Do It Yourself Systems Biology:
The Hidden Treasure in Pathway Analysis

Ally Perlina
Bioinformatics Scientist
Bldg TEN Rm 2405 x4784
aperlina@sbmri.org

TODAY

- Analyze data from your omics experiments/ public datasets or your favorite list
- Search the database for connections between diseases, processes, and pathways of interest
- No Data? No problem! You can test your hypothesis on the basis of what's already known by retrieving connections up-stream or downstream of specific molecular entities and seeing how they relate to your disease(s) or cellular processes of interest.

WHAT TYPES OF QUESTIONS CAN BE ANSWERED WITH METACORE?

- What are the most relevant biological pathways for my data?
- What is known about any particular gene/ protein/ compound? Which canonical pathways is it involved in? Which diseases are associated with it? Where is it expressed?
- What are the known interactions downstream of my favorite gene/protein/microRNA? How does my data reflect this?
- What are the differences or similarities between multiple experimental conditions or species/cell lines? Or between different data types?
- What are the most important genes in my gene list?
  - What are they interacting with?
  - Which genes are involved in my disease of interest?
  - What are some important hubs responsible for signal regulation in my data?
  - Are there known therapeutic targets in my list? How are they connected to each other and to my significant genes?

KNOWLEDGE BASE BEHIND METACORE

Data level

~3,000 journals

45 PhD's and 5 M.D.'s

Manually curated content

Annotation level

High quality

"small experiment evidence"

3 months training

3 levels of supervision

Data storage level

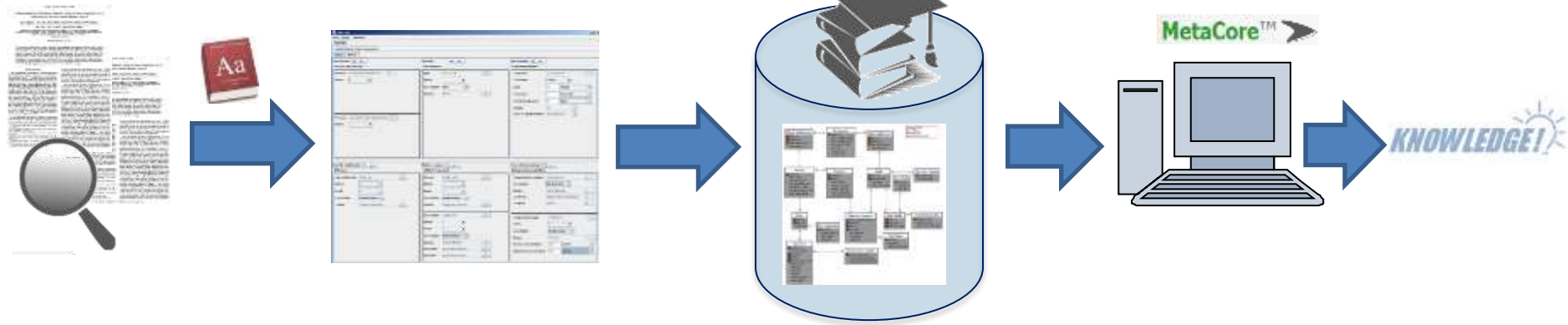
Oracle tables

Ontologies

Controlled vocabulary

Knowledge level/VALUE

GUI access to MetaBase content + analysis + reporting solutions



Systems Biology Content
+ Context at your fingertips

Sanford | Burnham
Medical Research Institute

RESEARCH WITHOUT SYSTEMS BIOLOGY

- Without System Biology researchers would have to spend decades trying to find important genes and proteins among terabytes of High-Throughput data;

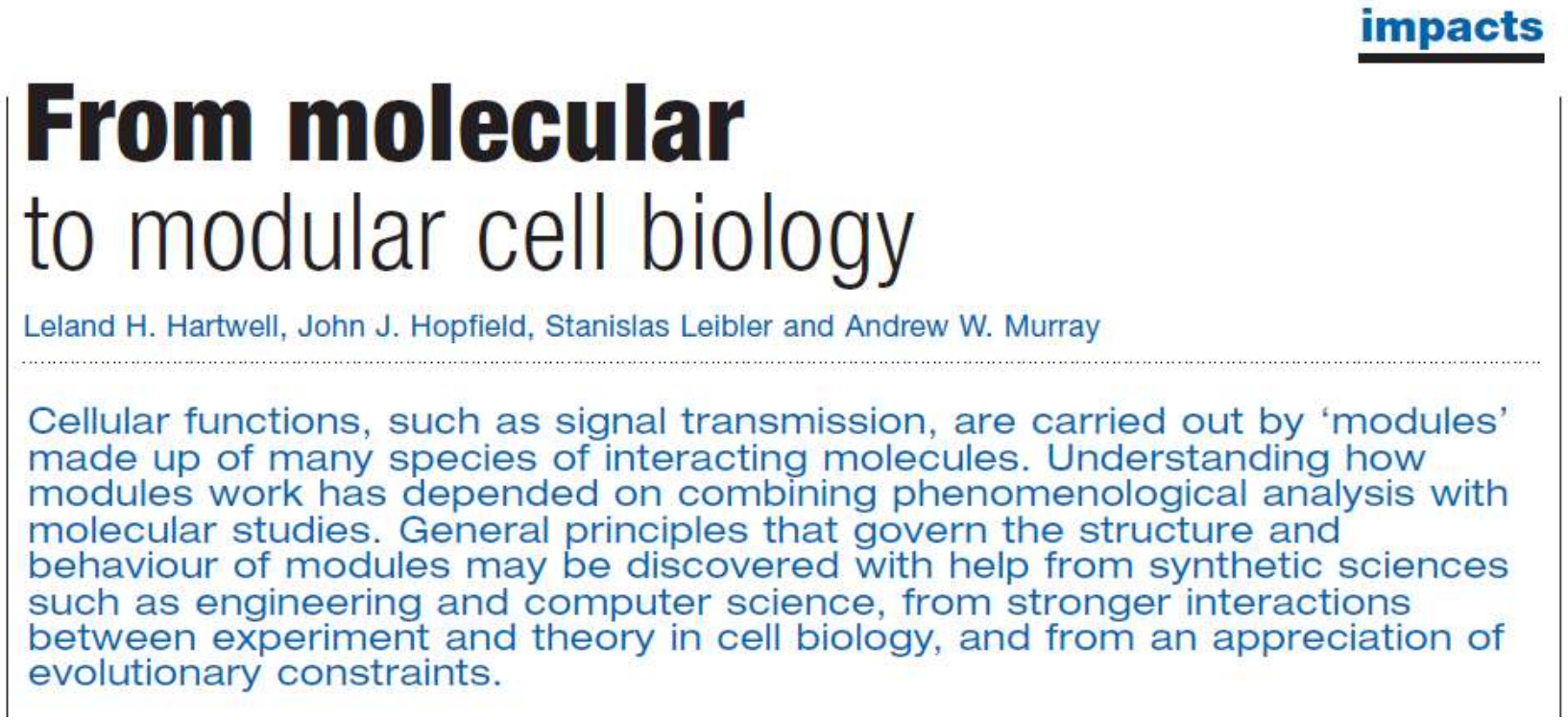
- Systems Biology deals with complexity by simplifying complex systems, summarizing them as components (nodes) and interactions (edges) between them.- Marc Vidal, Michael Cusick and Albert-Laszlo Barabasi, Cell 144, March, 2011;

- How?
  - By creating a cellular network and overlaying different High-Throughput data on it.

Sanford|Burnham
Medical Research Institute

SYSTEMS BIOLOGY RESEARCH: KEY PUBLICATIONS

- Hartwell et al (1999) article proposed study of "modules" instead of genes



impacts

From molecular
to modular cell biology

Leland H. Hartwell, John J. Hopfield, Stanislas Leibler and Andrew W. Murray

Cellular functions, such as signal transmission, are carried out by 'modules' made up of many species of interacting molecules. Understanding how modules work has depended on combining phenomenological analysis with molecular studies. General principles that govern the structure and behaviour of modules may be discovered with help from synthetic sciences such as engineering and computer science, from stronger interactions between experiment and theory in cell biology, and from an appreciation of evolutionary constraints.

SYSTEMS BIOLOGY RESEARCH: KEY PUBLICATIONS

- Kitano (2002) proposed research approaches for systems biology

REVIEW

Systems Biology: A Brief Overview

Hiroaki Kitano

To understand biology at the system level, we must examine the structure and dynamics of cellular and organismal function, rather than the characteristics of isolated parts of a cell or organism. Properties of systems, such as robustness, emerge as central issues, and understanding these properties may have an impact on the future of medicine. However, many breakthroughs in experimental devices, advanced software, and analytical methods are required before the achievements of systems biology can live up to their much-touted potential.

periments to identify specific interactions and conducting extensive literature surveys. Several attempts are under way to create a large-scale, comprehensive database on gene-regulatory and biochemical networks (*4*). Although such databases are useful sources of knowledge, many network structures remain to be identified. Substantial research has been done on expression profiling, in which clustering analy-

SYSTEMS BIOLOGY RESEARCH: KEY PUBLICATIONS

- Lee Hood et al (2004) proposed the use of systems biology in drug discovery and personalized medicine

VIEWPOINT

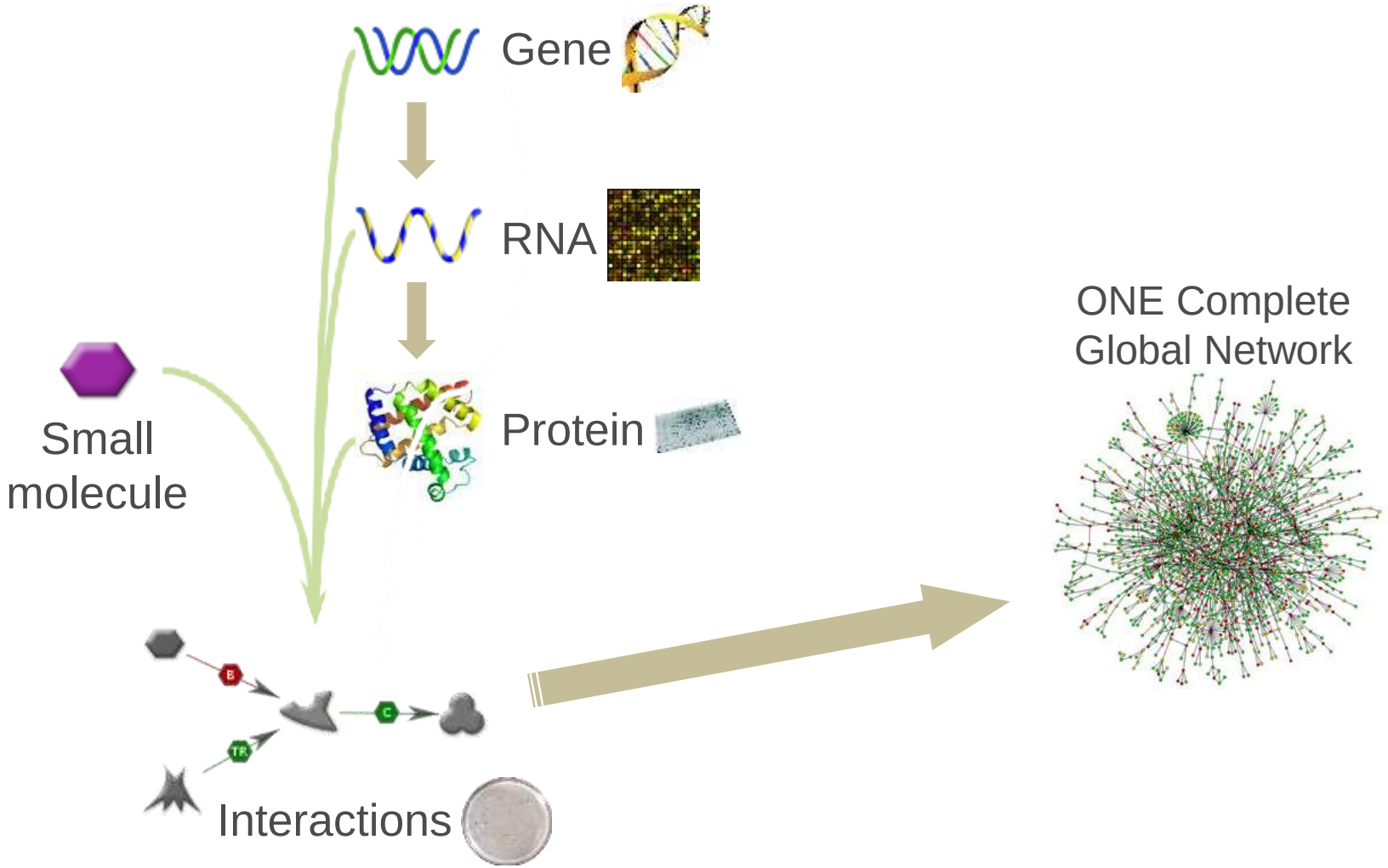
Systems Biology and New Technologies Enable Predictive and Preventative Medicine

Leroy Hood,[1]* James R. Heath,[2,3] Michael E. Phelps,[3] Biaoyang Lin[1]

Systems approaches to disease are grounded in the idea that disease-perturbed protein and gene regulatory networks differ from their normal counterparts; we have been pursuing the possibility that these differences may be reflected by multiparameter measurements of the blood. Such concepts are transforming current diagnostic and therapeutic approaches to medicine and, together with new technologies, will enable a predictive and preventive medicine that will lead to personalized medicine.

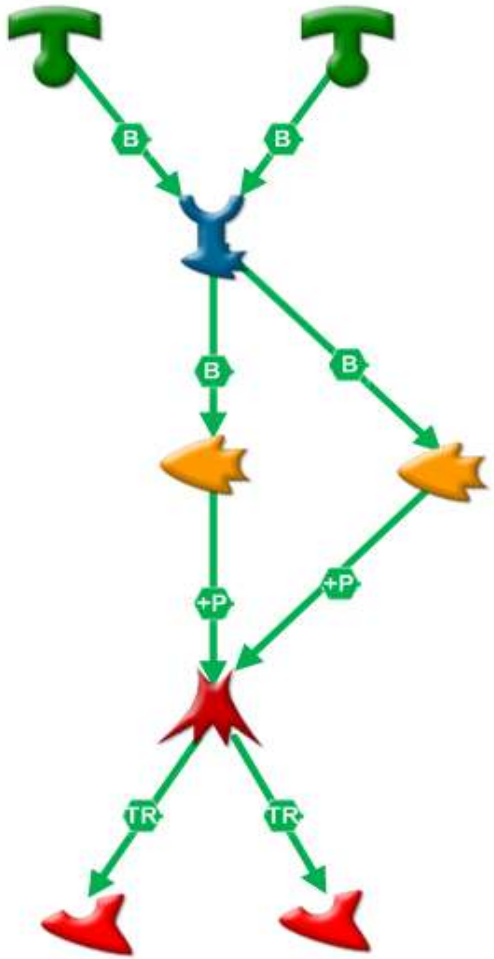
ic knockout strain had a distinct pattern of perturbed gene expression, with hundreds of mRNAs changing per knockout. About 15% of the perturbed mRNAs potentially encoded secreted proteins (*8*). If gene expression in diseased tissues also reveals patterns characteristic of pathologic, genetic, or envi-

SYSTEMS BIOLOGY DEALS WITH COMPLEXITY BY SIMPLIFYING - Marc Vidal



Sanford|Burnham Medical Research Institute

SIGNALING PATHWAYS IN NETWORK

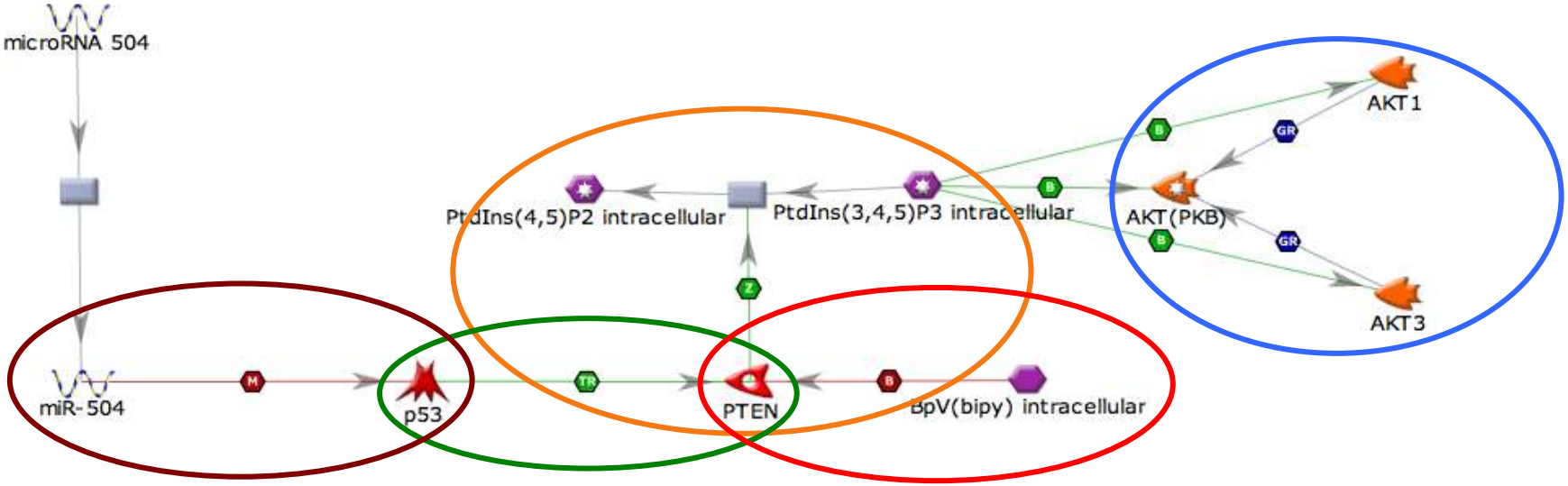


Sanford|Burnham
Medical Research Institute

What network properties are required for reconstruction of real biological pathways?

1. Rich semantics, controlled vocabulary:
2. Edges should be directed (otherwise we
3. Edges should have effect attribute to
4. Nodes should be specific (activation of

INTERACTIONS THAT CAN CONNECT ANY TYPE OF MOLECULAR ENTITIES



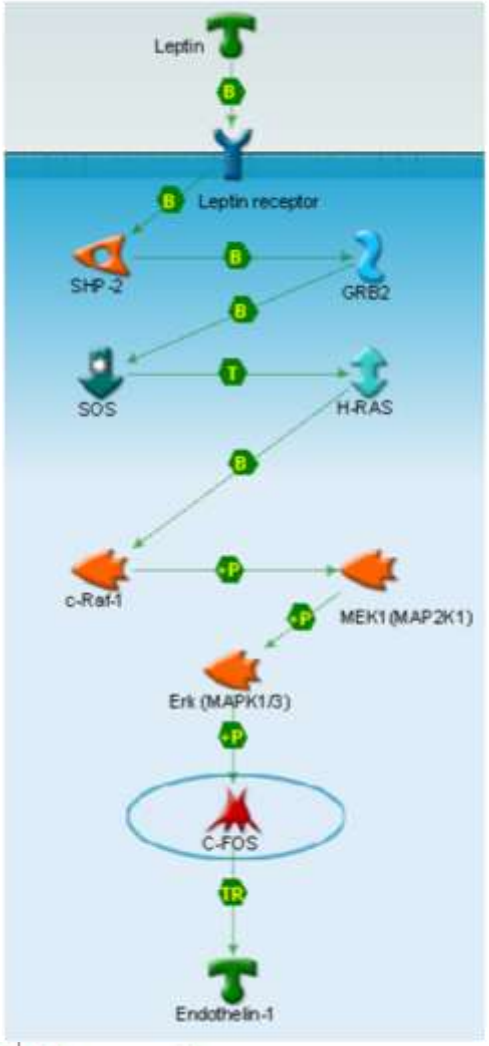
MetaCore global interaction network combines:

- Protein-protein interactions
- Compound-protein interactions, including drug-target
- Metabolic reactions (and transport)
- miRNA – protein interactions
- Complex-subunit and group-group member interactions (virtual)

Sanford|Burnham
Medical Research Institute

SIGNAL TRANSDUCTION PATHWAYS

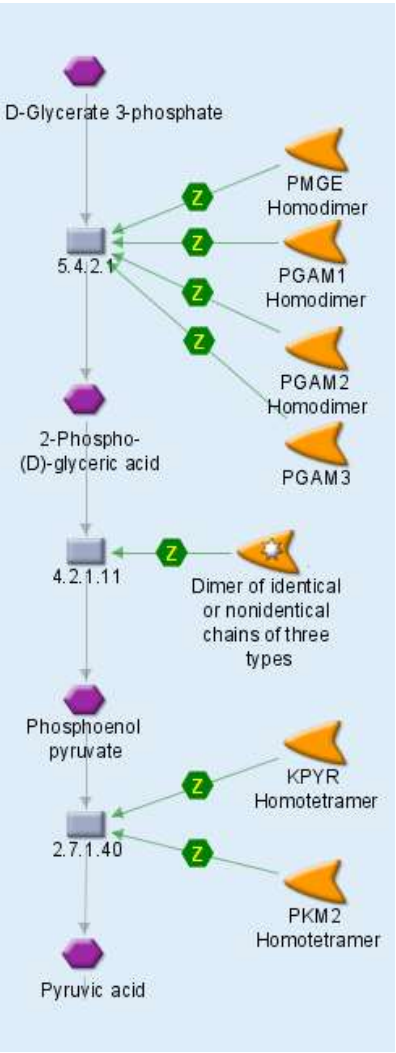
Ligand
↓
Receptor
↓
Adaptors
↓
Enzymes
↓
TF
↓
Target



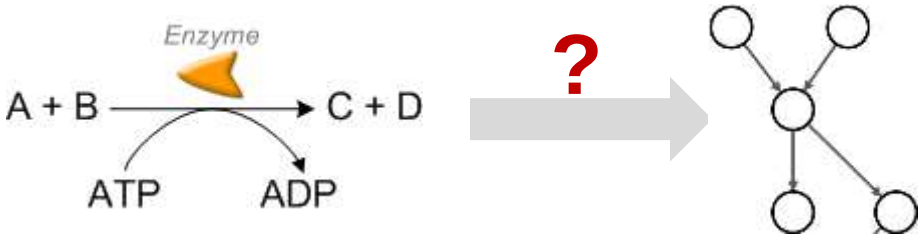
- Signal transduction is the intracellular transfer of information from external stimuli into nucleus through a signal pathway.
- Canonical signaling pathways are thoroughly described in literature.
- Pathway: Information transfer is **directed**
- Pathways interact with each other and with metabolic pathways in a complex manner, forming the global biological network

Sanford | Burnham
Medical Research Institute

METABOLIC PATHWAYS

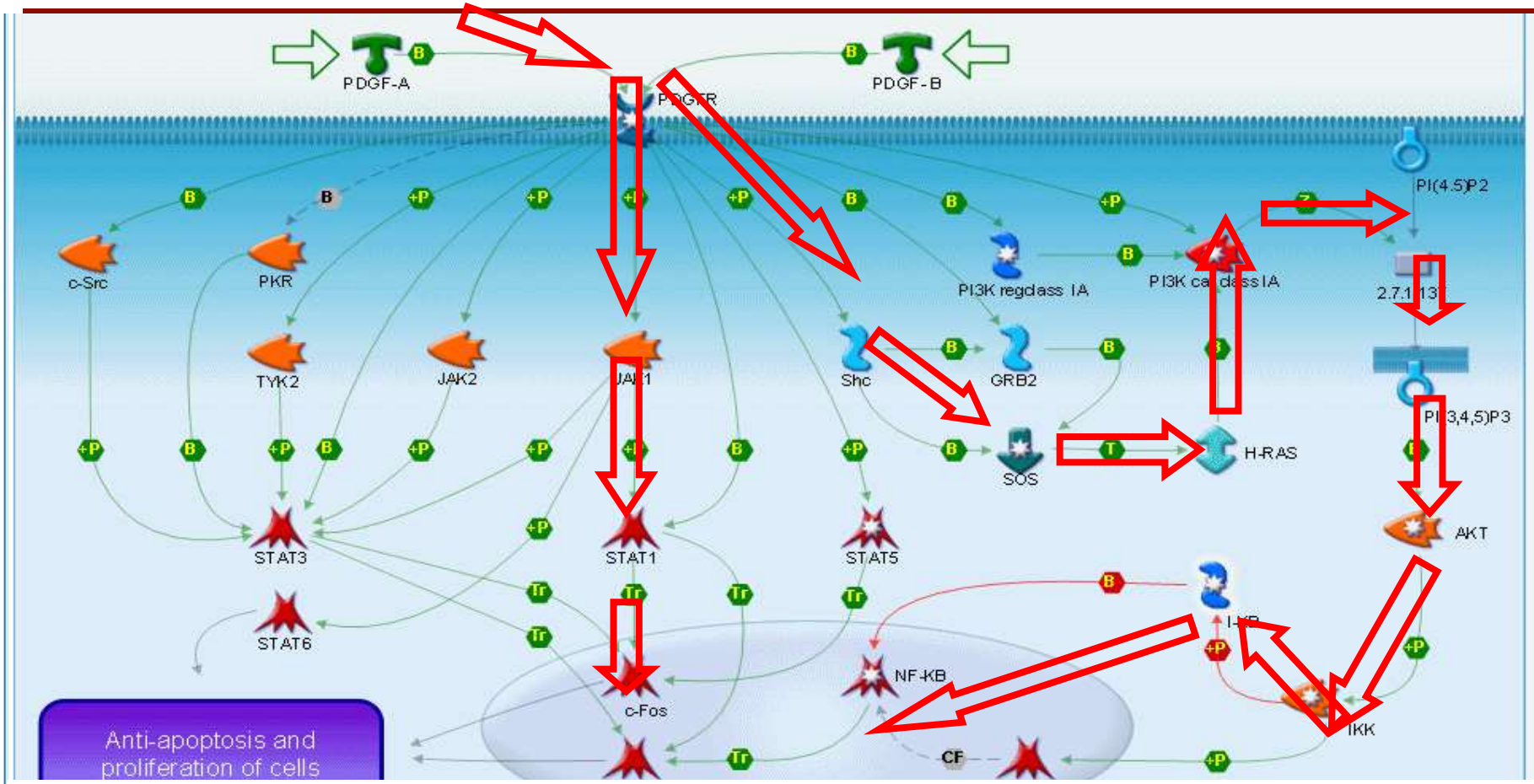


- Metabolic pathways are in an equilibrium state at any moment. They lack directedness;
- Usually metabolism gets depicted as a separate network without any relation to regulatory events;
- It is difficult to integrate metabolism into a binary network
- Metabolic pathways, however, are interconnected and integrated with signaling;



Sanford | Burnham
Medical Research Institute

ASSEMBLIES OF CANONICAL PATHWAYS ON MAPS

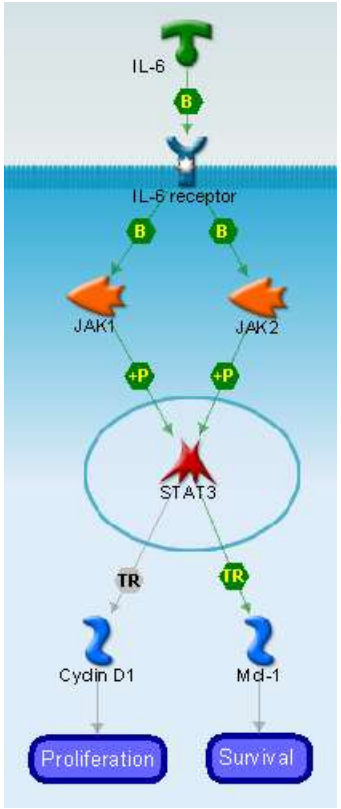


Canonical Pathway Maps may contain many individual pathways with the same theme organized based on most consensus literature data

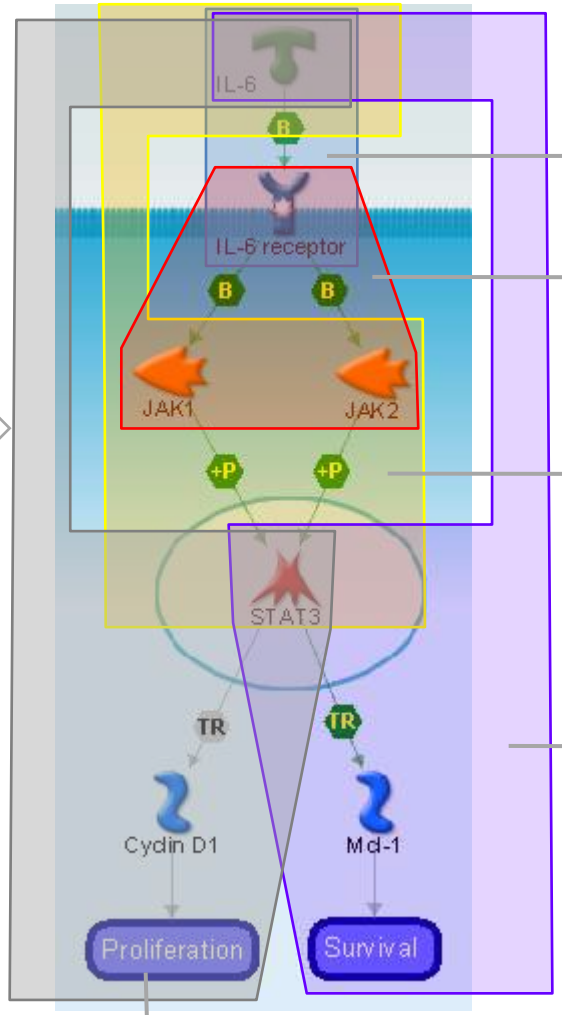
Sanford|Burnham
Medical Research Institute

BEHID EACH MAP: Compilation of different information types from different articles into one signaling pathway

IL-6 signaling pathway



Pathway is compiled of different data from number of articles



Pubmed 1: IL-6 activates IL-6 receptor

Pubmed 2: IL-6 receptor activates JAK1, JAK2

Pubmed 3: IL-6 activates JAK1, JAK2 which activate STAT3

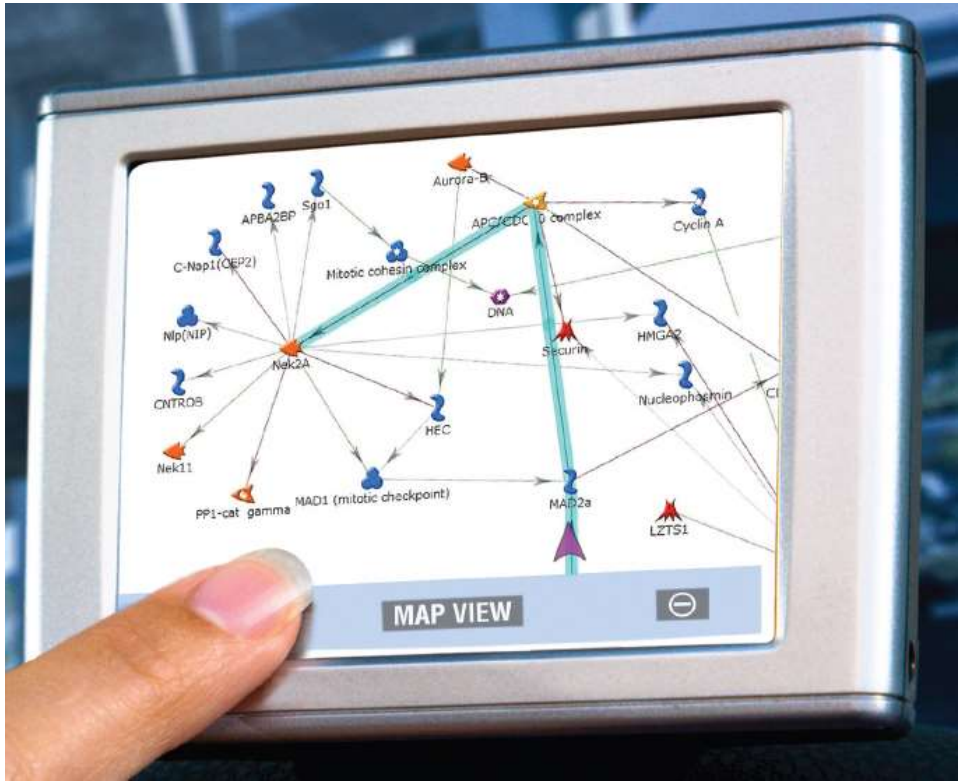
Pubmed 4: IL-6 through STAT3 activates expression of Mcl-1 leading to survival

Pubmed 5: IL-6 through STAT3 activates expression of Cyclin D1 leading to cell proliferation

Sanford|Burnham
Medical Research Institute

CAN WE USE IT AS A GPS IN PATHWAY ANALYSIS?

– Use the network building options to drive along your points of interest



Sanford|Burnham
Medical Research Institute

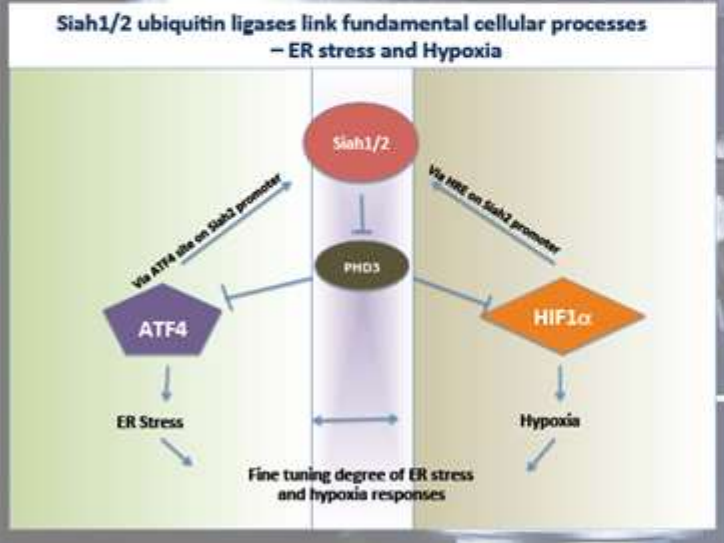
TEST HYPOTHESES AND EXPLORE THE DATA

From: paperoftheweek
To: ⊞ Labs (La Jolla); ⊞ Labs (Lake Nona)
Cc:
Subject: Paper of the Week
Sent: Mon 5/12/2014 9:30 AM

Click here to view paper online

Sanford-Burnham

Paper of the Week



PLOS | GENETICS

Fine tuning of the UPR by the ubiquitin ligases Siah1/2

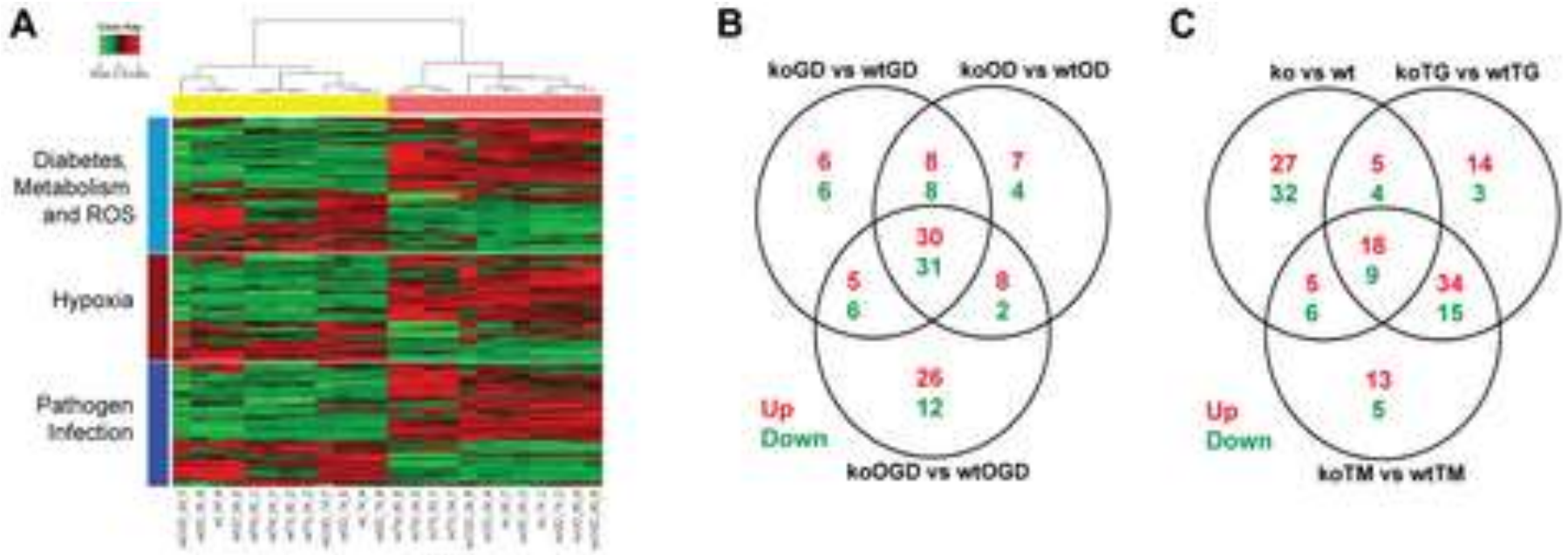
Ze'ev Ronai Lab
In collaboration with Randal Kaufman's lab

- Transcription of the ubiquitin ligases Siah1a/2 induced by PERK/ATF4 and Ire1/sXPB1, two key UPR sensors
- Induced Siah1a/2 augments ATF4 availability, constituting a feed-forward mechanism amplifying the UPR output
- Siah1/2 activation by- and contribution to- UPR occurs following severe ER stress – (i.e. ischemia) required for cell commitment to undergo apoptosis
- Siah1a+/–::Siah2–/– mice are partially protected against cerebral ischemia as evidenced by smaller infarct size and protection from neuronal death
- Siah1/2 constitutes an obligatory fine-tuning mechanism that predisposes cells to death under severe ER stress conditions

Send your newly accepted paper to paperoftheweek@sanfordburnham.org

Published online May 8, 2014

Figure 4. Siah1/2-dependent gene expression analysis confirms an ER stress signature.

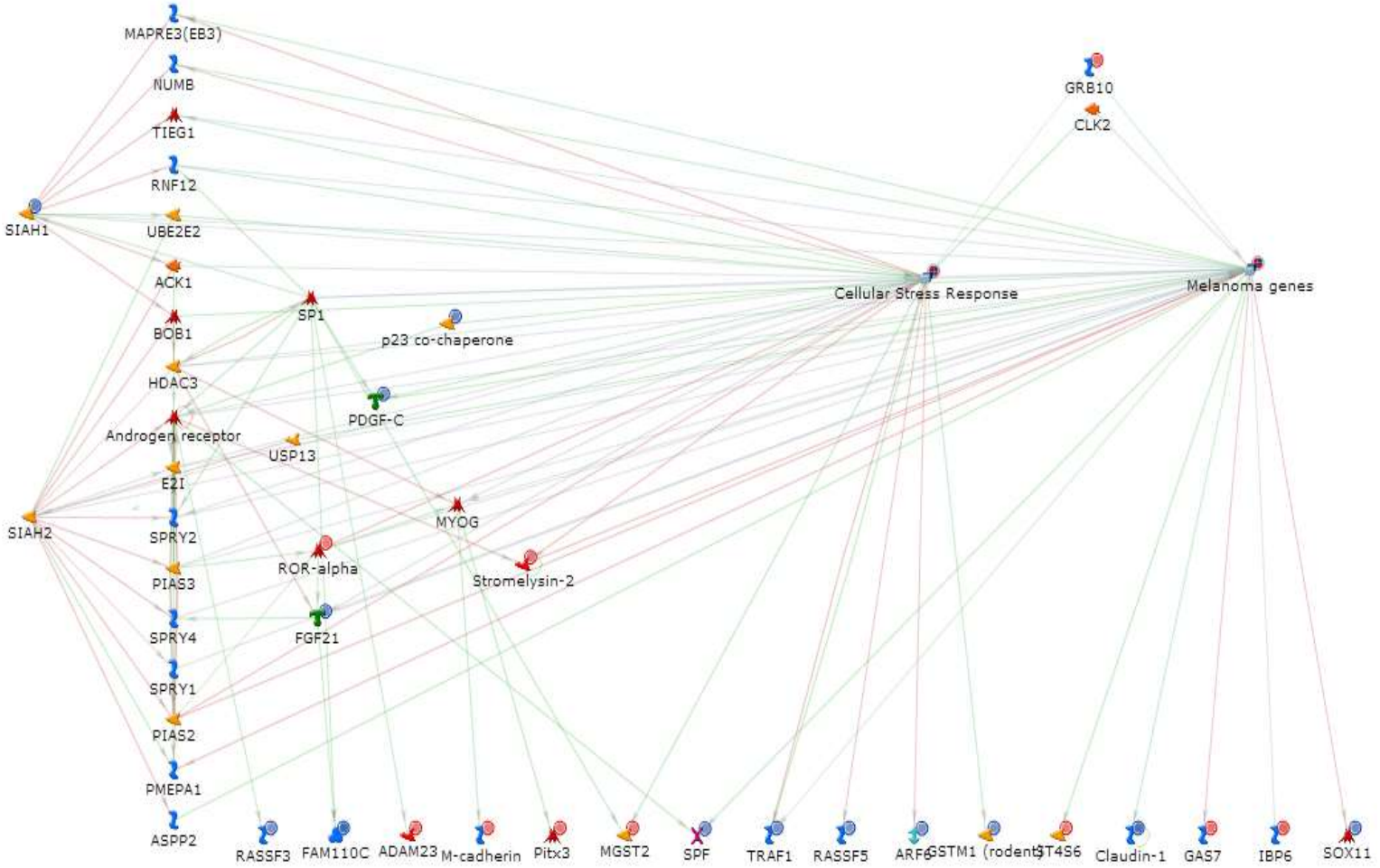


Scortegagna M, Kim H, Li J-L, Yao H, et al. (2014) Fine Tuning of the UPR by the Ubiquitin Ligases Siah1/2. PLoS Genet 10(5): e1004348. doi:10.1371/journal.pgen.1004348
http://www.plosgenetics.org/article/info:doi/10.1371/journal.pgen.1004348

PLOS | GENETICS

Sanford | Burnham
Medical Research Institute

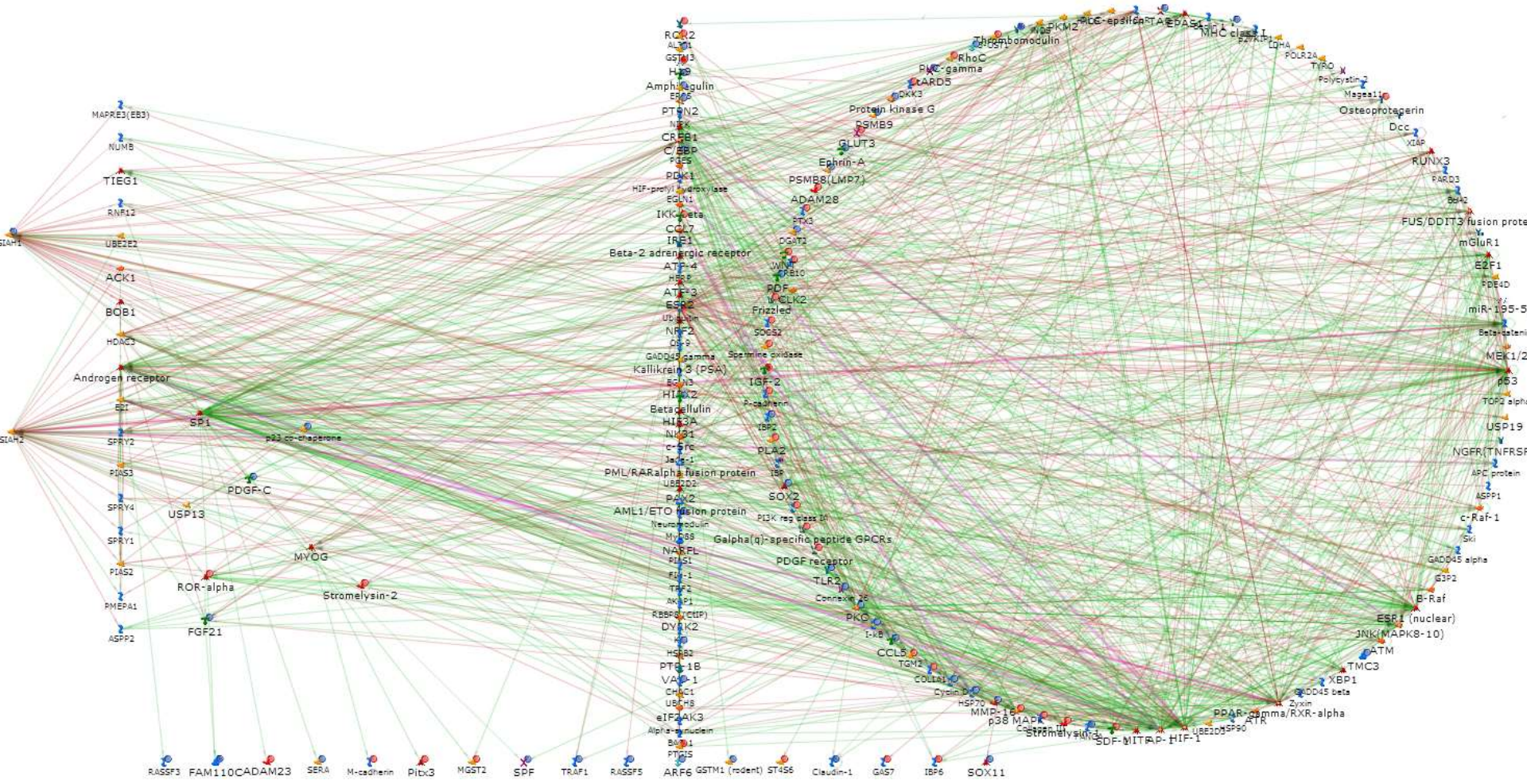
Connecting SIAH1 and SIAH2 to Stress Response and Melanoma



Genes with blue or red circles tagged onto them are significant based on the data from the PloS paper tables 1-4

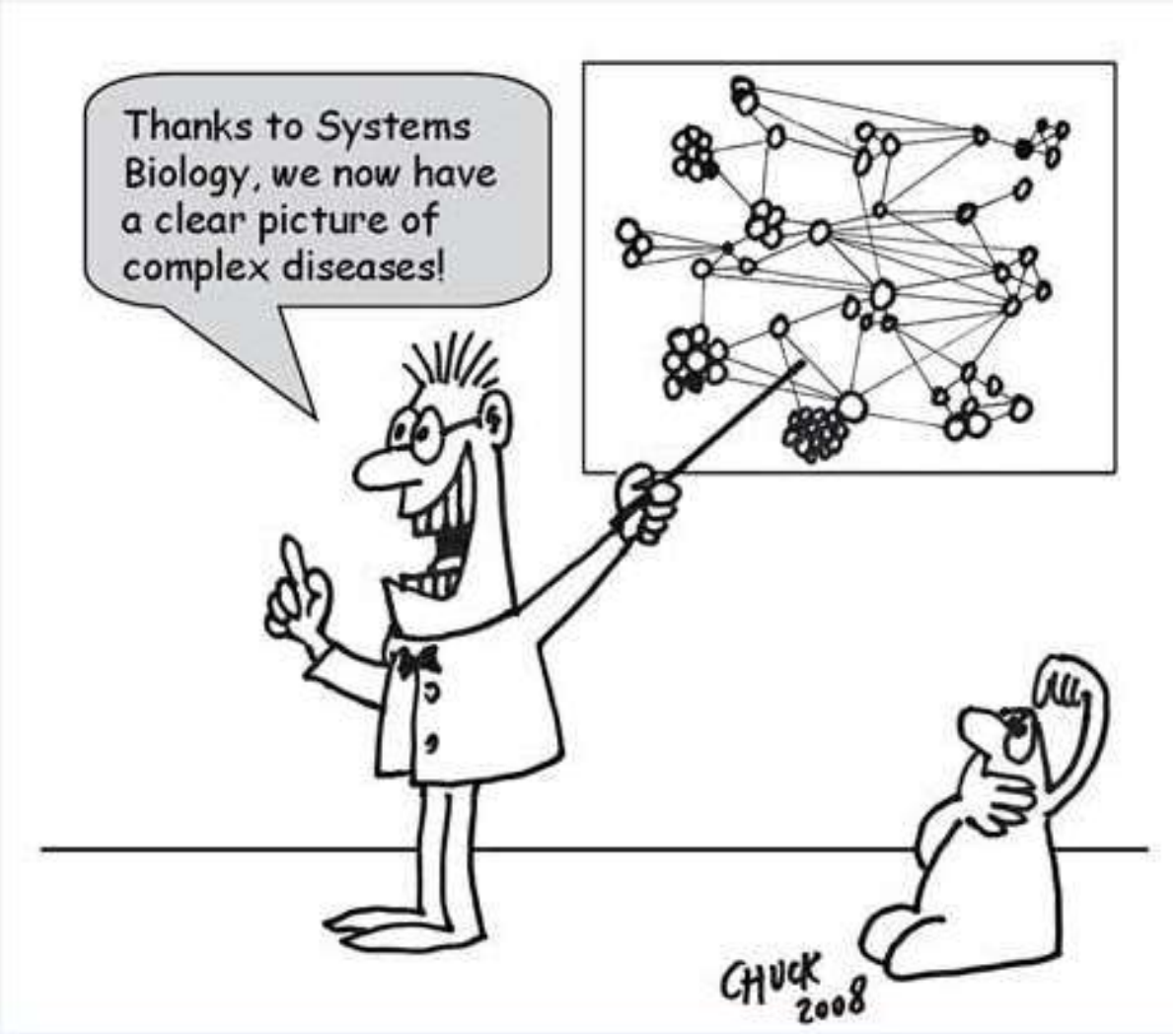
Sanford | Burnham
Medical Research Institute

The HYPOTHESIS BLUEPRINT: Expanding the groups reveals which stress response genes have known connections with SIAH ½ and Melanoma genes



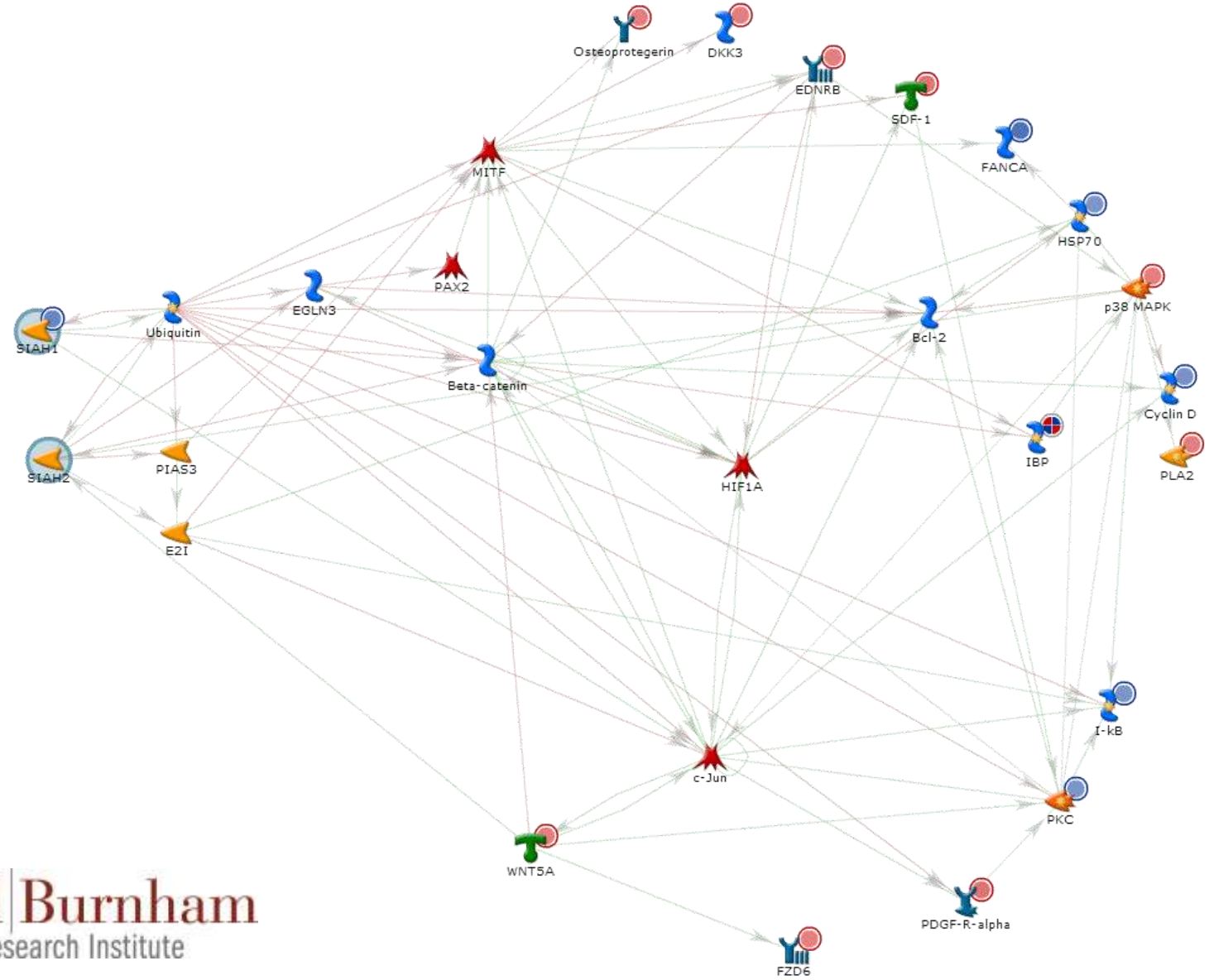
Sanford | Burnham
Medical Research Institute

IF ONLY IT WERE THAT EASY



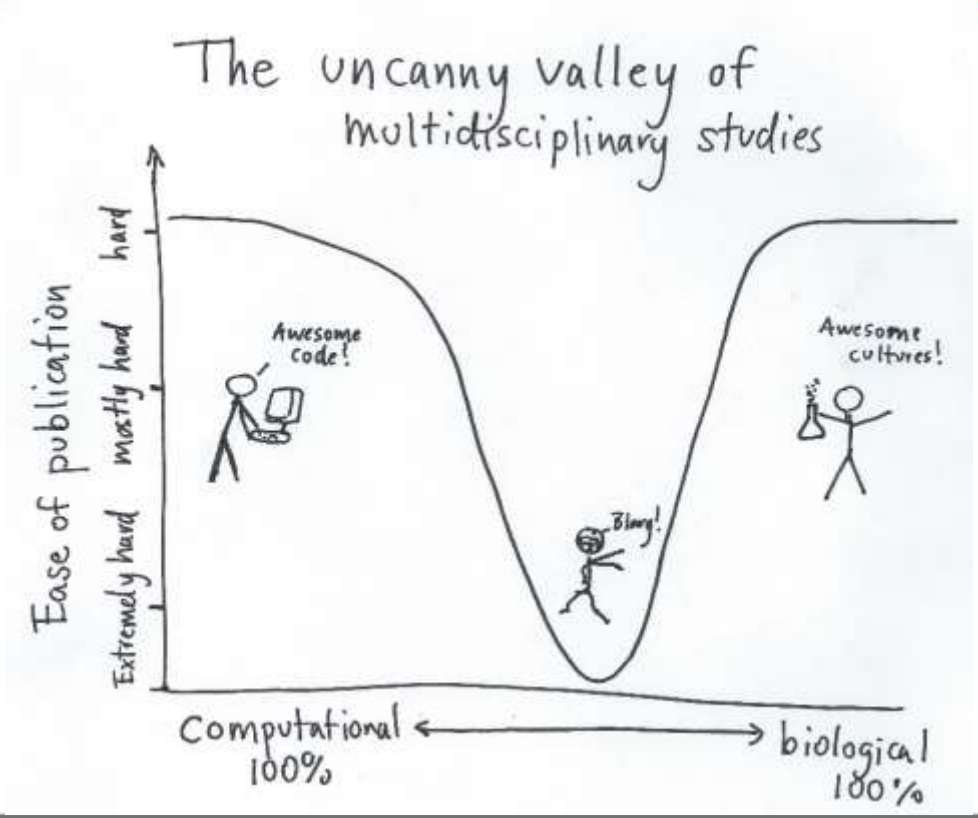
Sanford|Burnham
Medical Research Institute

EXAMPLE OF DRILLING DOWN TO MECHANISMS THAT LEAD TO SIGNIFICANT GENES relevant to MITF or HIF (important in melanoma):



Sanford|Burnham
Medical Research Institute

EXTRACTING TESTABLE BIOLOGICAL MECHANISMS FROM HIGH-THROUGHPUT DATA IS NOT TRIVIAL, BUT IT IS NOT POSSIBLE WITHOUT SYSTEMS BIOLOGY



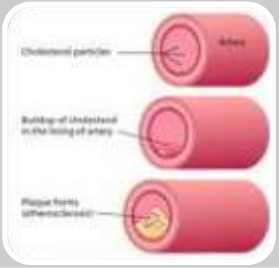
Sanford|Burnham
Medical Research Institute
From Research, The Power To Cure

EXAMPLES OF MULTI-OMIC INTEGRATED ANALYSES

TNO – ATHEROSCLEROSIS DRUG PROJECT

Start with Data

- Gene expression and metabolomic data
- Mice treated with different cholesterol lowering drugs
- On different cholesterol diets prior to treatment
- Analyzed in MetaCore



Generate hypothesis

- Atherogenic processes not only driven by changes in metabolism but also by deregulation of inflammatory and extracellular signaling pathways
- Experiment
  - Look how cholesterol plaques form, how big and how they are affected by different drugs or by diet alone-high or low cholesterol

Conclusion

- 2 different branches of atherogenesis need to be addressed for drugs to work
- Only those drugs that effect metabolism, inflammation and adhesion will show most promise

INTEGRATIVE ANALYSIS OF METABOLOMICS AND SIGNALING (TNO PROJECT)

Kleemann *et al. BMC Systems Biology* 2011, **5**:125
http://www.biomedcentral.com/1752-0509/5/125

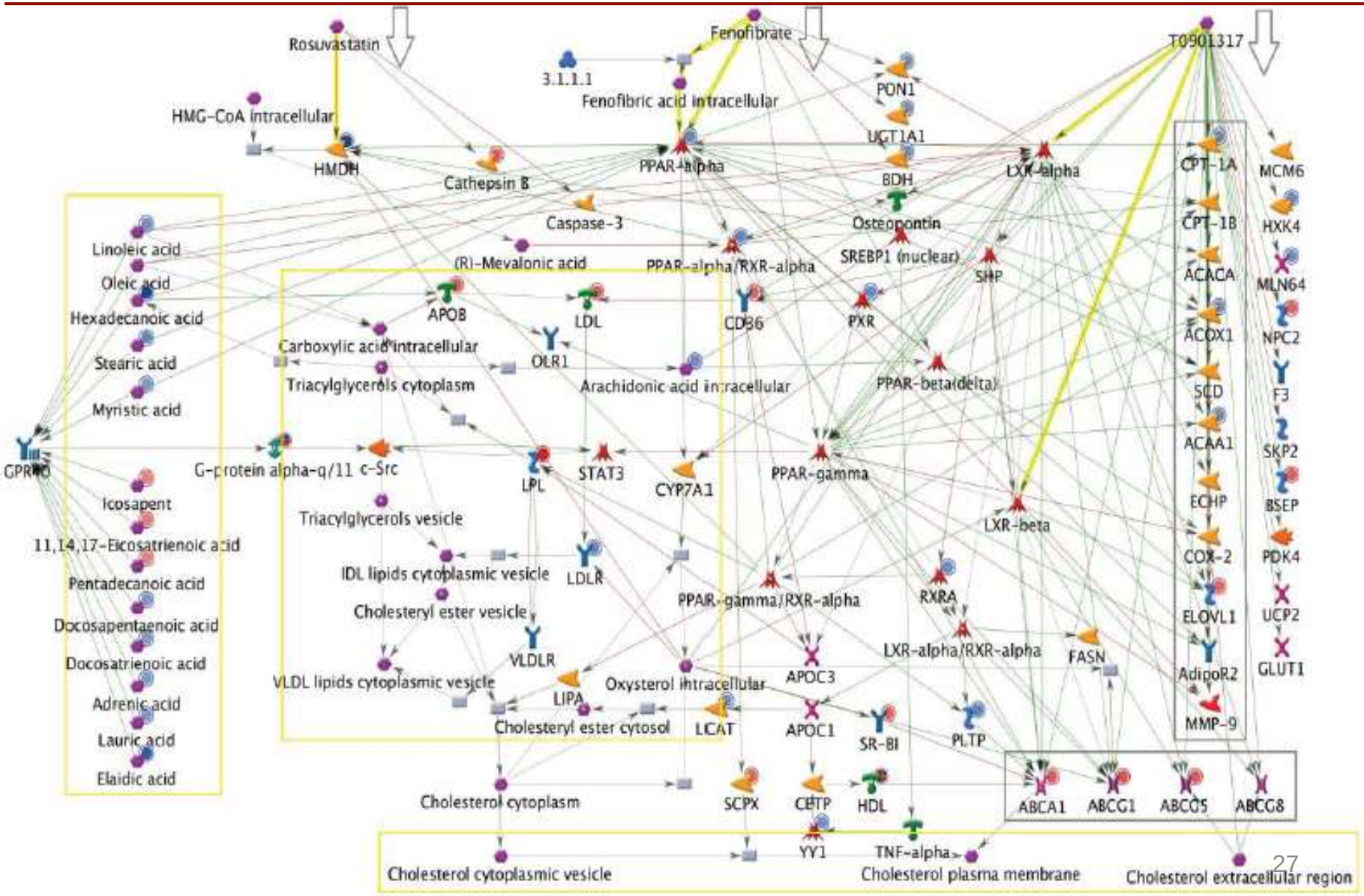
RESEARCH ARTICLE Open Access

A Systems Biology Strategy for Predicting Similarities and Differences of Drug Effects: Evidence for Drug-specific Modulation of Inflammation in Atherosclerosis

Robert Kleemann[1*], Svetlana Bureeva[3], Ally Perlina[3], Jim Kaput[4,5], Lars Verschuren[1,2], Peter Y Wielinga[1], Eva Hurt-Camejo[6], Yuri Nikolsky[3], Ben van Ommen[2] and Teake Kooistra[1]

Sanford|Burnham
Medical Research Institute

RESULTS OF CONCURRENT ANALYSIS OF METABOLOMICS AND SIGNALING - TNO



CYSTIC FIBROSIS FOUNDATION

- Proteomic Data was taken from Bill Balch's publication

Cell

Hsp90 Cochaperone Aha1 Downregulation Rescues Misfolding of CFTR in Cystic Fibrosis

Xiaodong Wang,[1,6,7] John Venable,[1,6] Paul LaPointe,[1,6] Darren M. Hutt,[1,6] Atanas V. Koulov,[1] Judith Coppinger,[1] Cemal Gurkan,[1] Wendy Kellner,[1] Jeanne Matteson,[1] Helen Plutner,[1] John R. Riordan,[5] Jeffery W. Kelly,[2,3] John R. Yates, III,[1,*] and William E. Balch[1,4,*]

Sanford|Burnham
Medical Research Institute

Table from the Cell Publication

Table 1. The CFTR ER-Associated Folding Proteome

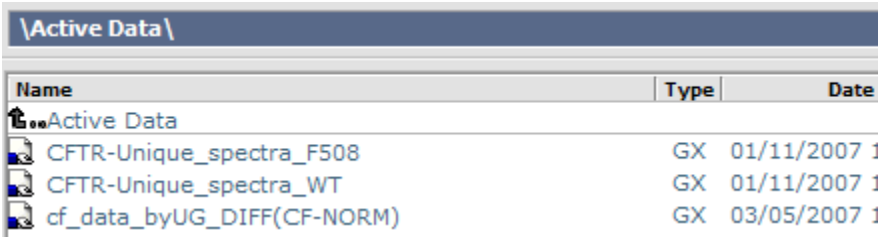
		ΔF508 CFTR			WT CFTR		
RefSeq AC	Protein Name	Sequence Coverage (%)	Unique Spectra	Total Spectra	Sequence Coverage (%)	Unique Spectra	Total Spectra
NM_000492	CFTR	57	229	2481	60	333	4172
NM_024351	Hsc70	60	66	369	48	39	132
NM_022310	GRP78[a]	40	30	85	33	17	37
NM_021979	Hsp70-2	16	23	57	8	7	17
NM_001746	calnexin[a]	16	13	52	10	4	7
NM_008302	Hsp90β	36	24	49	21	11	18
NM_010481	GRP75	34	23	40			
NM_005348	Hsp90α	39	24	37	6	4	5
NM_022934	DnaJ-like protein	34	11	35			
NM_001539	Hsp40-A1 (Hdj2)	31	10	33	40	13	25
NM_005345	Hsp70-1A	19	12	20	8	4	6
NM_004282	BAG-2	28	7	20	10	2	2
NM_005880	Hsp40-A2 (Hdj3)	32	9	19	28	6	16
NM_002155	Hsp70B′	10	8	14	3	3	5
NM_013559	Hsp105	10	5	6			
NM_013686	TCP1	10	3	5	9	3	5
NM_010223	FKBP8	17	4	5			
NM_013863	BAG-3	12	4	5			
NM_016737	Hop	11	4	4			
NM_016742	Cdc37	8	2	3			
NM_000942	cyclophilin B[a]	13	2	2			
NM_006601	p23	16	2	2			
NM_012111	Aha1	15	7	15	34	10	20
NM_009037	reticulocalbin[a]	27	5	8	50	14	19
NM_011992	reticulocalbin 2[a]	9	3	4	22	6	12
NM_001219	calumenin[a]				7	3	3

Indicated are the interacting proteins in BHK cells, their percentage sequence coverage, number of unique spectra, and number of total spectra as detected by mass spectrometry in cell lines examined (Figure 1).
[a] ER lumenal chaperones.

Sanford | Burnham
Medical Research Institute

CF Expression Data

- Gene Expression was provided by CFF members
- This example represents differential gene expression between 4 CF patients' arrays and 4 normals



- BOTH TYPES OF FILES CAN BE SIMULTANEOUSLY USED FOR ENRICHMENT ANALYSIS AND NETWORK CONSTRUCTION, AND WILL BE VISUALIZED CONCURRENTLY (top 2 = proteomic, 3rd file = gene expression):

Sanford|Burnham
Medical Research Institute

Enrichment by Maps:

Statistically significant maps

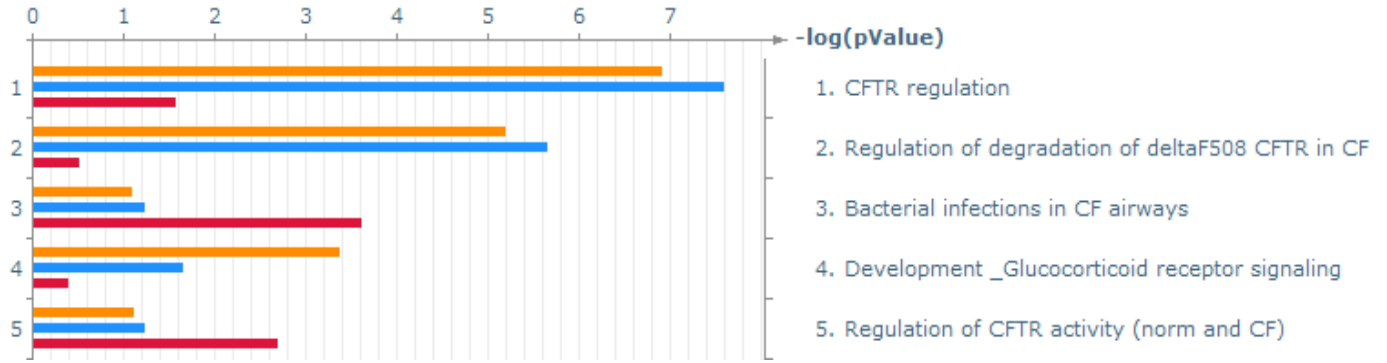
Experiments

#	Experiment	Network objects
1	(1) CFTR-Unique_spectra_F508	25
2	(2) CFTR-Unique_spectra_WT	18
3	(3) cf_data_byUG_DIFF(CF-NORM)	3068

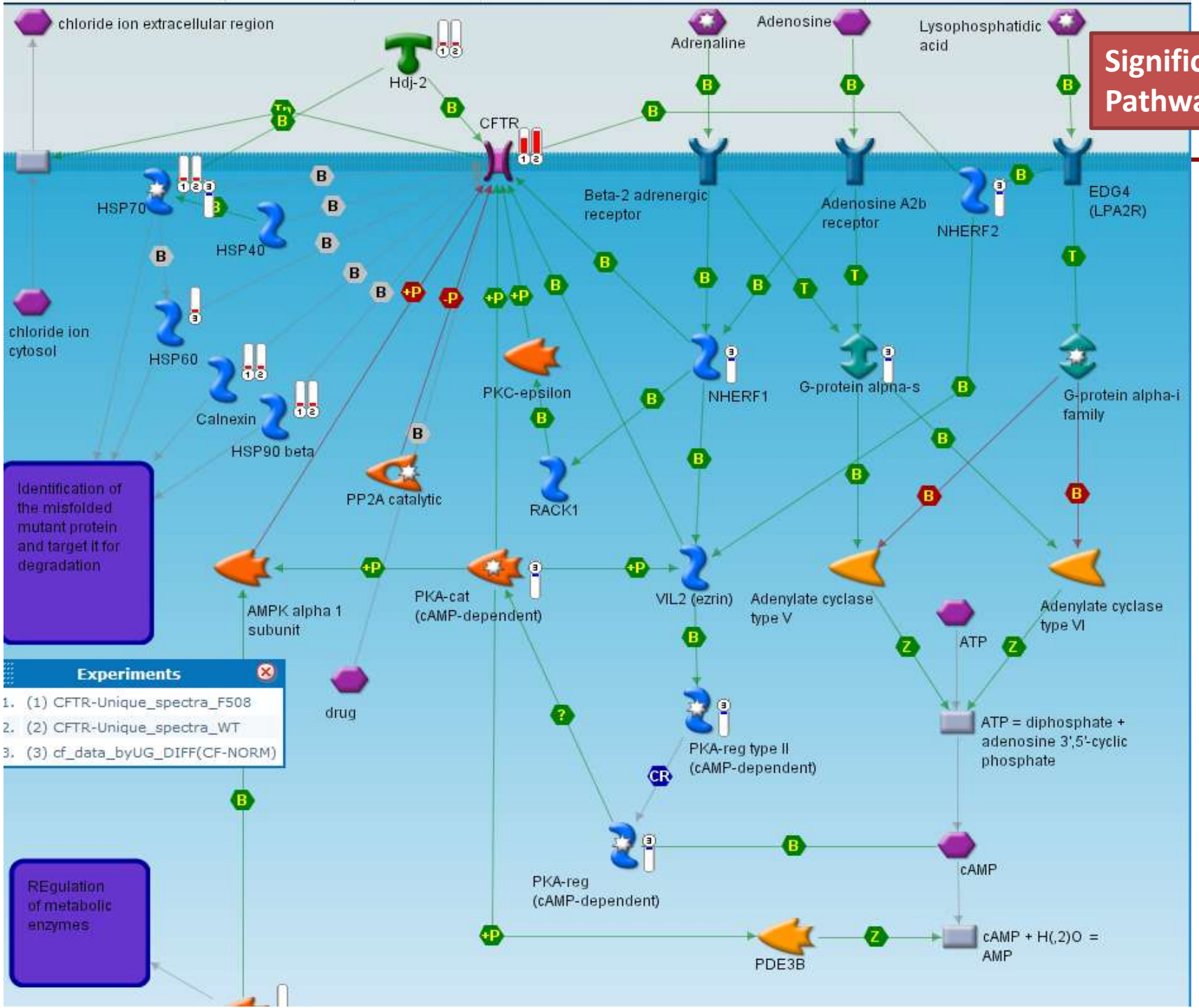
Maps

Left click on the bar shows map.

Expand | Collapse | Export to image

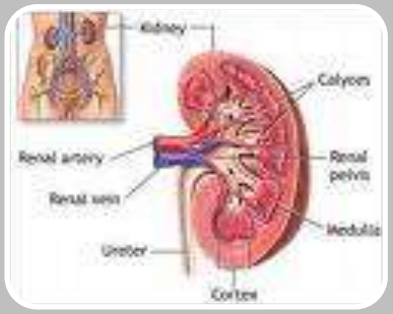


Sanford|Burnham Medical Research Institute



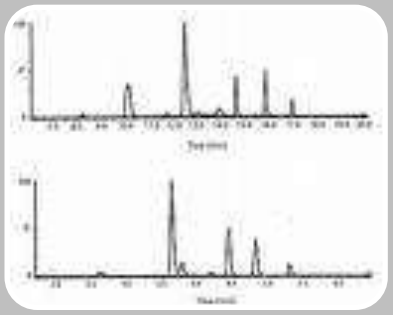
Significant Pathway Map

Merck Project (Published)



Hypothesis

- Regular metabolic intermediates measured in urine can serve as biomarkers of kidney toxicity



Experimental design

- Metabolomic profiles collected from different groups of mice treated with various kidney toxins were compared and analyzed in context of pathways
- As a result hypothesis became more apparent that there is a relationship between key metabolites and certain transporters and transcription factors. To test this, expression profiles were also applied and showed correlations with metabolomic data

Conclusion and new finding

- This concordance between kidney expression and urine metabolism corroborated the hypothesis of relationship between proximal kidney tubule injury and differential effects on metabolic and signaling pathways

Merck KIDNEY TOX PATHWAY ANALYSIS – Clin Res in Toxicology

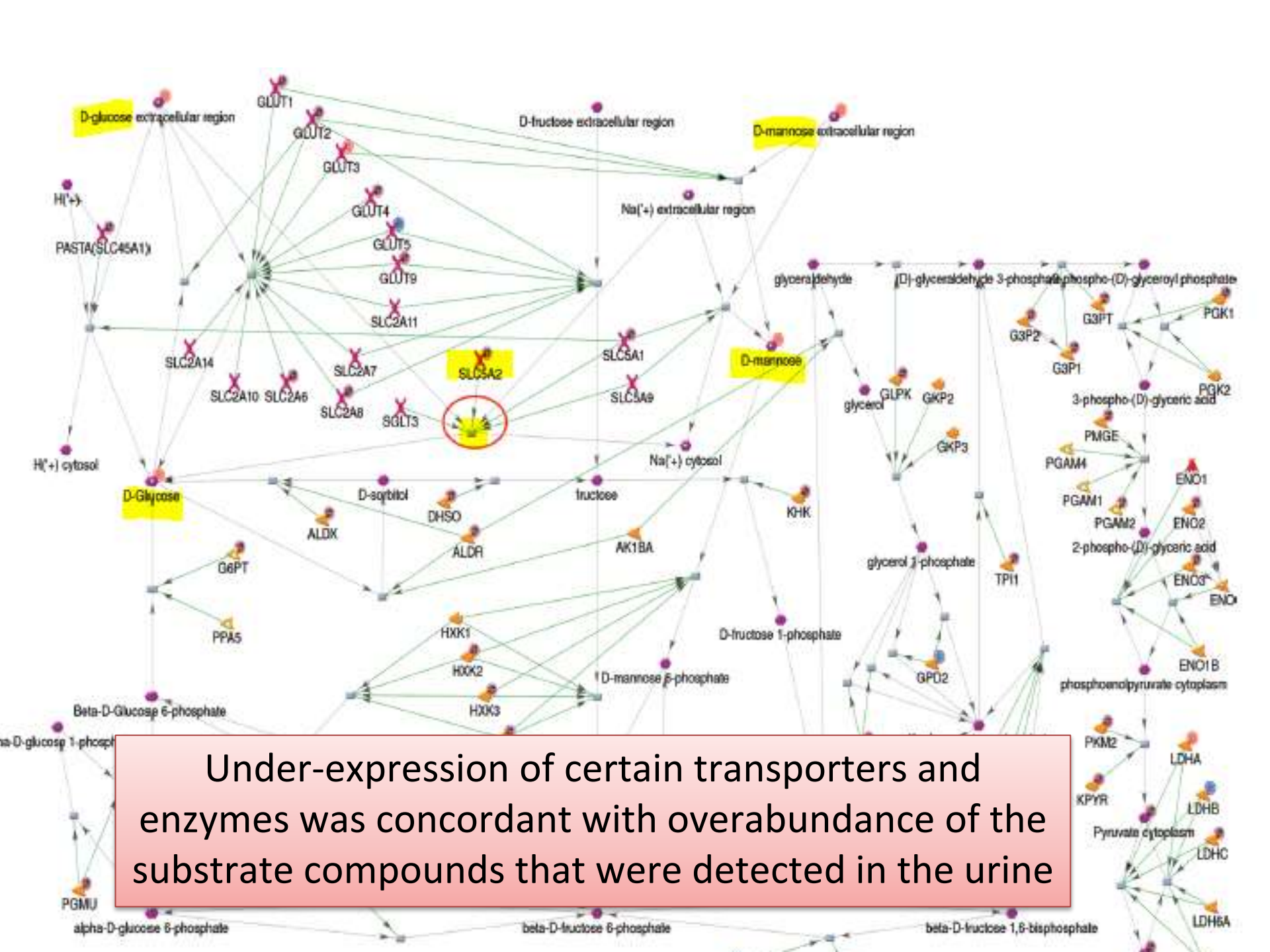
Integrated Pathway Analysis of Rat Urine Metabolic Profiles and Kidney Transcriptomic Profiles To Elucidate the Systems Toxicology of Model Nephrotoxicants

Ethan Yixun Xu,*,† Ally Perlina,‡ Heather Vu,† Sean P. Troth,† Richard J. Brennan,‡ Amy G. Aslamkhan,† and Qiuwei Xu*,†

Department of Safety Assessment, Merck Research Laboratories, West Point, Pennsylvania 19486, and GeneGo Inc., St. Joseph, Michigan 49085

Received February 15, 2008

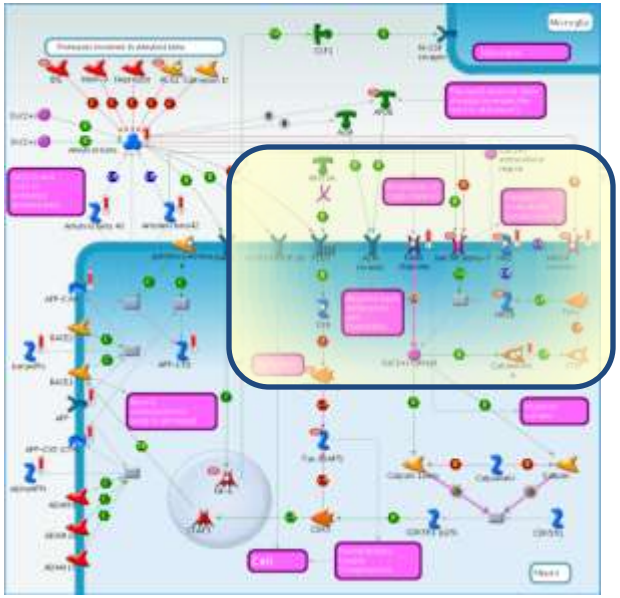
In this study, approximately 40 endogenous metabolites were identified and quantified by ^{1}H NMR in urine samples from male rats dosed with two proximal tubule toxicants, cisplatin and gentamicin. The excreted amount of a majority of those metabolites in urine was found to be dose-dependent and exhibited a strong correlation with histopathology scores of overall proximal tubule damage. MetaCore pathway analysis software (GeneGo Inc.) was employed to identify nephrotoxicant-associated biochemical changes via an integrated quantitative analysis of both urine metabolomic and kidney transcriptomic profiles. Correlation analysis was applied to establish quantitative linkages between pairs of individual metabolite and gene transcript profiles in both cisplatin and gentamicin studies. This analysis revealed that cisplatin and gentamicin treatments were strongly linked to declines in mRNA transcripts for several luminal membrane transporters that handle each of the respective elevated urinary metabolites, such as glucose, amino acids, and monocarboxylic acids. The integrated pathway analysis performed on these studies indicates that cisplatin- or gentamicin-induced renal Fanconi-like syndromes manifested by glucosuria, hyperaminoaciduria, lactic aciduria, and ketonuria might be better explained by the reduction of functional proximal tubule transporters rather than by the perturbation of metabolic pathways inside kidney cells. Furthermore, this analysis suggests that renal transcription factors HNF1α, HNF1β, and HIF-1 might be the central mediators of drug-induced kidney injury and adaptive response pathways.



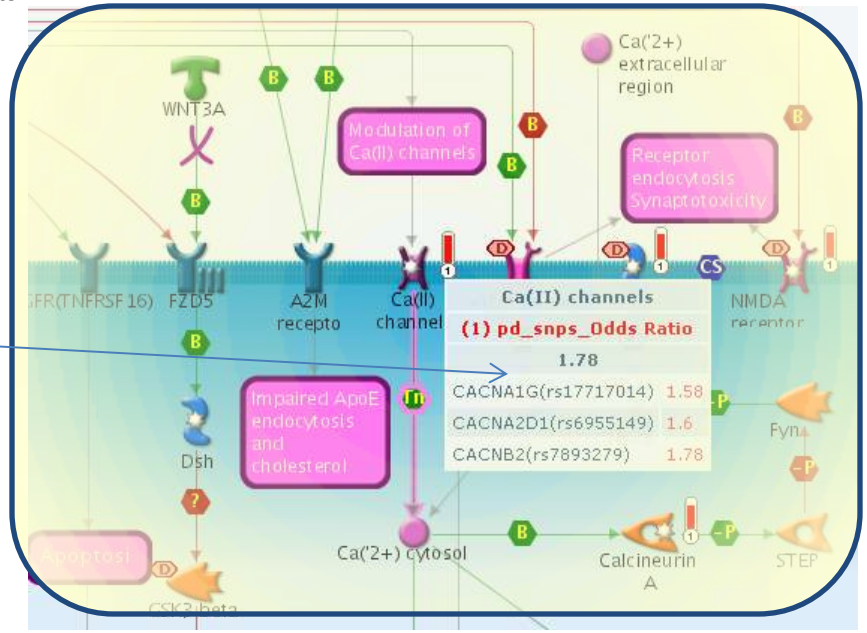
Under-expression of certain transporters and enzymes was concordant with overabundance of the substrate compounds that were detected in the urine

Sample SNP analysis (AD Case Study)

Canonical Pathway Maps enriched with SNP data from Alzheimer's disease patients

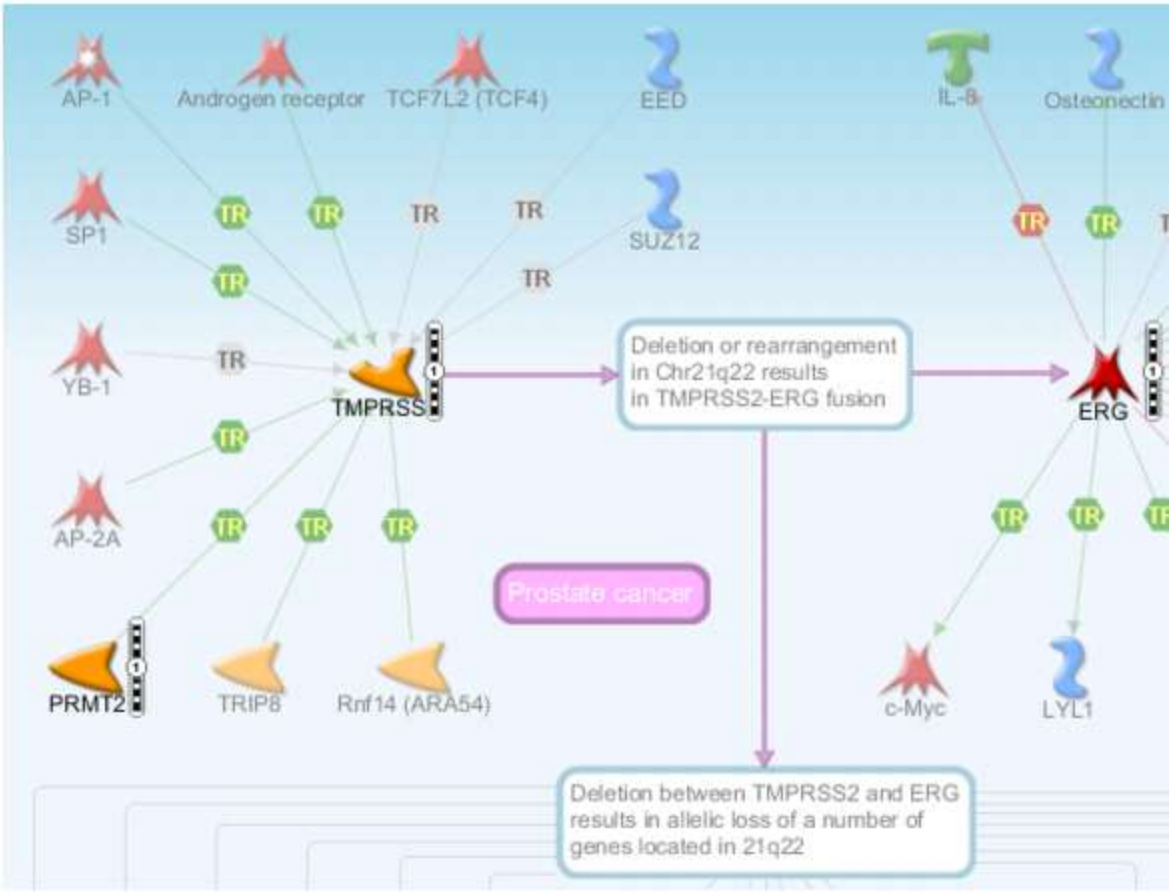


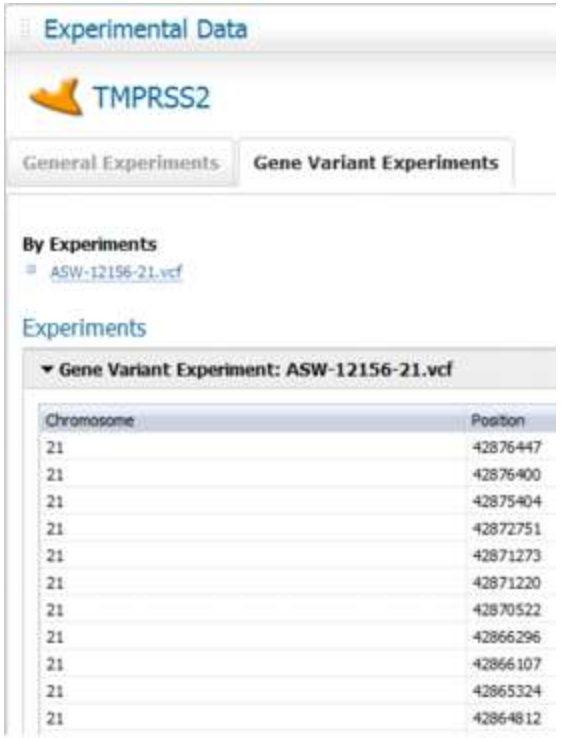
Overlaid SNP data zoomed in with rsIDs for the experiment



The partial canonical pathway map above highlights the genes the significant SNPs are linked to (red thermometers); t
his map represents Alzheimer's disease: extracellular Amyloid beta toxicity. Also evident on the map in figure 4 are the know Parkinson's disease biomarkers which are curated by GeneGo, these proteins have red hexagons next to them marked with a "D" The only known Parkinson's biomarker in this map that has a significant SNP associated with it is GRIN2A and this protein is a subunit of NR2 (NR2 is a group of protein, which may belong to NMDA receptor. Active NMDA receptors are composed of at least one NR1subunit, the obligatory subunit for channel activity, and one or more than one type of NR2 (NR2A, NR2B, NR2C, NR2D) subunits which confer variability in the properties of the receptors)..

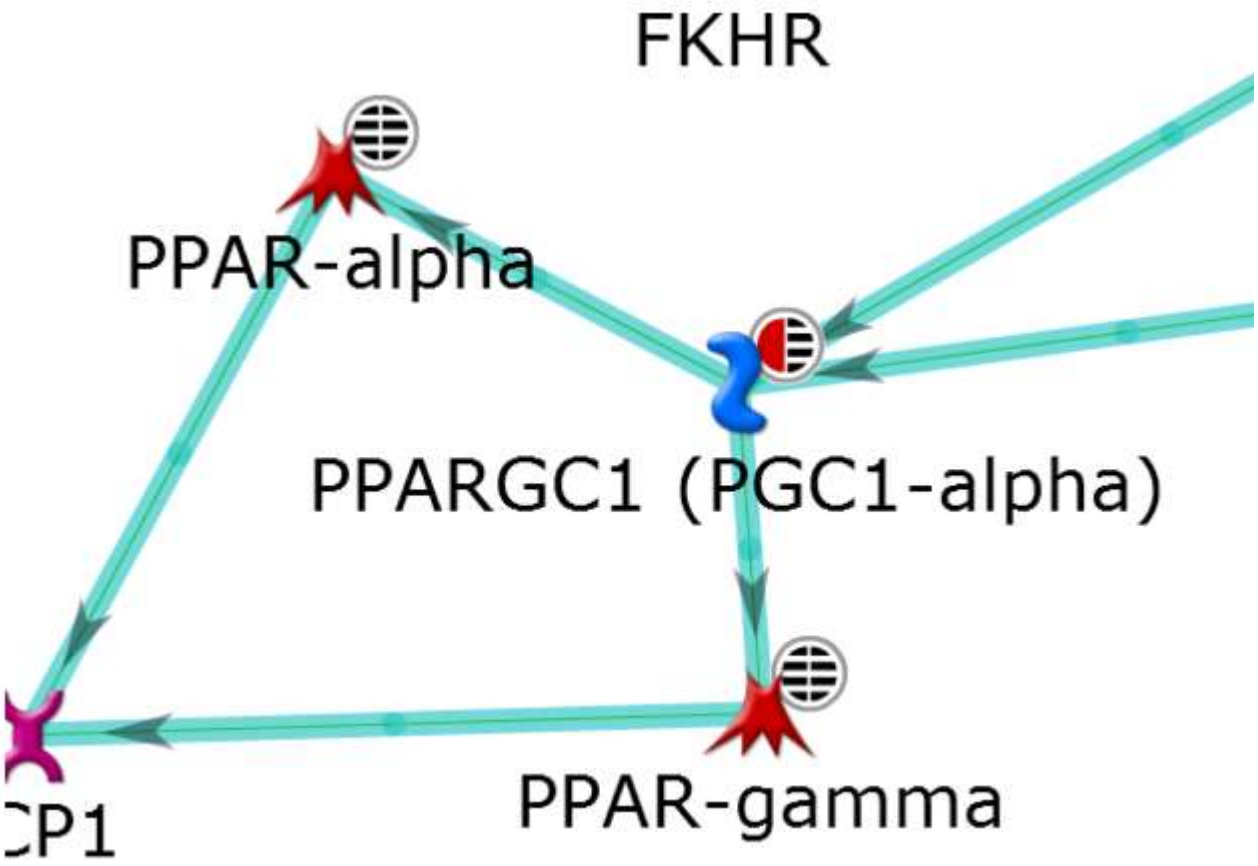
Sample Gene Variant Data (START WITH .VCF FILES)





Sanford|Burnham
Medical Research Institute

Sample Gene Variant Data Integrated with Microarray Data on a Network



Sanford|Burnham
Medical Research Institute

Interactive Software Analysis Session and Q&A

HIGHLIGHTS FROM THE LIVE SESSION

SEARCHING AND BROWSING:

- Learned how to do simple searches (based on the audience input) and retrieve information about up/downstream interactions, disease associations, miR targeting, and annotated aberrations for any gene/protein
- Briefly used advanced search to query all phosphorylated protein interactions that are part of apoptotic processes in humans (based on a question from the audience)

DATA UPLOAD:

- Went over different parsers and ID types, and how to start working with the data

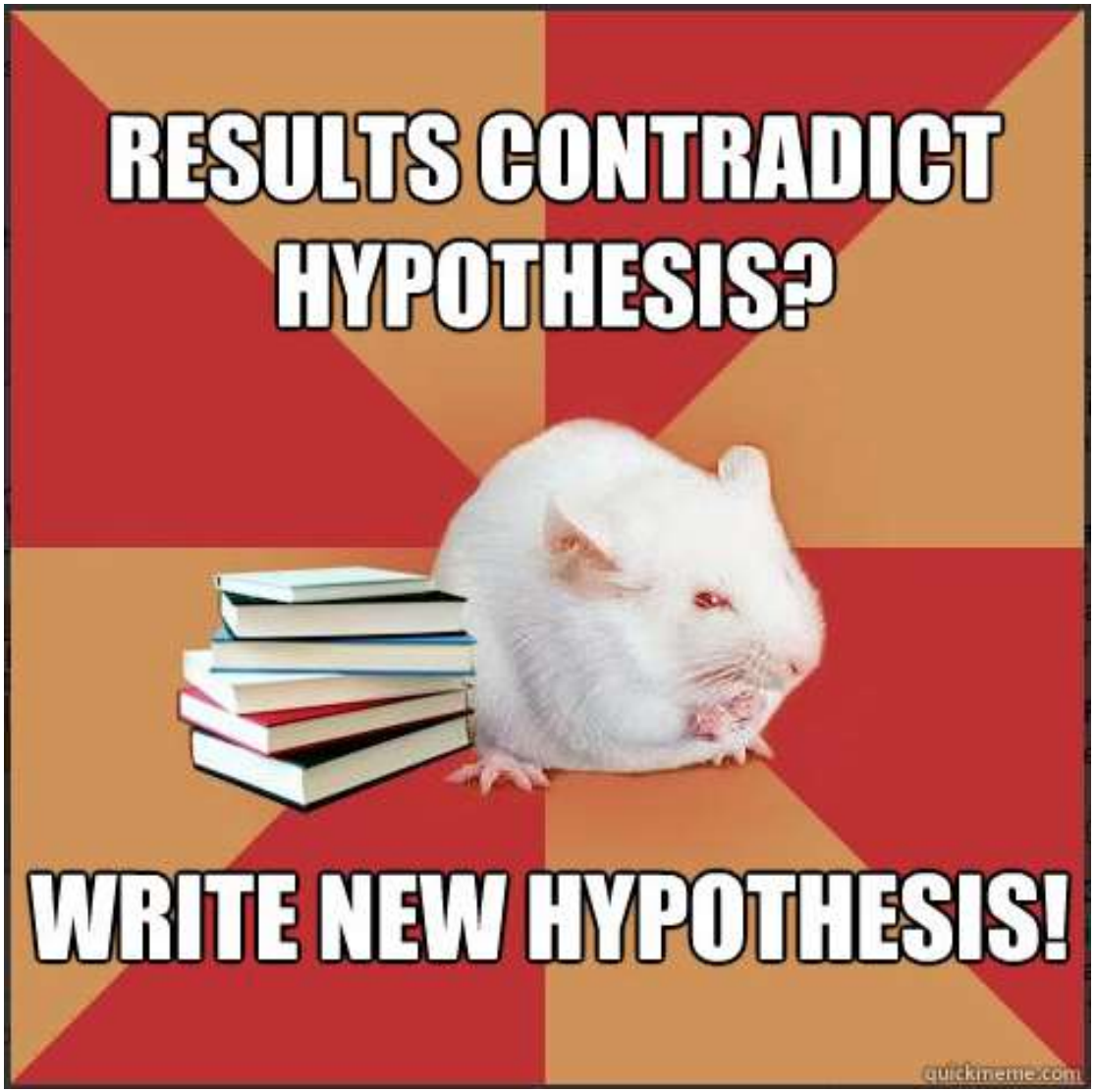
WORKFLOW REPORTS and ENRICHMENT ANALYSIS:

- Briefly covered enrichment analysis with maps (and how to display drugs for targets on maps) and under ONE-CLICK ANALYSIS and Analyze Single Experiment Workflow for the PLOS Genetics data using the “Workflows and Reports” section

NETWORK ANALYSIS:

- Learned how to build a network using the “Trace Pathways” option to connect TRAF3 with IRF3 (suggestions from the audience) obtained from search results and separately starting from the POS Genetics paper differentially expressed genes (Table 1.) connected to SIAH1/2 within 2 interactions
- Learned how to use Filters on networks to display only disease-related genes or only genes that are part of some cellular processes of choice; how to display and toggle different data
- Pathways visualizations on networks (the GPS-like feature)

THANK YOU!


RESULTS CONTRADICT HYPOTHESIS?
WRITE NEW HYPOTHESIS!
quickmeme.com