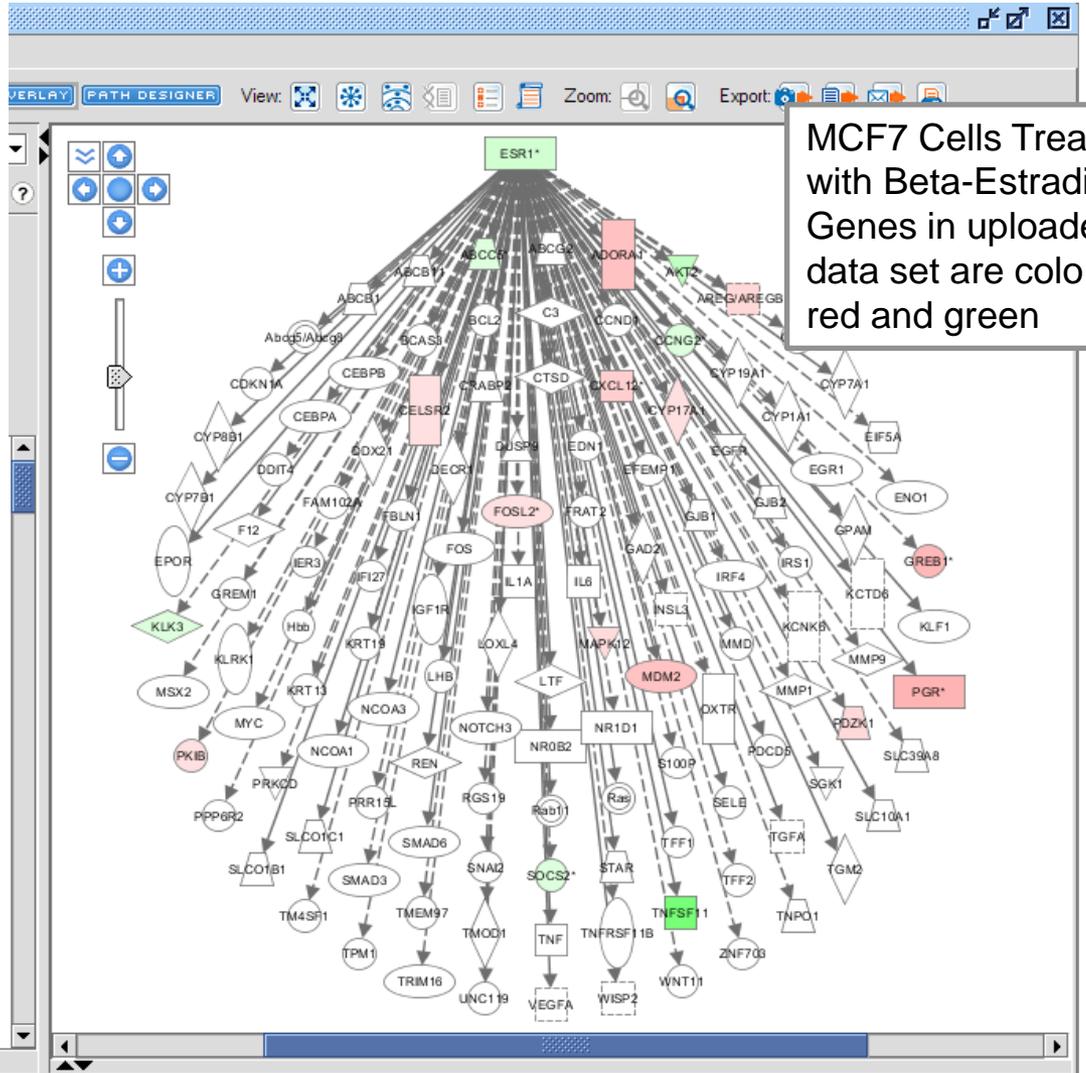
Use expression values from a Dataset or Analysis

Current Analysis/Dataset/List: Day 10
[Change Analysis/Dataset/List](#)

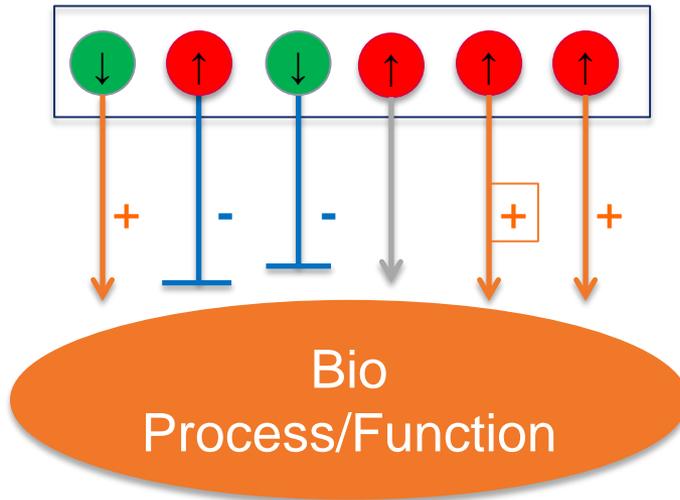
© 2000-2012 Ingenuity Systems, Inc. All rights reserved.

Identify likely upstream regulators and their activity state

- 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



Downstream Effect z-score



Differential Gene Expression
(Uploaded Data)

effect genes have on process or function based literature

-1 -1 1 0 1 1

Predicted Effect-

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

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

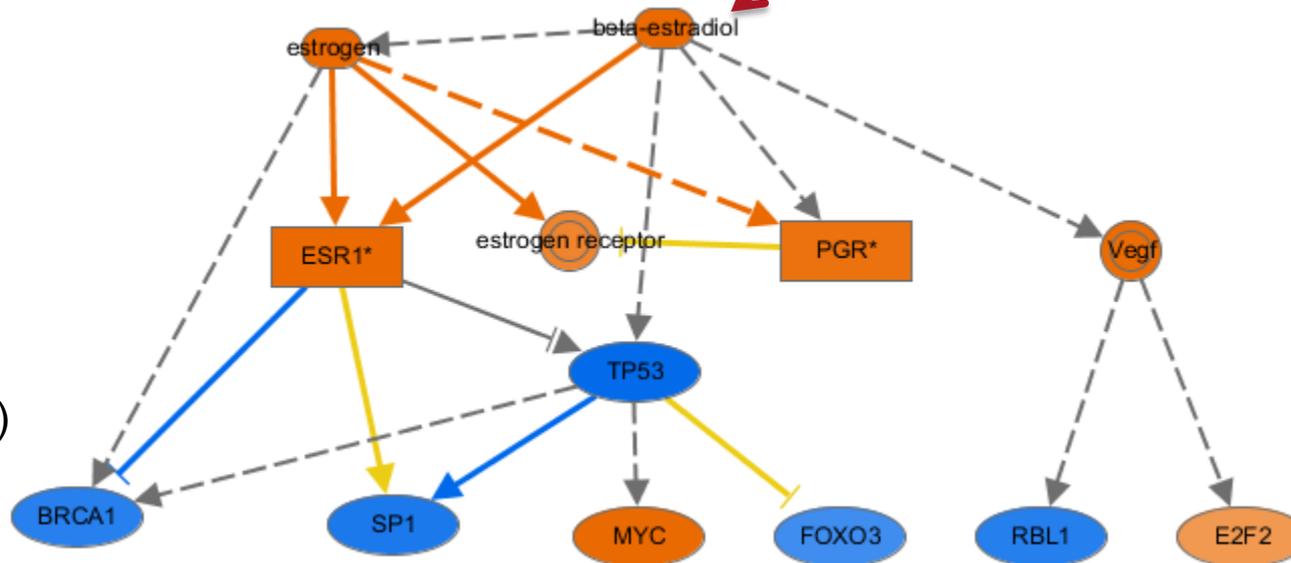
- “z-score” is statistical measure of correlation between relationship direction and gene expression.
- z-score > 2 or < -2 is considered significant

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

Create *de novo* pathways of regulators and genes

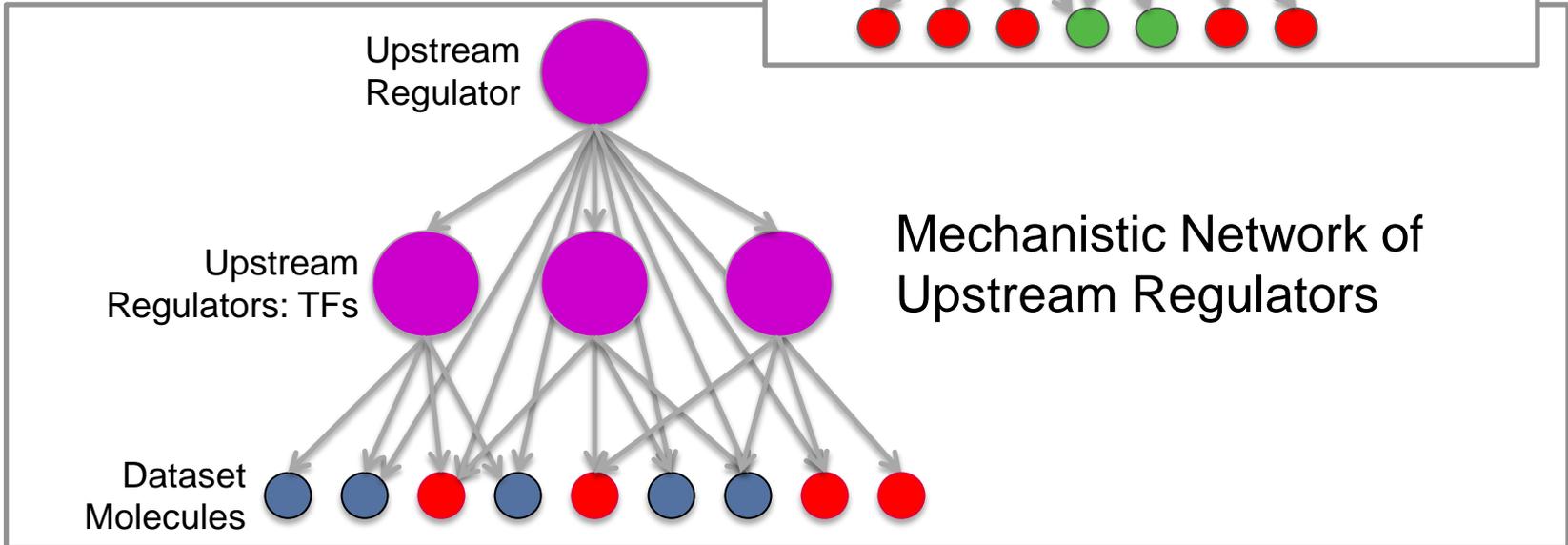
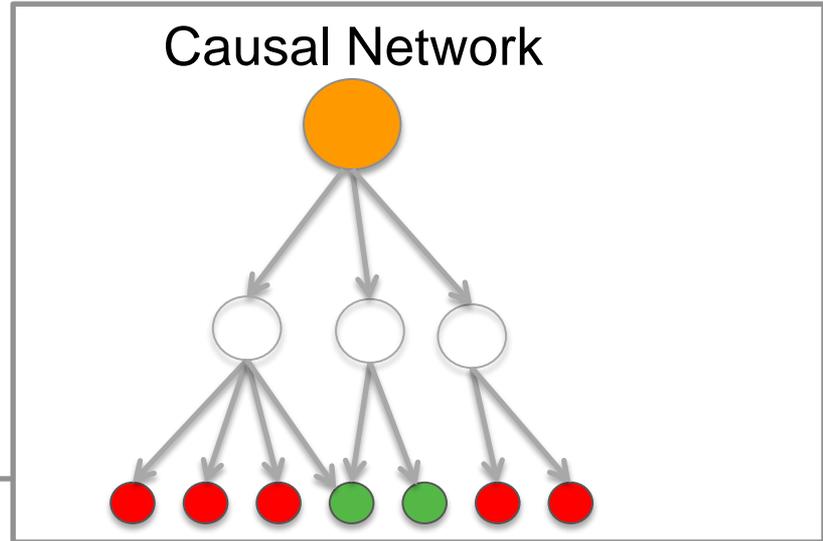
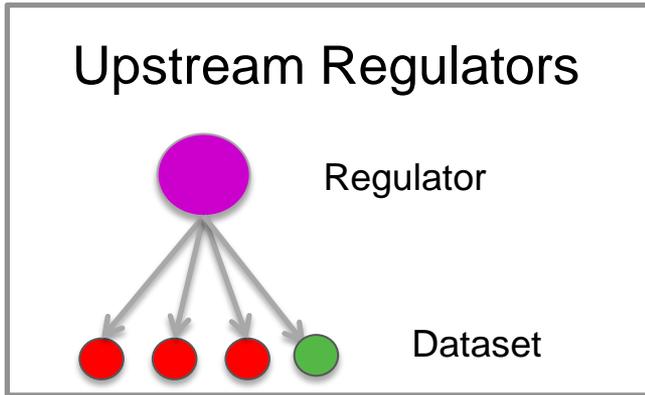
Upstream Regulator	Fold Change	Molecule Type	Predicted Activation	Activation z-sc...	p-value of overlap	Target molecules in	Mechanistic Net...
<input type="checkbox"/> beta-estradiol		chemical - endogen	Activated	6.097	1.24E-26	↓ABCA1, ↓... ...all 122	186 (13)
<input type="checkbox"/> Mek		group	Activated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

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



Single- vs. Mechanistic- vs. Causal Networks

Leveraging the network to create more upstream regulators

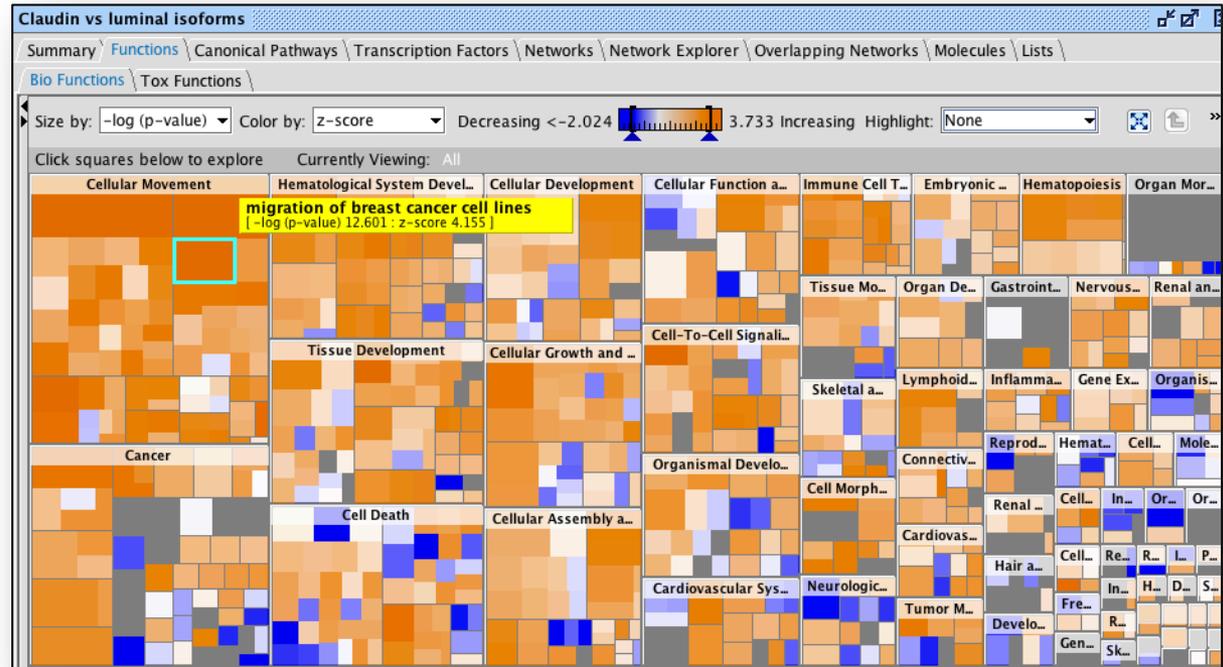


Downstream Effects Analysis

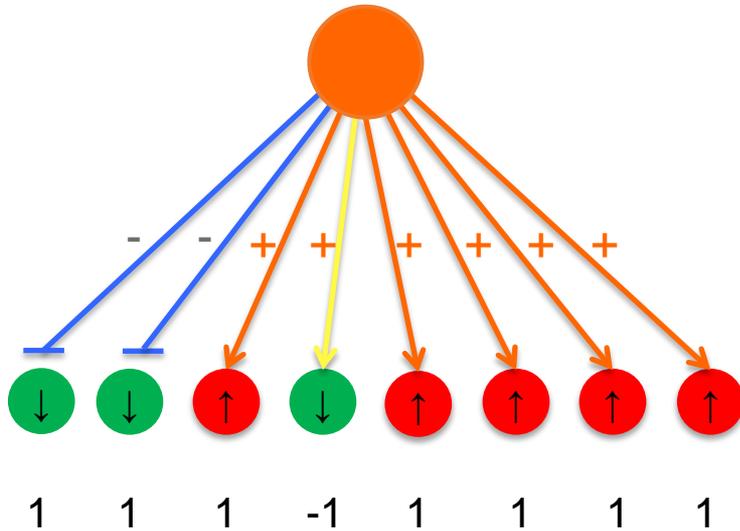
A novel approach to visualize and predict biological impact of gene expression changes

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



← Every possible TF & Upstream Regulator in the Ingenuity Knowledge Base is analyzed

← Literature-based effect TF/UR has on downstream genes

← Differential Gene Expression (Uploaded Data)

← Predicted activation state of TF/UR:

1 = Consistent with activation of UR

-1 = Consistent with inhibition of UR

$$z = \frac{x}{\sigma_x} = \frac{\sum_i x_i}{\sqrt{N}} = \frac{N_+ - N_-}{\sqrt{N}} = (7-1)/\sqrt{8} = 2.12 \text{ (= predicted activation)}$$

- z-score is a statistical measure of the match between expected relationship direction and observed gene expression
- z-score > 2 or < -2 is considered significant

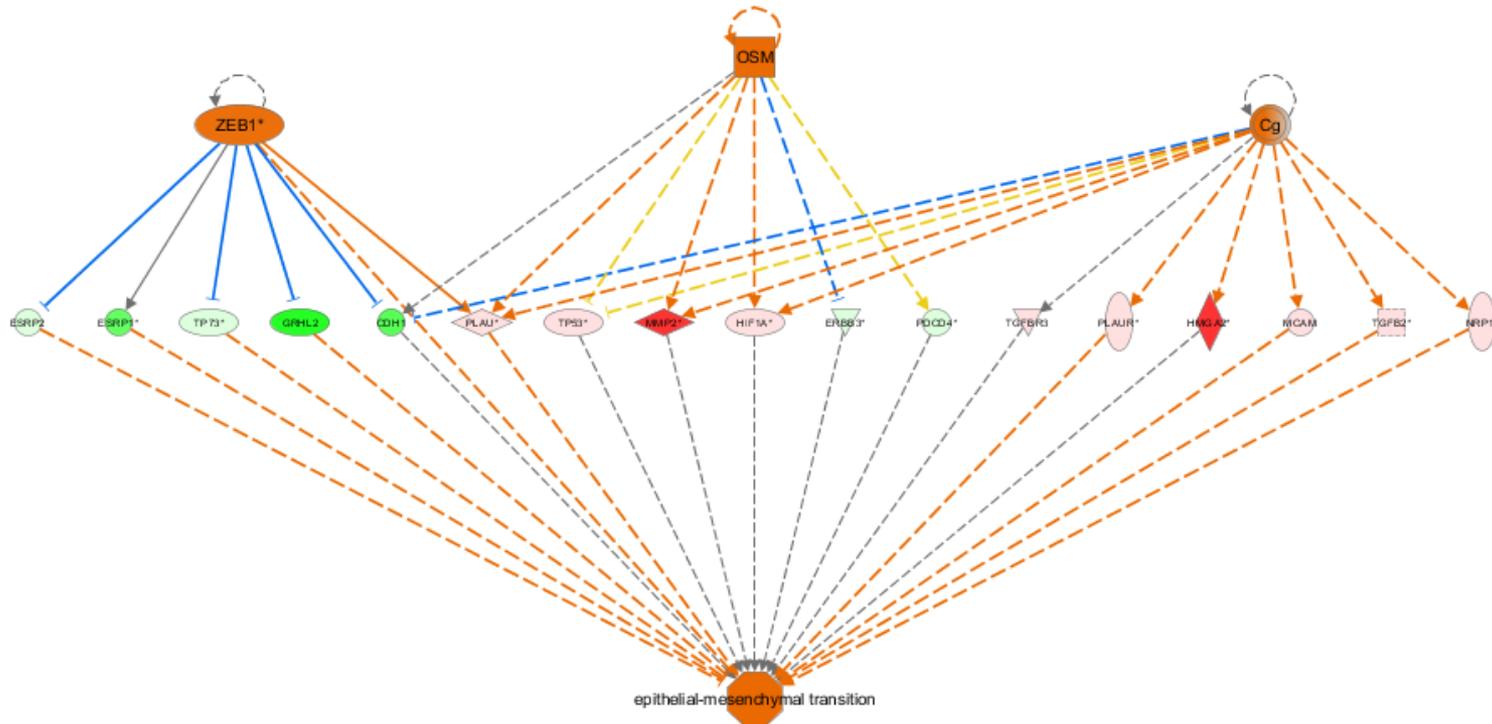
Note that the actual z-score is weighted by the underlying findings, the relationship bias, and dataset bias

Regulator Effects

New in IPA Spring 2014 Release (End of March)

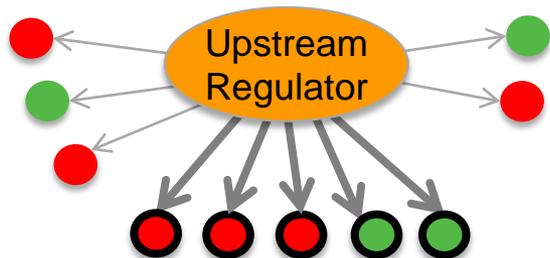
Hypothesis for how a phenotype, function or disease is regulated in the dataset by activated or inhibited upstream regulators

- 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

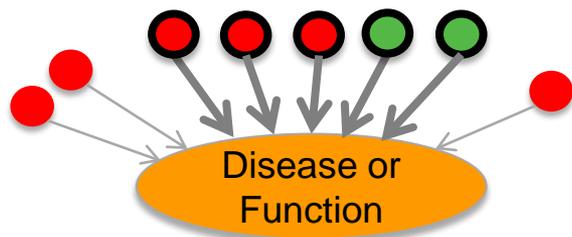


Hypotheses for how activated or inhibited upstream regulators cause downstream effects on biology

Upstream Regulator Analysis



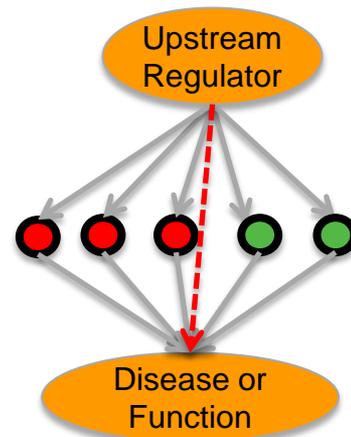
Targets in the dataset



Downstream Effects Analysis

Algorithm
 →
 First iteration

Simplest Regulator Effects result

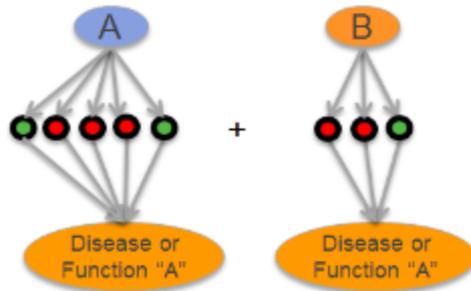


Displays a relationship between the regulator and disease/function if it exists

Causally consistent networks score higher

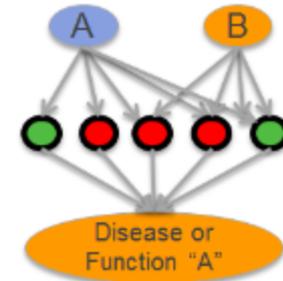
The algorithm runs iteratively to merge additional regulators with diseases and functions

Upstream Regulator Analysis



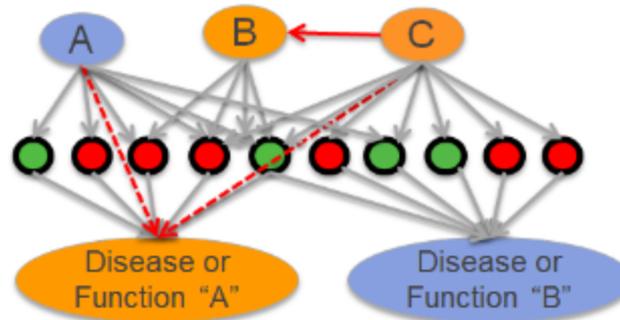
Algorithm
Next
iteration

Regulator Effects



Downstream Effects Analysis

Additional
iterations



Displays relationship
between the regulators as
well as regulators and
functions if they exist

Gene/Protein Expression 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