LOGIC MODELING OF

CELL SIGNALING NETWORKS

J Saez-Rodriguez – Molecular Systems Biology 5: 331 [2009]

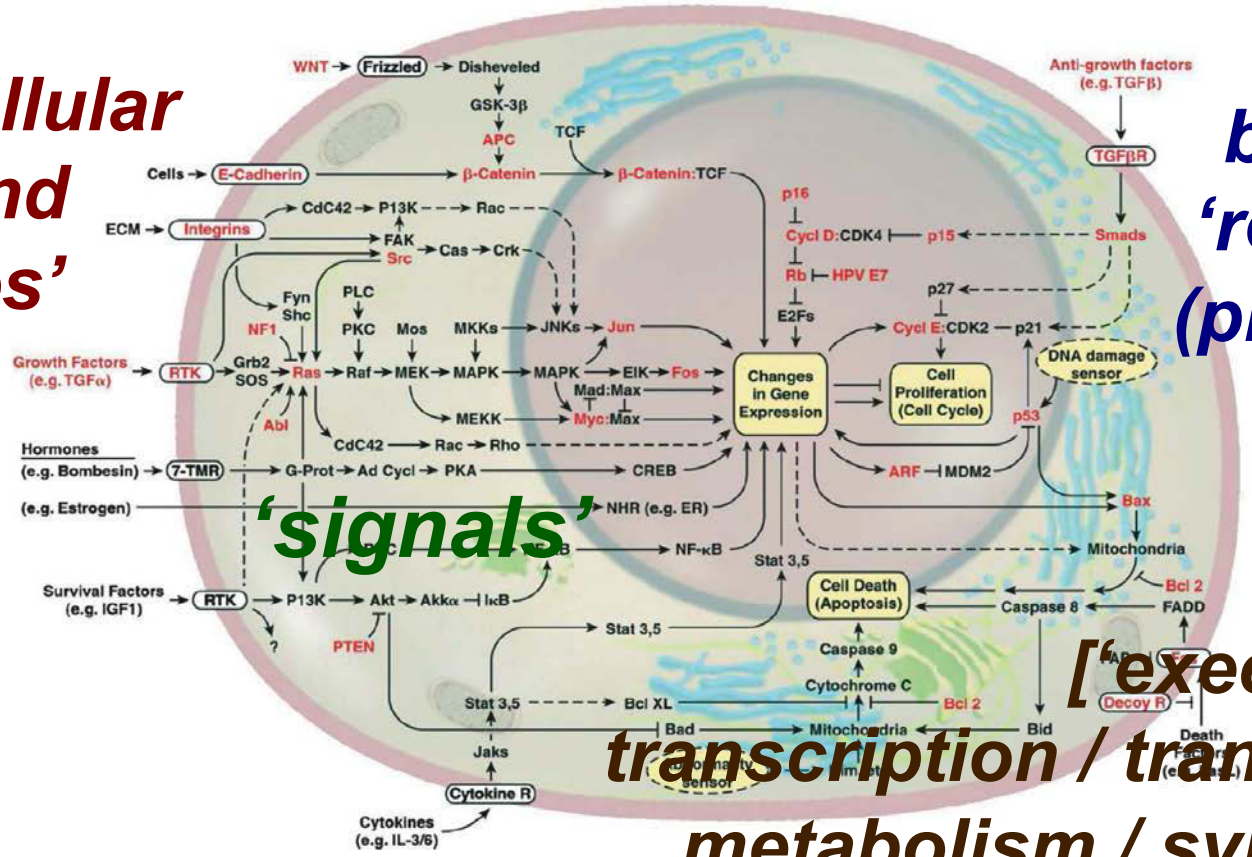
MK Morris – Biochemistry 49: 3216 [2010]

J Saez-Rodriguez – Cancer Research 71: 5400 [2011]

MK Morris – PLoS Computational Biology 7: e1001099 [2011]

Central Topic:
Regulation of Mammalian Cell Behavior by Receptor-Mediated Signaling

extracellular ligand 'cues'



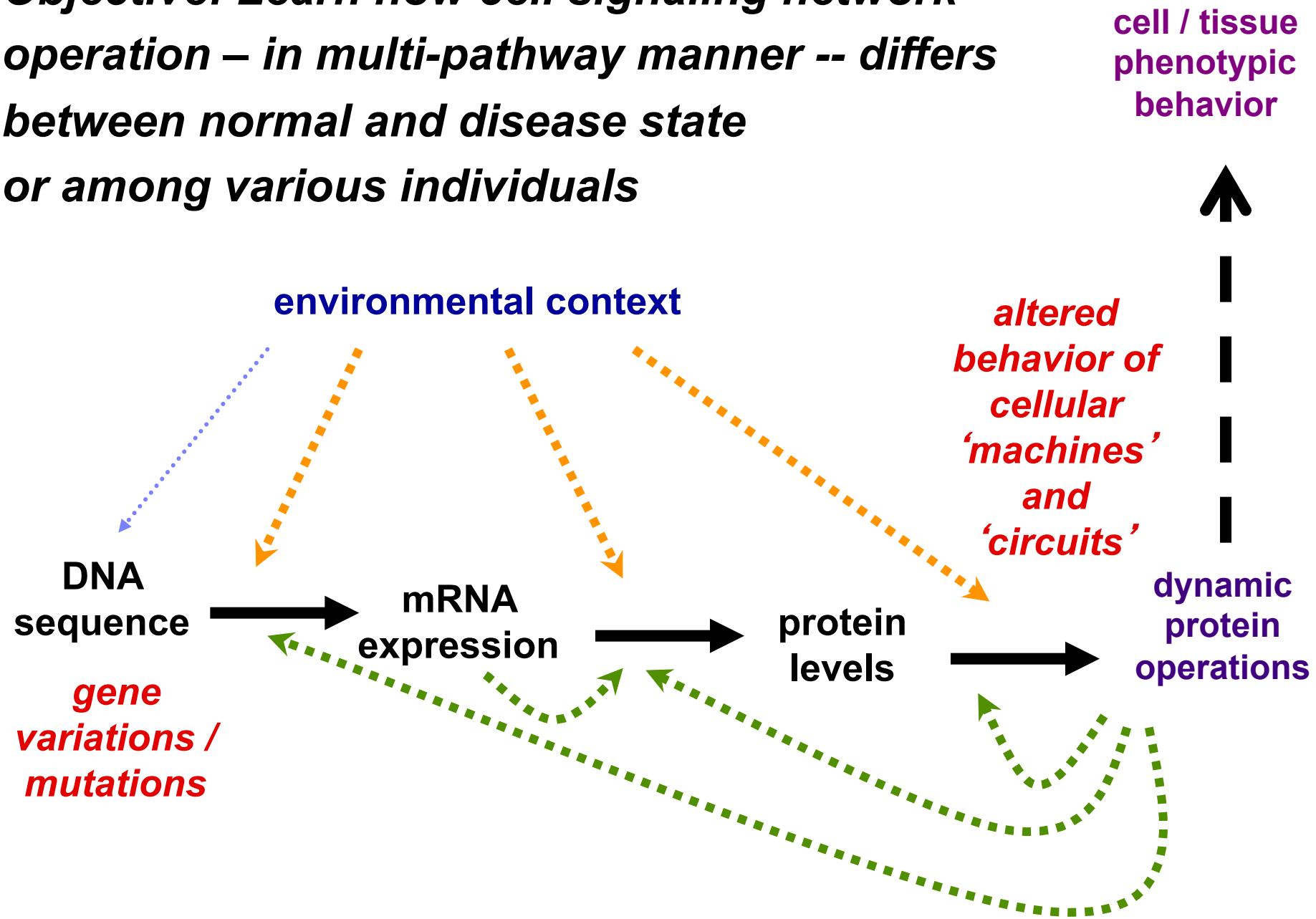
'signals'

behavior 'response' (phenotype)

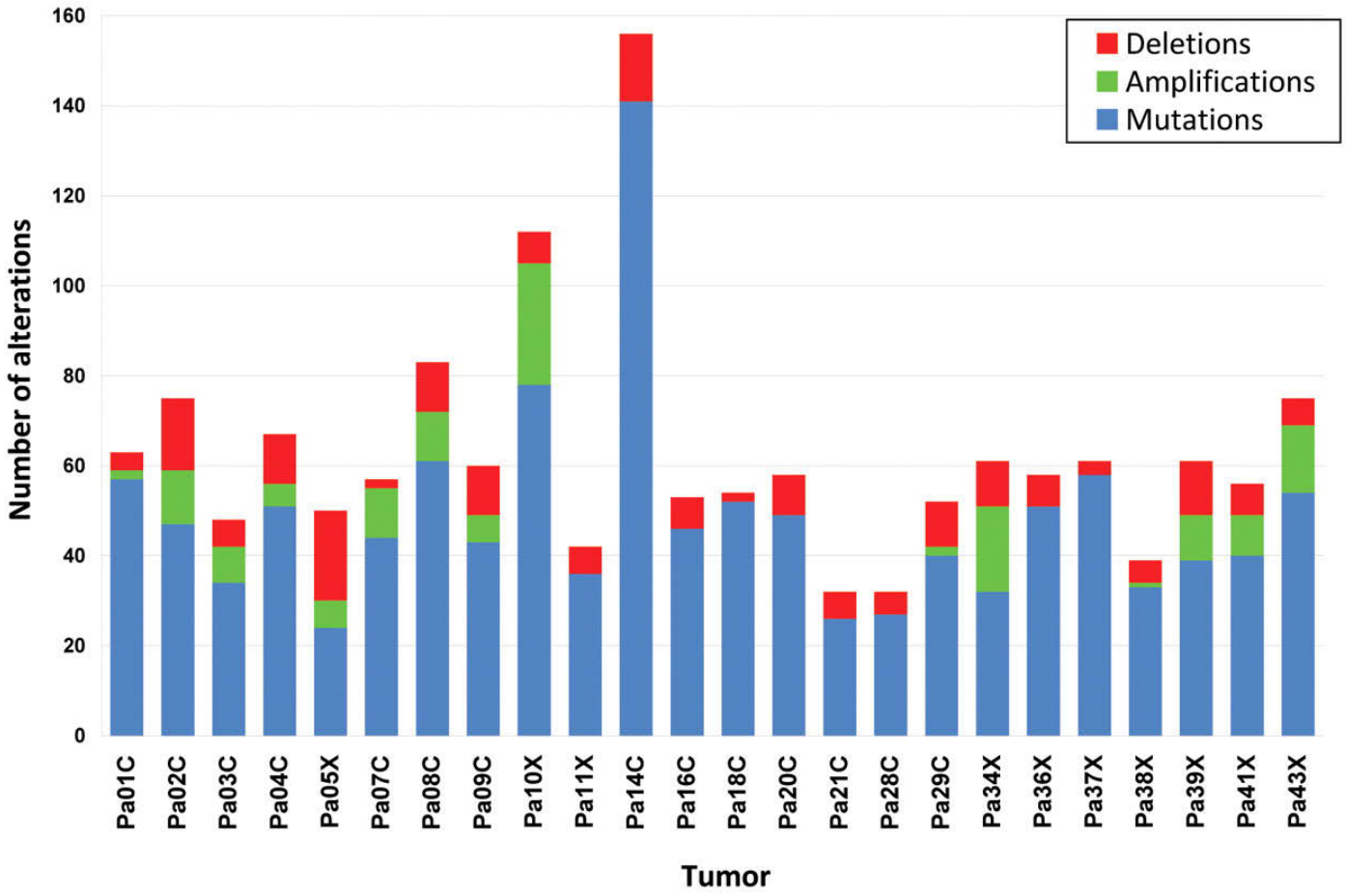
['execution'– transcription / translation, metabolism / synthesis, cytoskeleton / motors]

Source: Hanahan, Douglas, and Robert A. Weinberg. "The Hallmarks of Cancer." *Cell* 100, no. 1 (2000): 57-70.

Objective: Learn how cell signaling network operation – in multi-pathway manner -- differs between normal and disease state or among various individuals



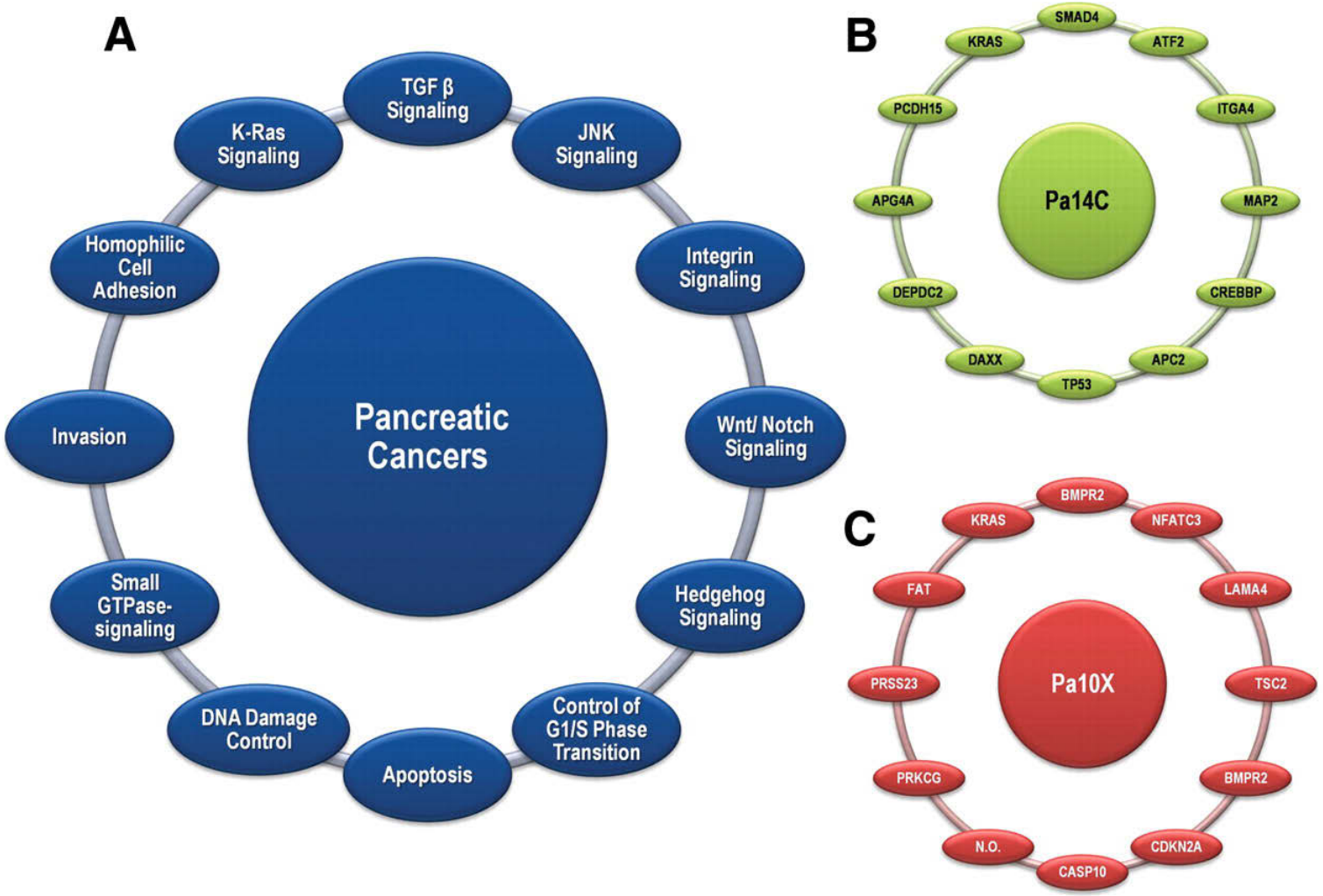
Example: Myriad -- and highly diverse -- genetic alterations (amplifications, deletions, mutations) across pancreatic tumors… (as well as in breast, colon, brain)



Source: Jones, Siân, Xiaosong Zhang, et al. "Core Signaling Pathways in Human Pancreatic Cancers Revealed by Global Genomic Analyses." *Science* 321, no. 5897 (2008): 1801-6.

[Jones *et al.*, Science (2008)]

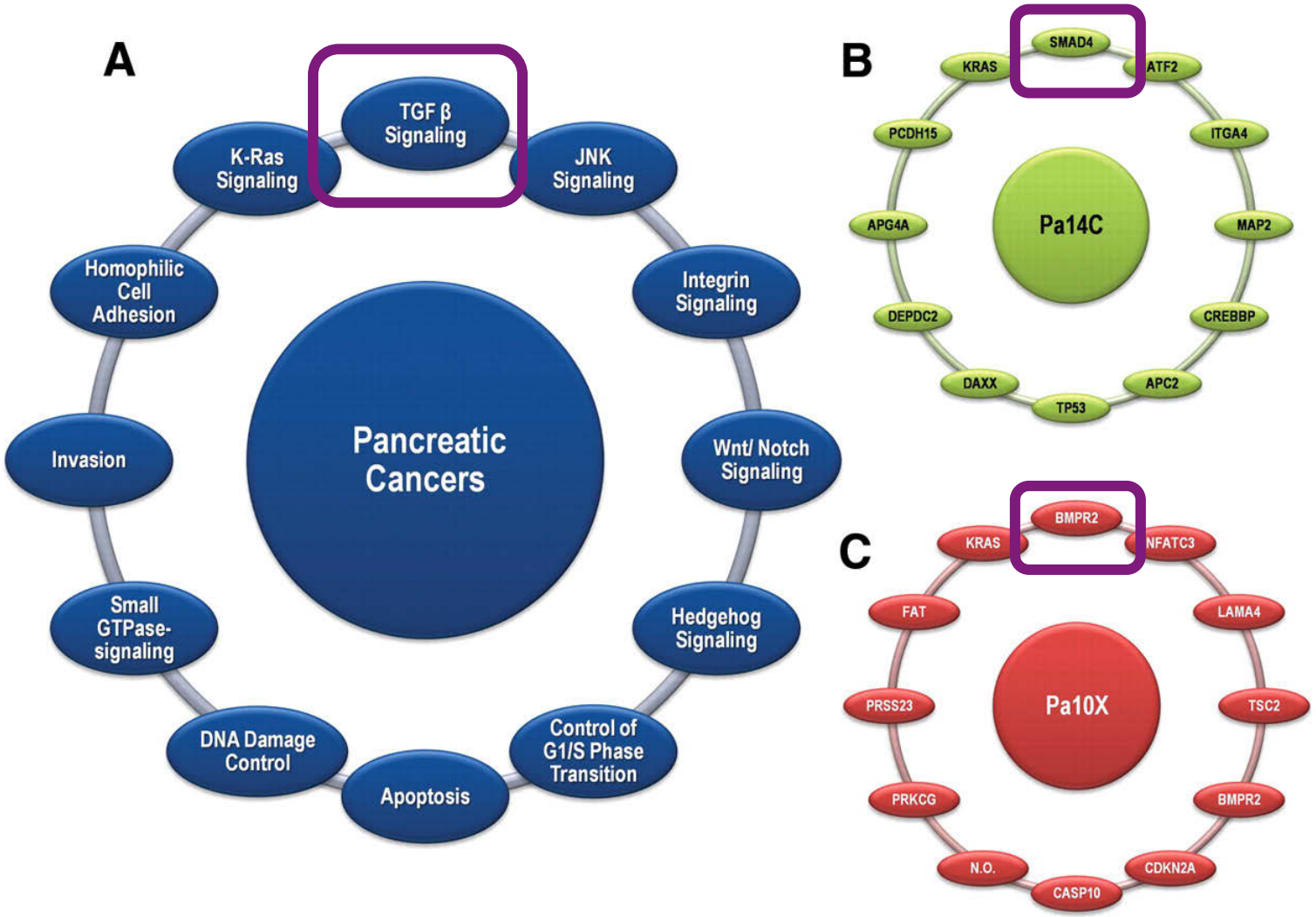
…but diverse mutations lead to dysregulation of a limited set of key pathways at protein level



Source: Jones, Siân, Xiaosong Zhang, et al. "Core Signaling Pathways in Human Pancreatic Cancers Revealed by Global Genomic Analyses." *Science* 321, no. 5897 (2008): 1801-6.

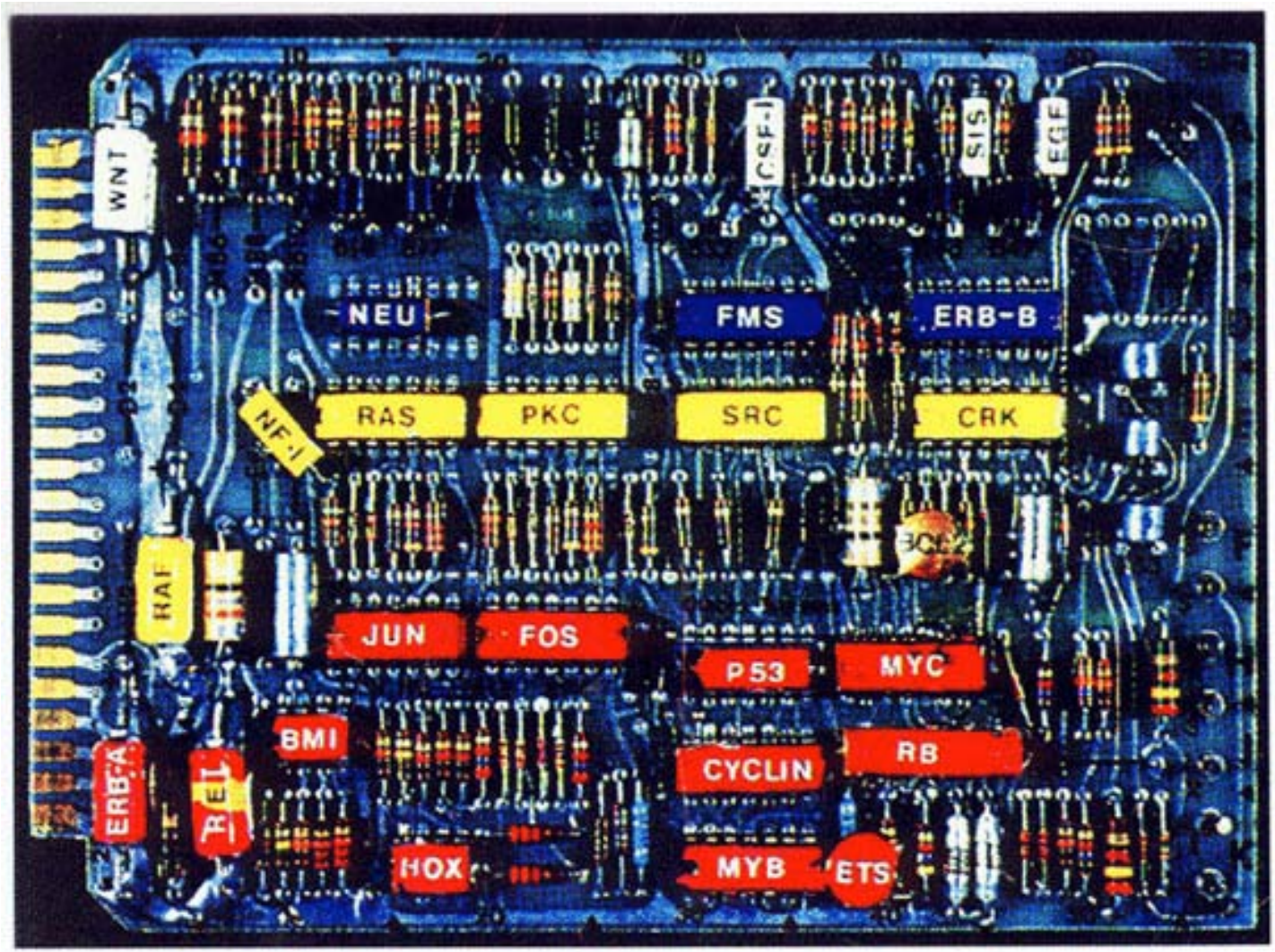
[Jones *et al.*, Science (2008)]

…but diverse mutations lead to dysregulation of a limited set of key pathways at protein level



Source: Jones, Siân, Xiaosong Zhang, et al. "Core Signaling Pathways in Human Pancreatic Cancers Revealed by Global Genomic Analyses." *Science* 321, no. 5897 (2008): 1801-6.

[Jones *et al.*, Science (2008)]



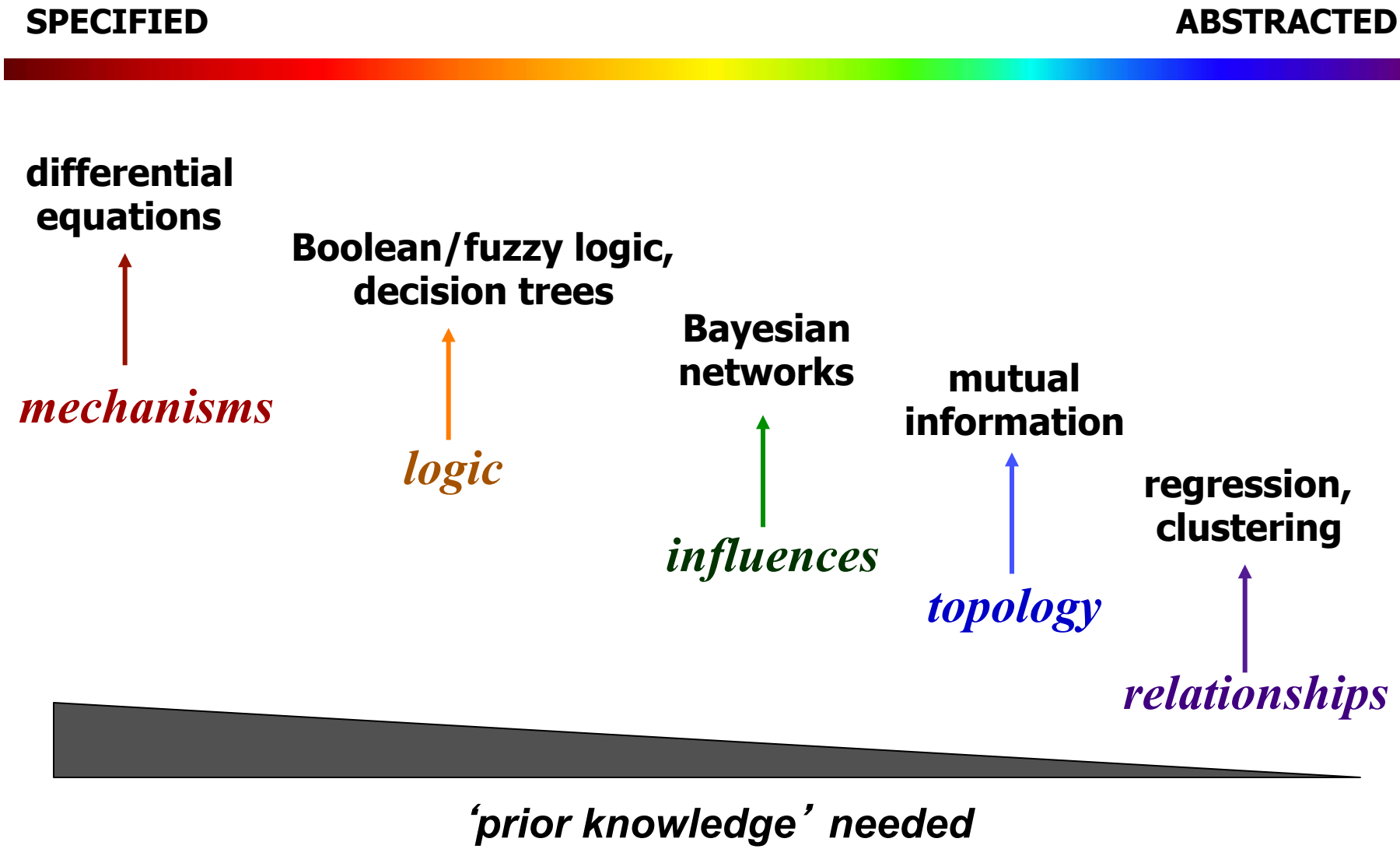
Cell Signaling "Circuitry"

Need to advance from Metaphor to Model

Varmus, H., and R. A. Weinberg. "The Genetic Elements Governing Cancer: Tumor Suppressor Genes." *Genes and The Biology of Cancer* (1993): 101-9.

Varmus & Weinberg, Genes & the Biology of Cancer [1993]

Spectrum of Computational Modeling Methods



Pathway / Interactome Databases hold substantial prior knowledge

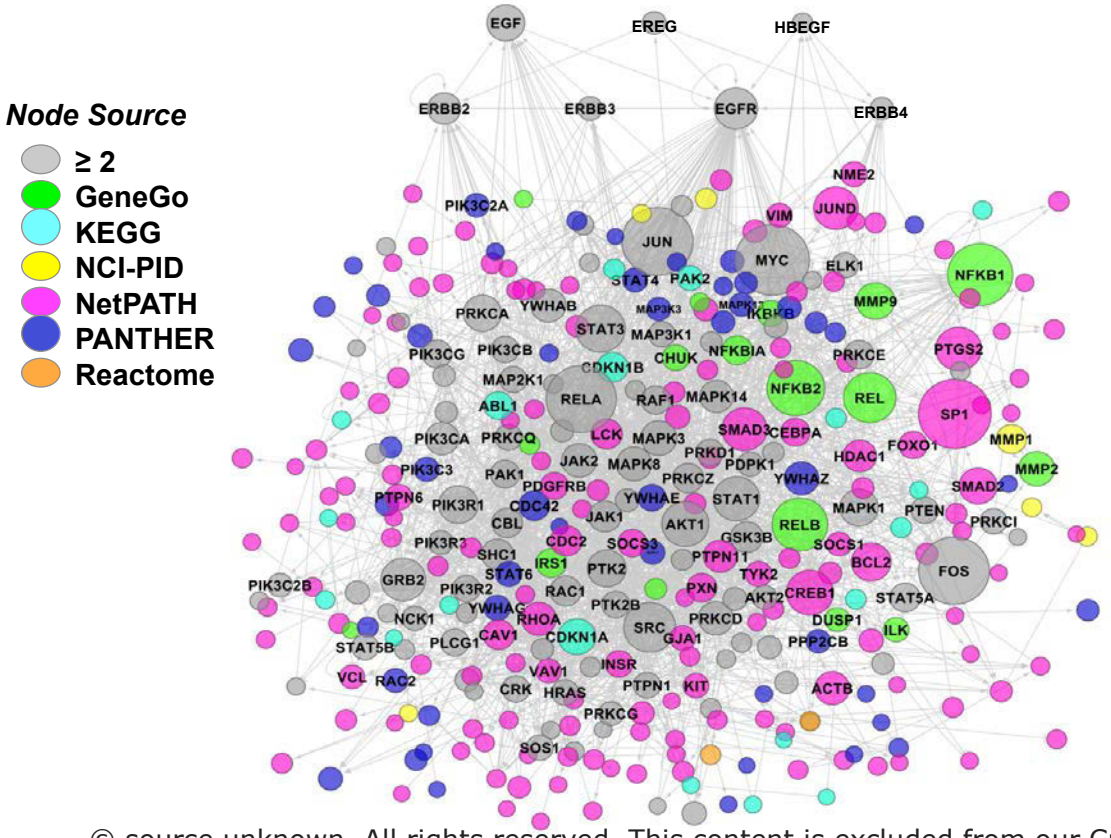
Pathway Databases (Nodes)

Database	Pathways	Relevant	No. Genes	Format
GeneGO	700+	55	804	Table
PANTHER	165	14	1,025	SBML
CellMap (NetPATH)	20	12	625	BioPAX / SIF
Reactome	1081	4	173	BioPAX / SIF
NCI-PID	104	28	459	BioPAX / SIF
KEGG	1000+	8	564	-
SUMMARY		***120***	***2,054***	

Interactome Databases (Edges)

Database	Type	No. Edges	Graph type
i2D v1.71	Protein-Protein (Exp)	11,327	Undirected
STRING	Integrated Text mining	35,033	Mixture
GeneGo	Curated	11,994	Directed, Signed
Cell Map	Curated	12,933	Mixture
NCI-PID	Curated	14,58	Mixture
Reactome	Curated	6,930	Mixture
SUMMARY		***68,067***	***Mixture***


Pathway / Interactome Databases hold substantial 'prior knowledge' for integrative analysis of multi-pathway network effects; but, there is need to move forward from illustration to prediction



© source unknown. All rights reserved. This content is excluded from our Creative Commons license. For more information, see http://ocw.mit.edu/help/faq-fair-use/.

Shortcomings:

- ***Typically diverse with respect to specificity and context – i.e., cell type, genomic content, and/or environmental conditions***
- ***Do not readily permit 'input-output' calculation of network operating behavior, and thus difficult to relate to phenotype and/or interventions***

I2D INTEROLOGOUS INTERACTION DATABASE

PANTHER Classification System

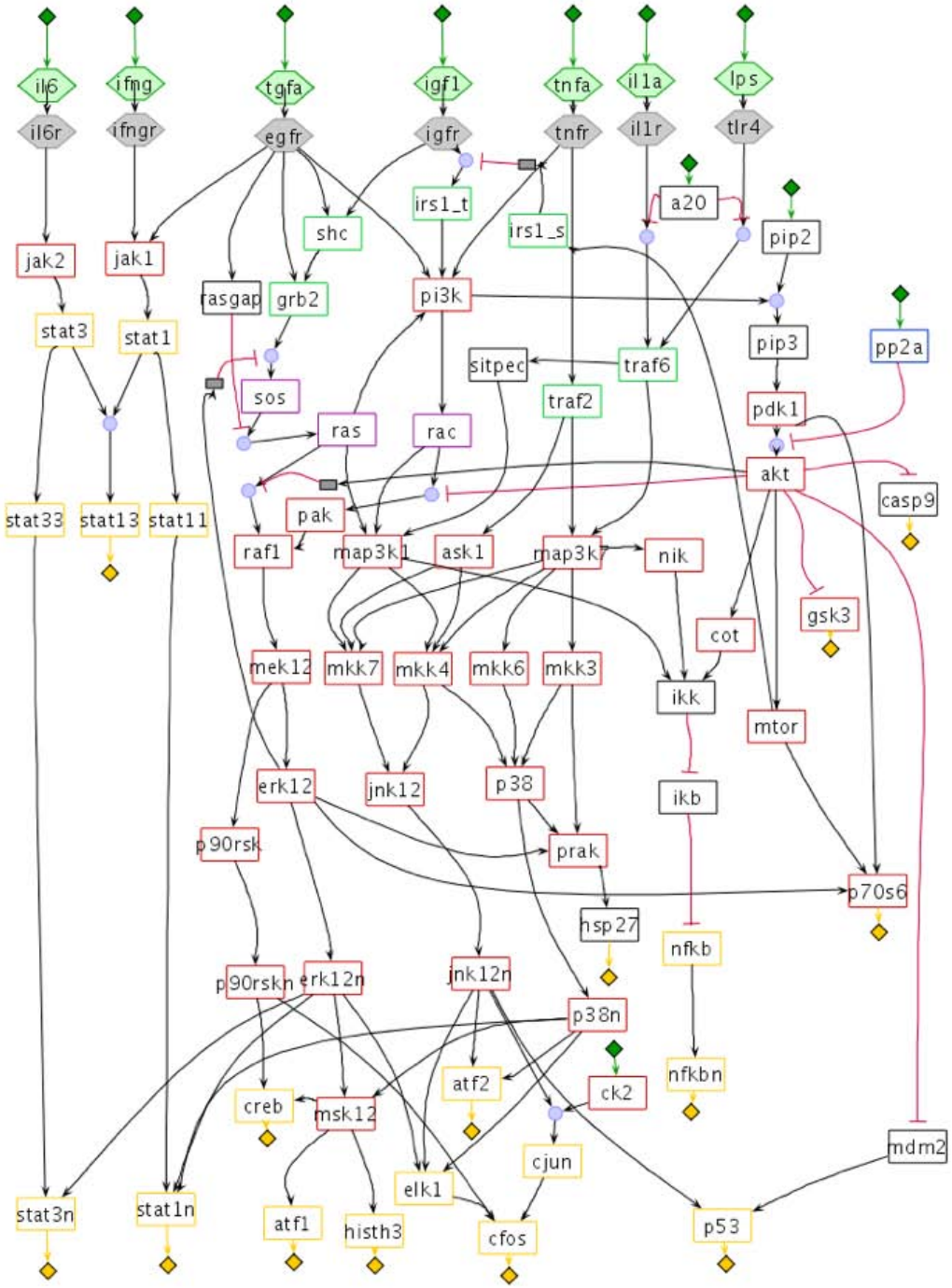
KEGG Kyoto Encyclopedia of Genes and Genomes

NATIONAL CANCER INSTITUTE nature PathwayInteractionDatabase

gene

NetPath

STRING



Central Goal:

Establish methodology for converting from qualitative cell pathway topology 'maps' to quantitatively computable network models

Approach:

Employ logic-based modeling framework, to train qualitative 'prior knowledge' maps to quantitative empirical data for system context and multi-pathway comparisons of interest

More Detailed Insights from Stronger Modeling Analysis -- integrating empirical data with prior knowledge using network logic approach

Generic Pathway Map

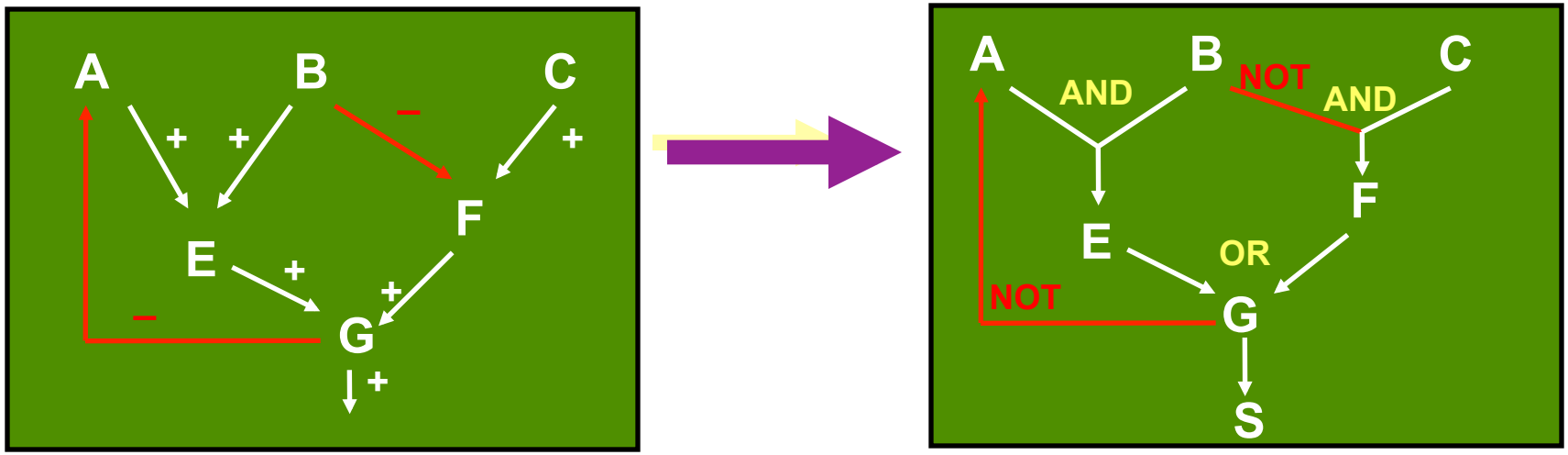
(*e.g.*, Ingenuity)

nodes (=compounds), signed directed edges (activation +, inhibition -)

Network Logic Model

Boolean operators:

AND / OR / NOT

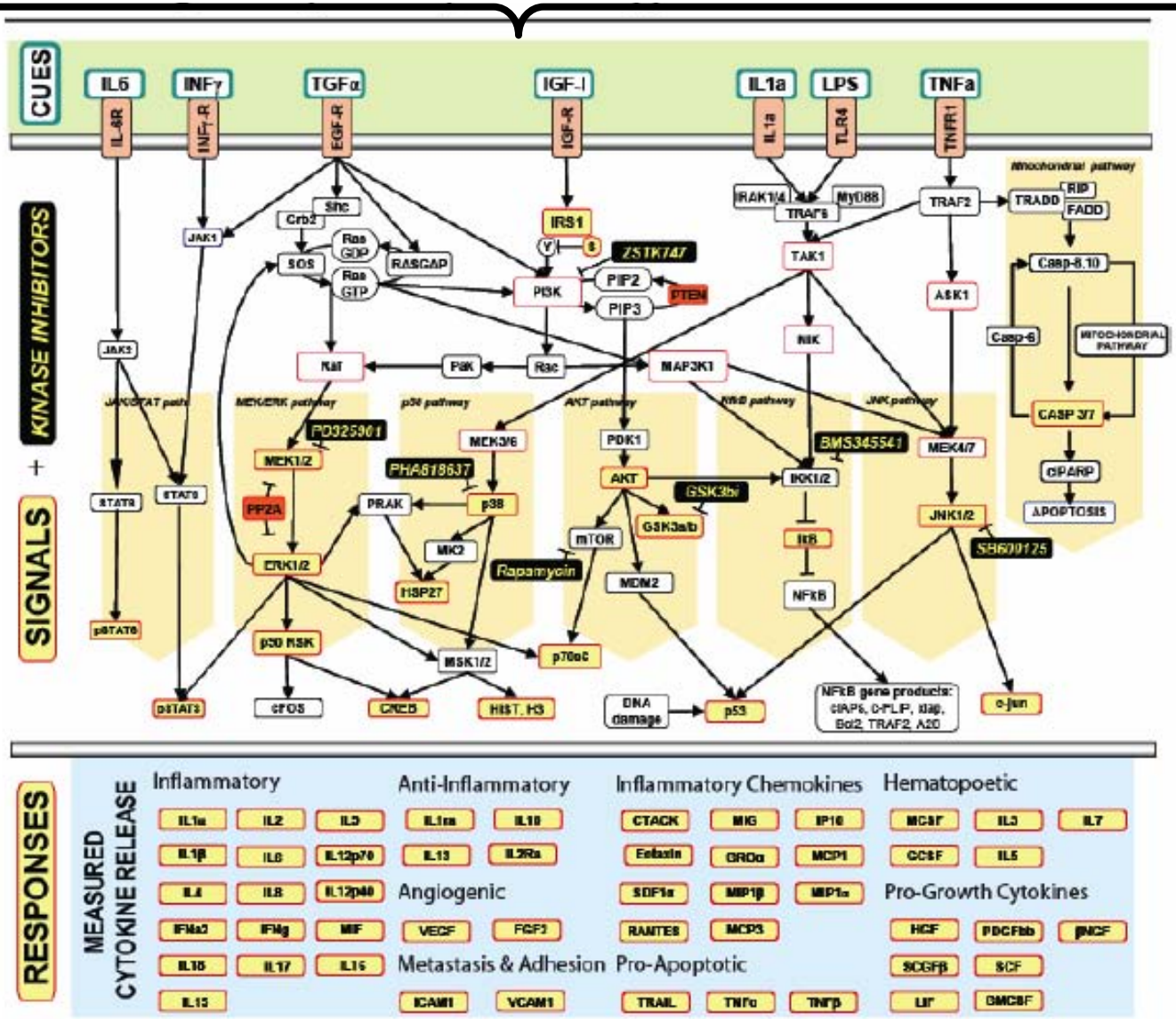


Example Study: Comparative Hepatocytic Cell Signaling Network Operation in Inflammation Context

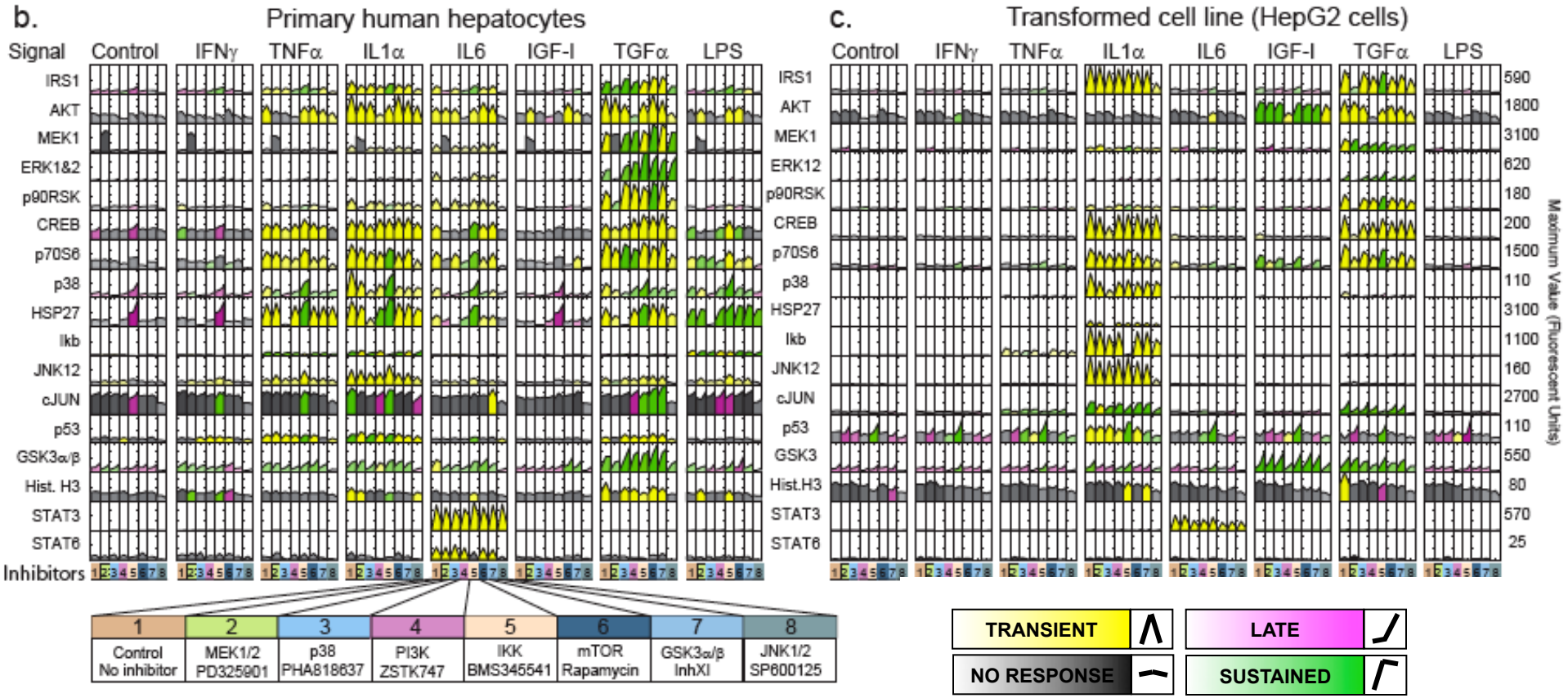
hepatocellular lines

Primary Hepatocytes | **HepG2** | **Huh7** | **Hep3B** | **Focus**

[7 ligands + control] X [7 inhibitors + control] X 17 signals = ~ 1000 measurements for each cell-type and time-point and replicate

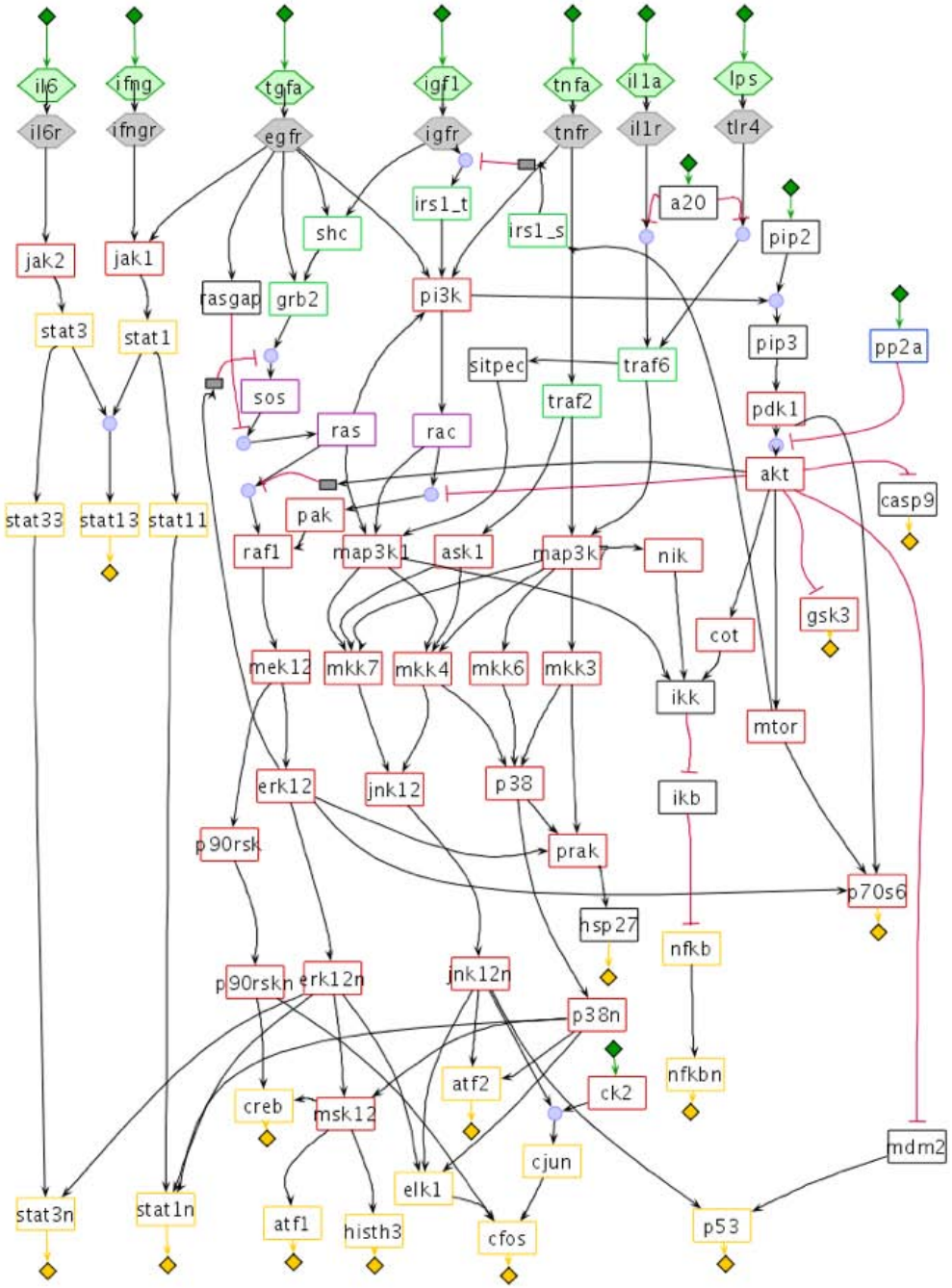


Multi-Pathway Phosphoproteomic Data – primary human hepatocytes, HepG2 hepatocellular line



Time-points: 0, 30 min, 3 hrs

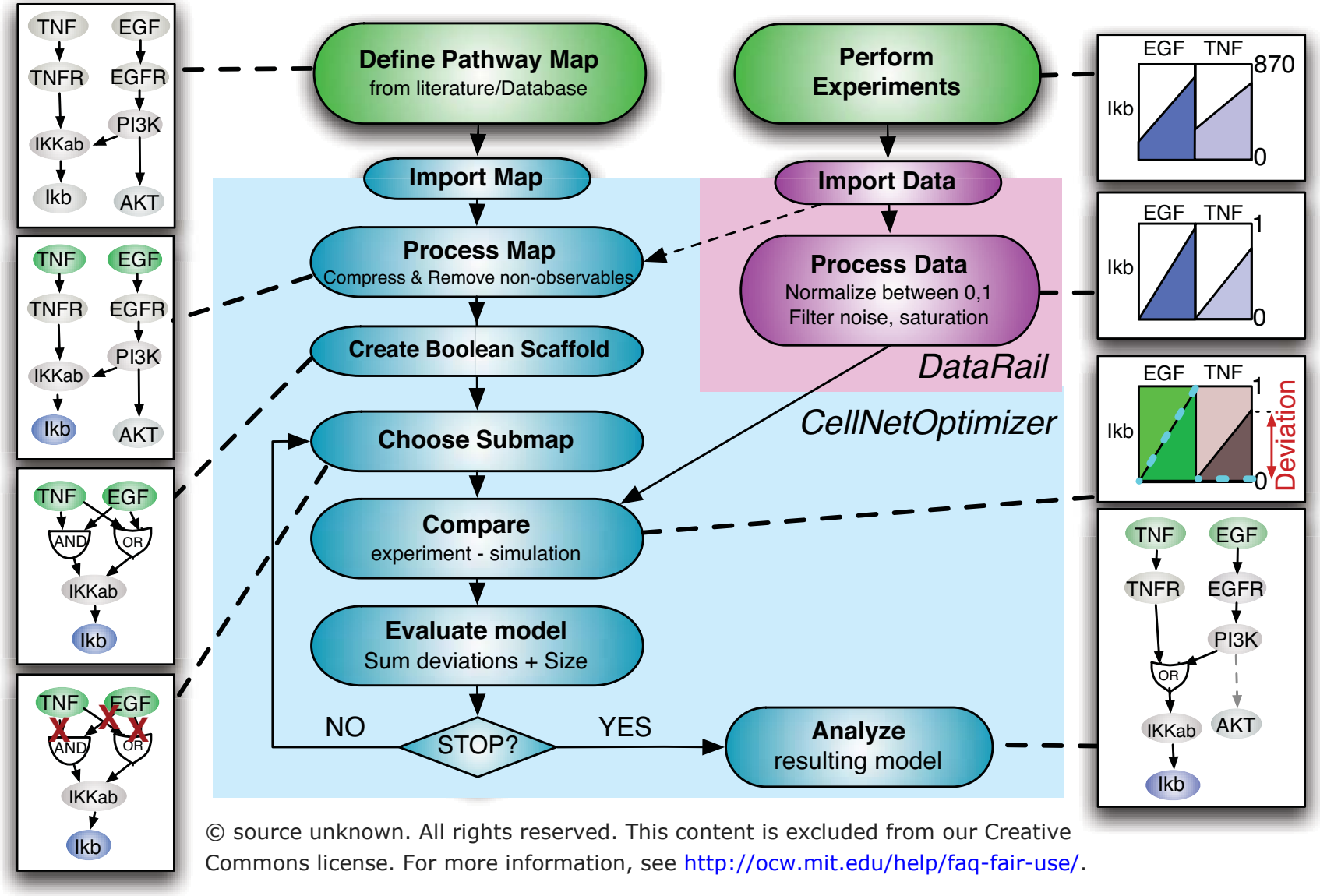
-- also cell death, proliferation index, and production of ~50 cytokines for each condition



Example Database Pathway Map from literature curation, for network responses to our cytokine and growth factor treatments

from *Ingenuity* supplemented by some literature knowledge for key receptors -- 82 nodes, 116 edges

Training Prior Pathway Map Knowledge on Context-Specific Empirical Signaling Data



Automated Development of Logic Network Models from Fit of Generic Pathway Map to Experimental Data as an Optimization Problem

Objective Function

$$\theta = \theta_f + \alpha \cdot \theta_S$$

Fit to data

$$\theta_f = \sum_{l=1}^{S} \sum_{K=1}^{M} (Bi_{kl}^{M} - Bi_{kl}^{E})^2$$

$\in \{0,1\}$ $\in [0,1)$

Relative importance Fit vs. Size

Size of model

$$\theta_S = \sum_{k=1}^{n} \nu_k P_k$$

- **minimize Objective Function (*θ*) across model variants (P),**
- **trading off model-data error and model size;**
- **α ascertained by Pareto optimum for false-positive vs false-negative trade-offs**
- **obtain family of best-fit models (within 1% of Objective Function optimum)**

Automated Development of Logic Network Models from Fit of Generic Pathway Map to Experimental Data as an Optimization Problem

Genetic Algorithm

1. **Initialize a population of model variants (from Ingenuity scaffold or from random scaffolds)**
2. **Evaluate objective function (model-vs-data error plus model-size penalty) for each individual in the population**
3. **Generate next generation of population using Elite Survival, Fitness Selection, Mutation, and Crossover**
4. **Assess whether stop criterion is fulfilled, or iterate back to step 2**
5. **Model pruning to reduce model size without detriment to model-vs-data error**
6. **100 runs for each value of model-size penalty α**

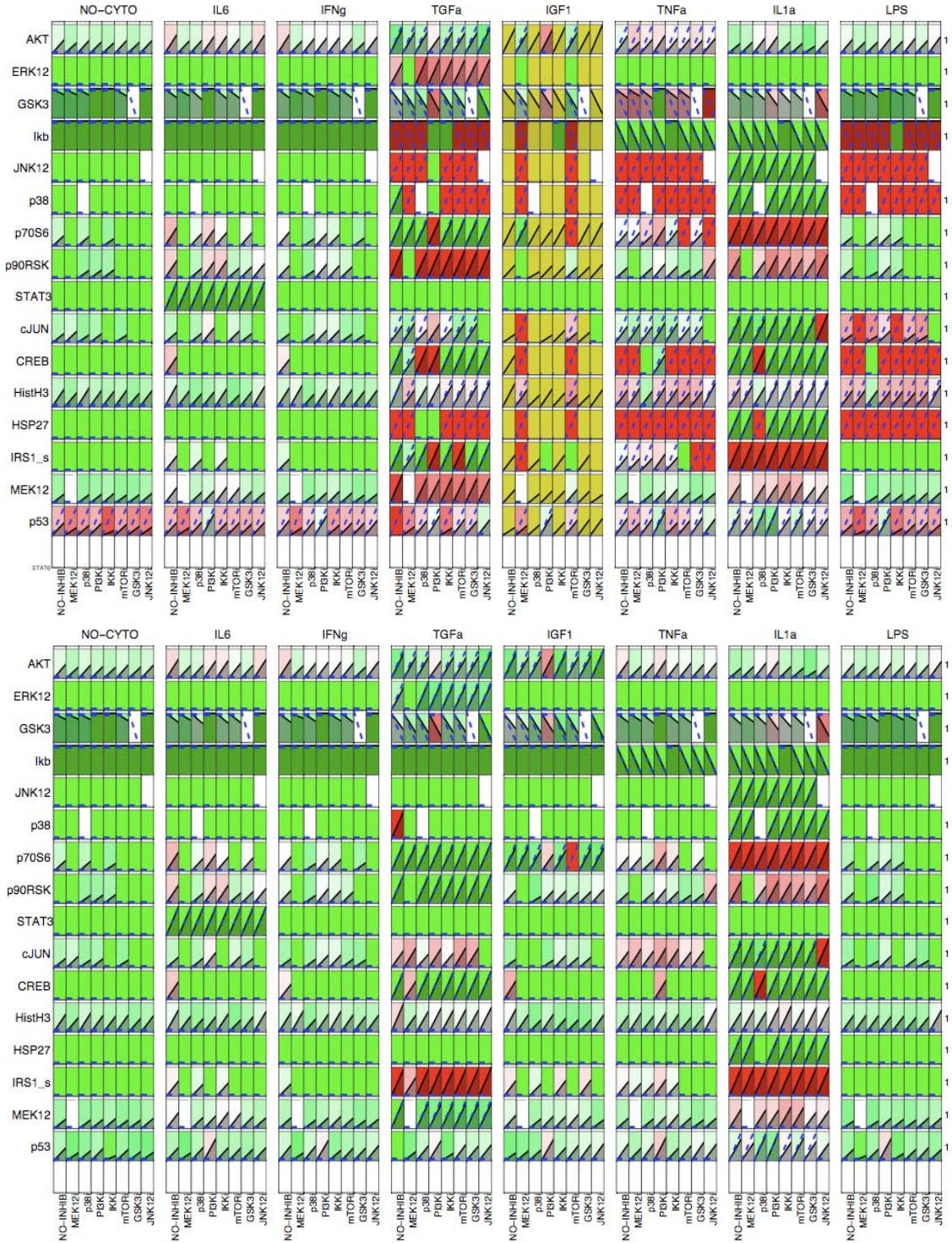


Illustration for HepG2 cell line

– improvement in data fit from best-fit original scaffold model to best-fit trained model

Training data fit to ~9% error, substantially improved from original scaffold model fit of >45% error

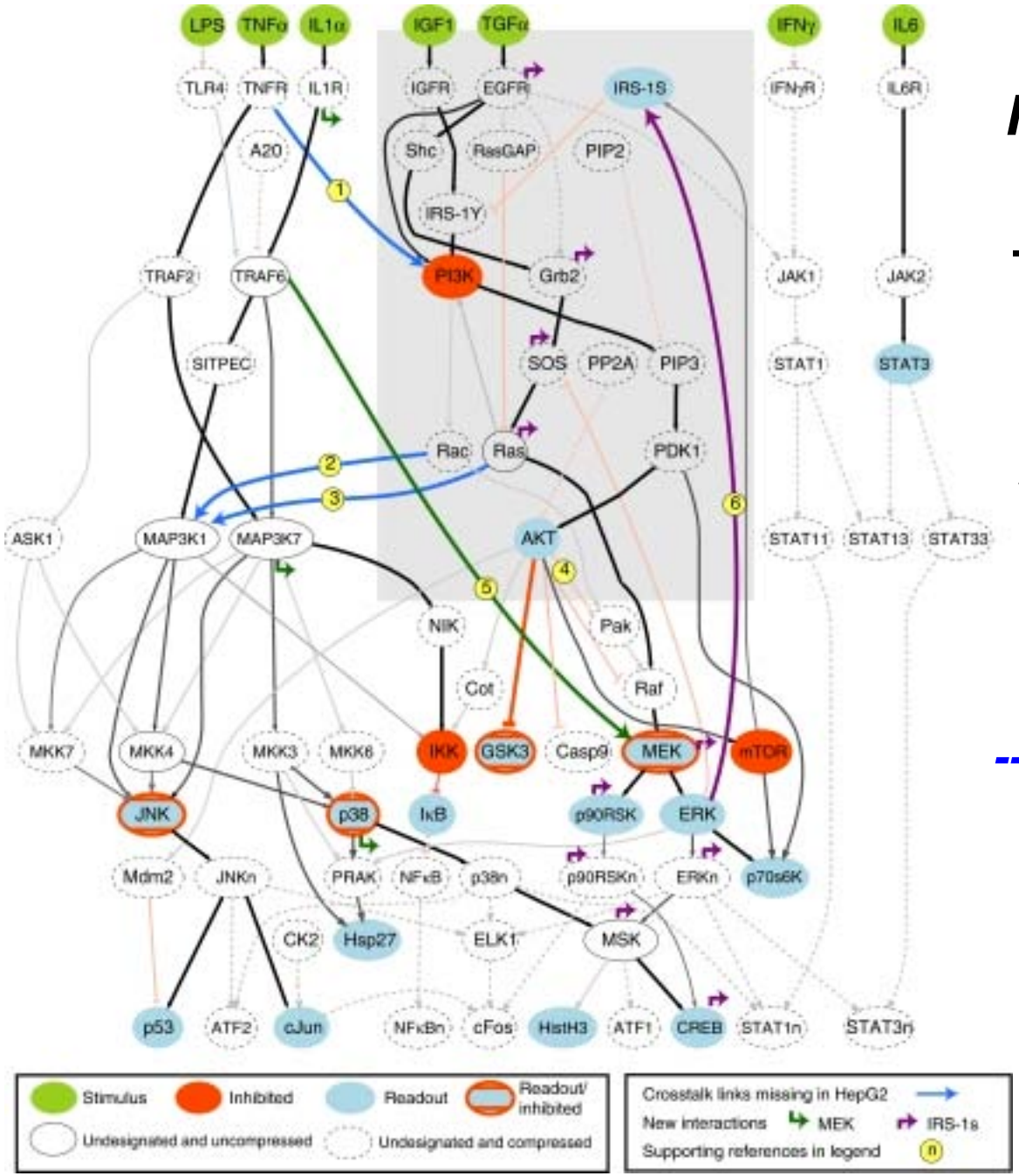


Illustration for HepG2 cell line

– consensus model from fit of empirical data to initial prior knowledge scaffold

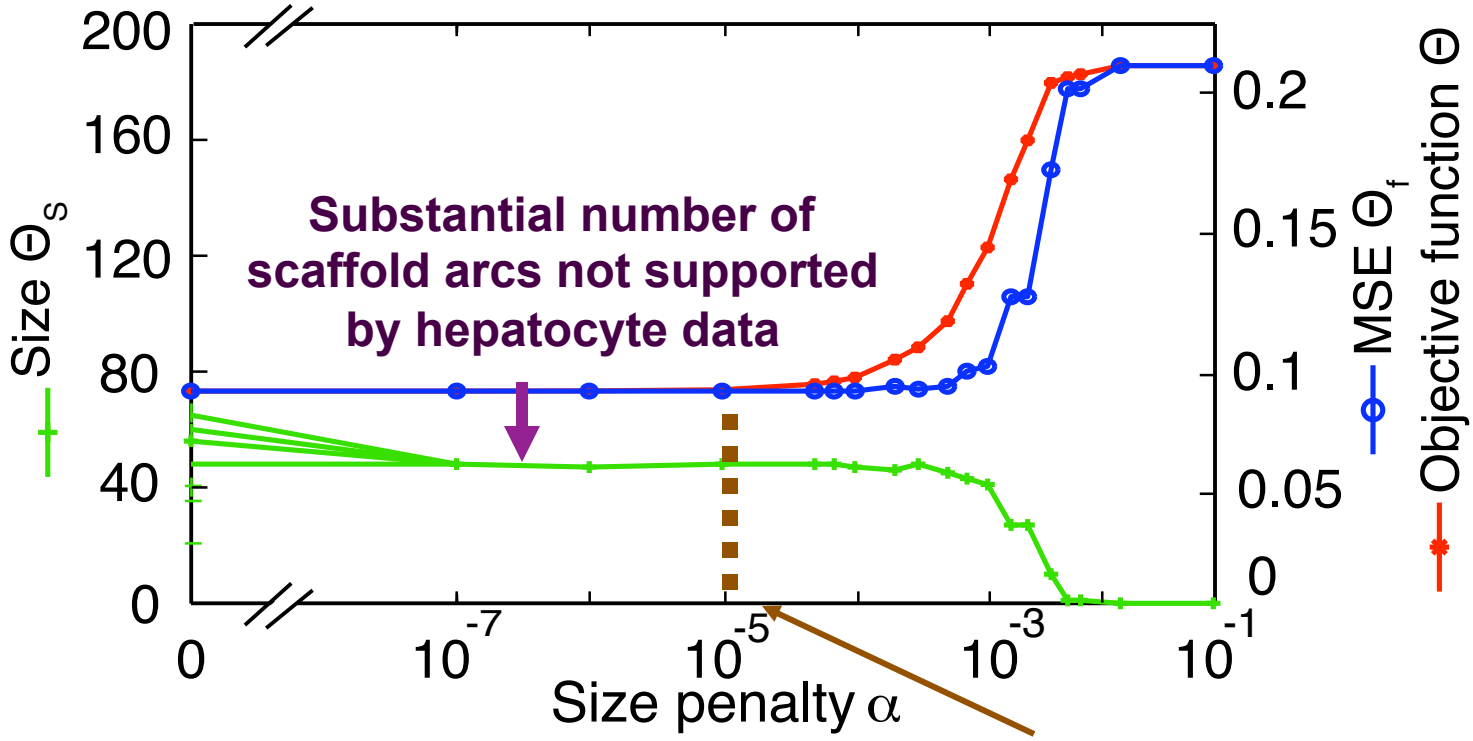
-- additional arcs needed to improve model fit, support in literature though not in prior knowledge scaffold

-- arcs present in other cell line models but not in HepG2

Source: Saez-Rodriguez, Julio, Leonidas G. Alexopoulos, et al. "Discrete Logic Modelling as a Means to Link Protein Signalling Networks with Functional Analysis of Mammalian Signal Transduction." *Molecular Systems Biology* 5, no. 1 (2009).

Model size is fairly insensitive to size penalty

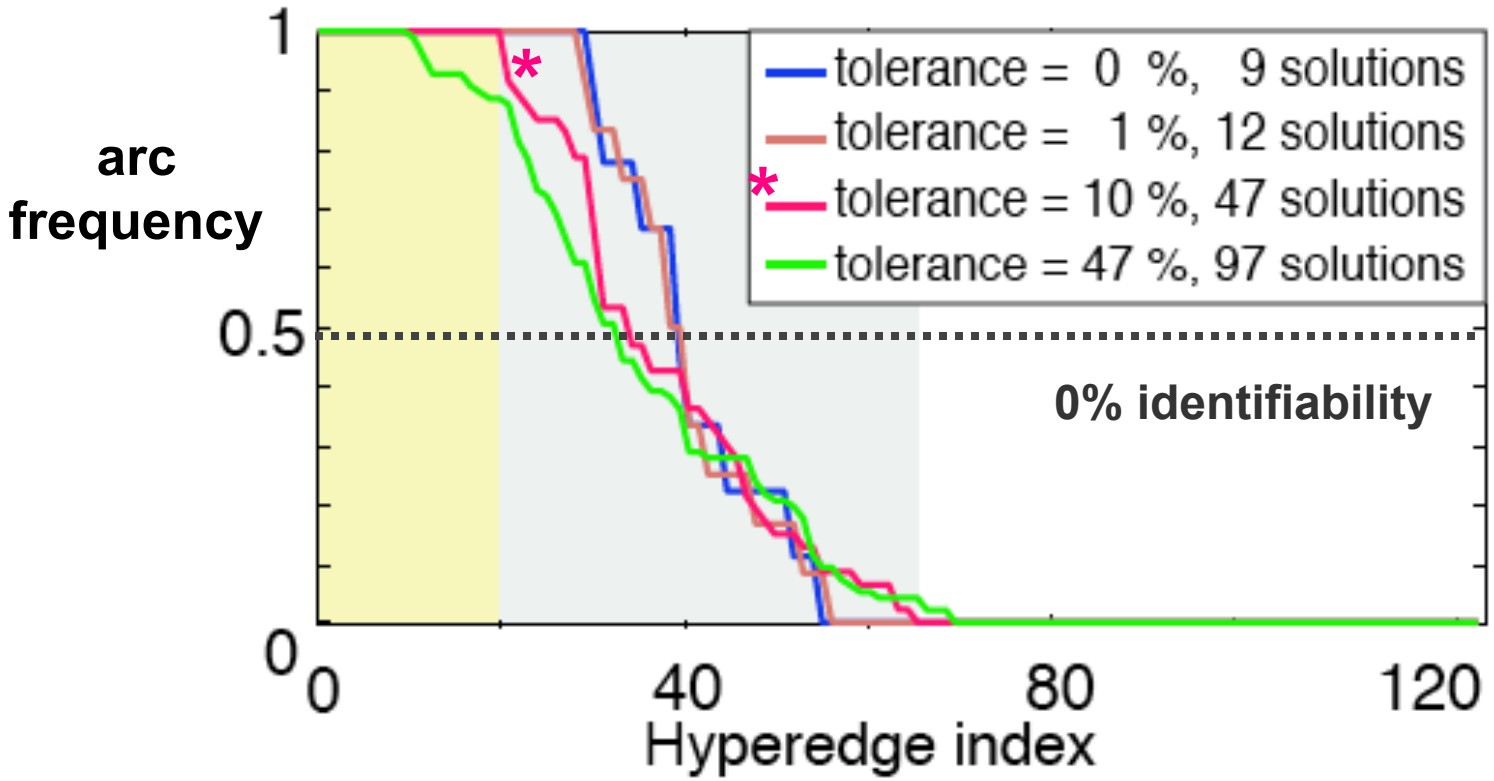
Objective function = Fit of data (MSE) + α Size



selected model-size penalty, for maximal predictive capability

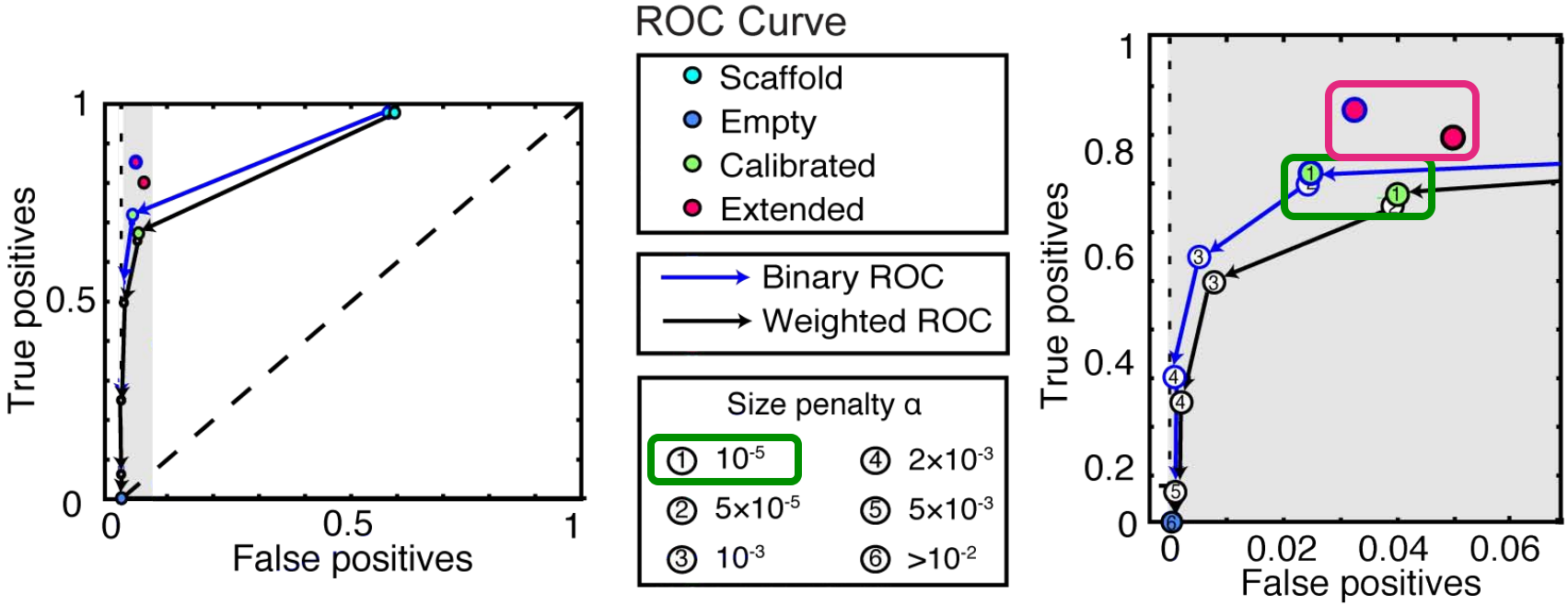
Models can only be partially identified
-- thus model families are best outcome

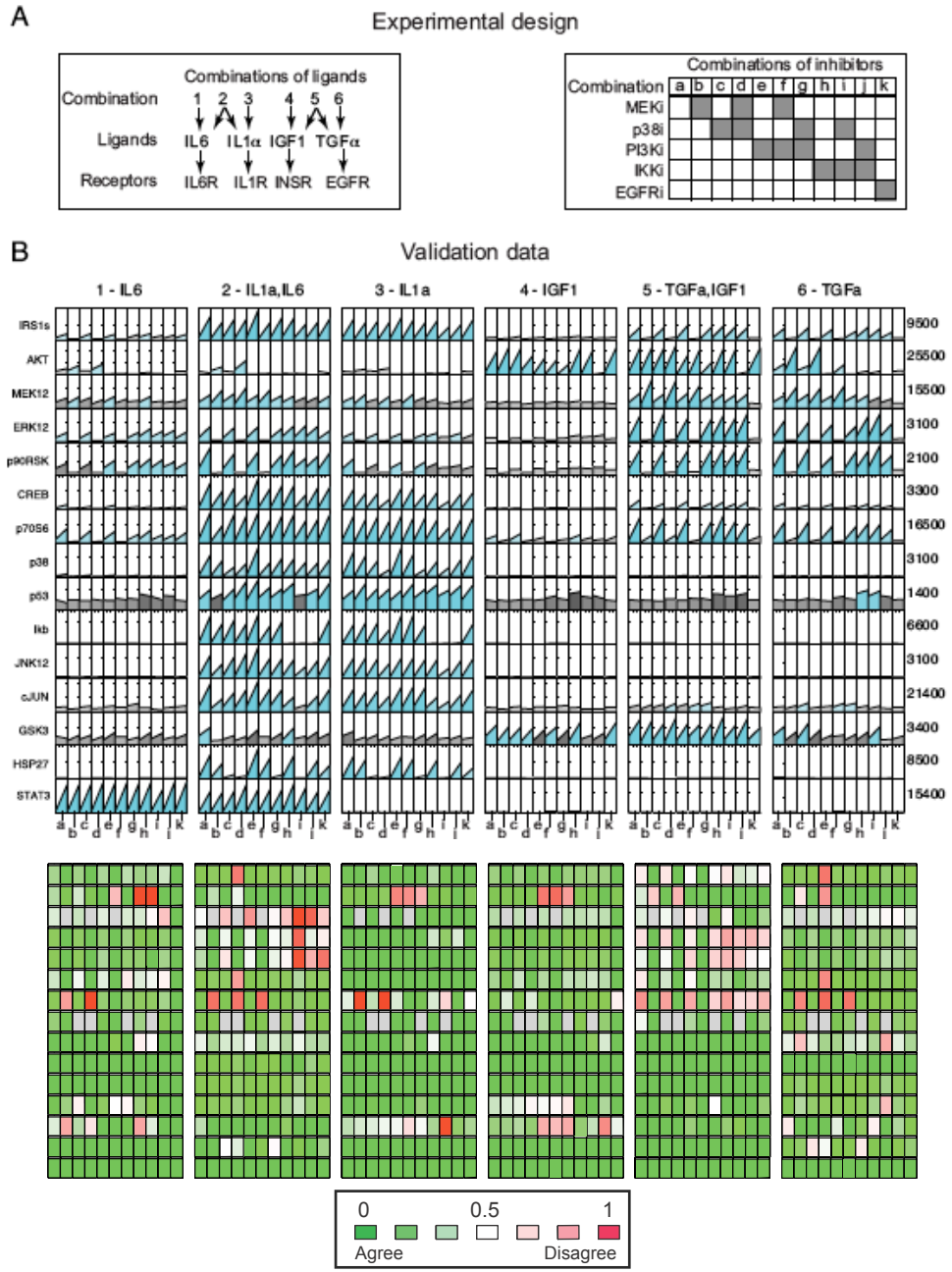
Frequency of Arc Distribution for Error Tolerance-Related (*i.e.*, beyond exptl uncertainty) Model Families



Trade-off between False Negatives and False Positives

- **Receiver Operating Characteristic (ROC) curve [ratio of true positives (1-false negatives) vs. false positives] for different values of the size penalty α**
- **Optimal choice of size penalty ($\alpha=10^{-5}$) corresponds to most predictive model**
- **Extended model (*i.e.*, with added arcs) decreases false negatives but increases false positives**

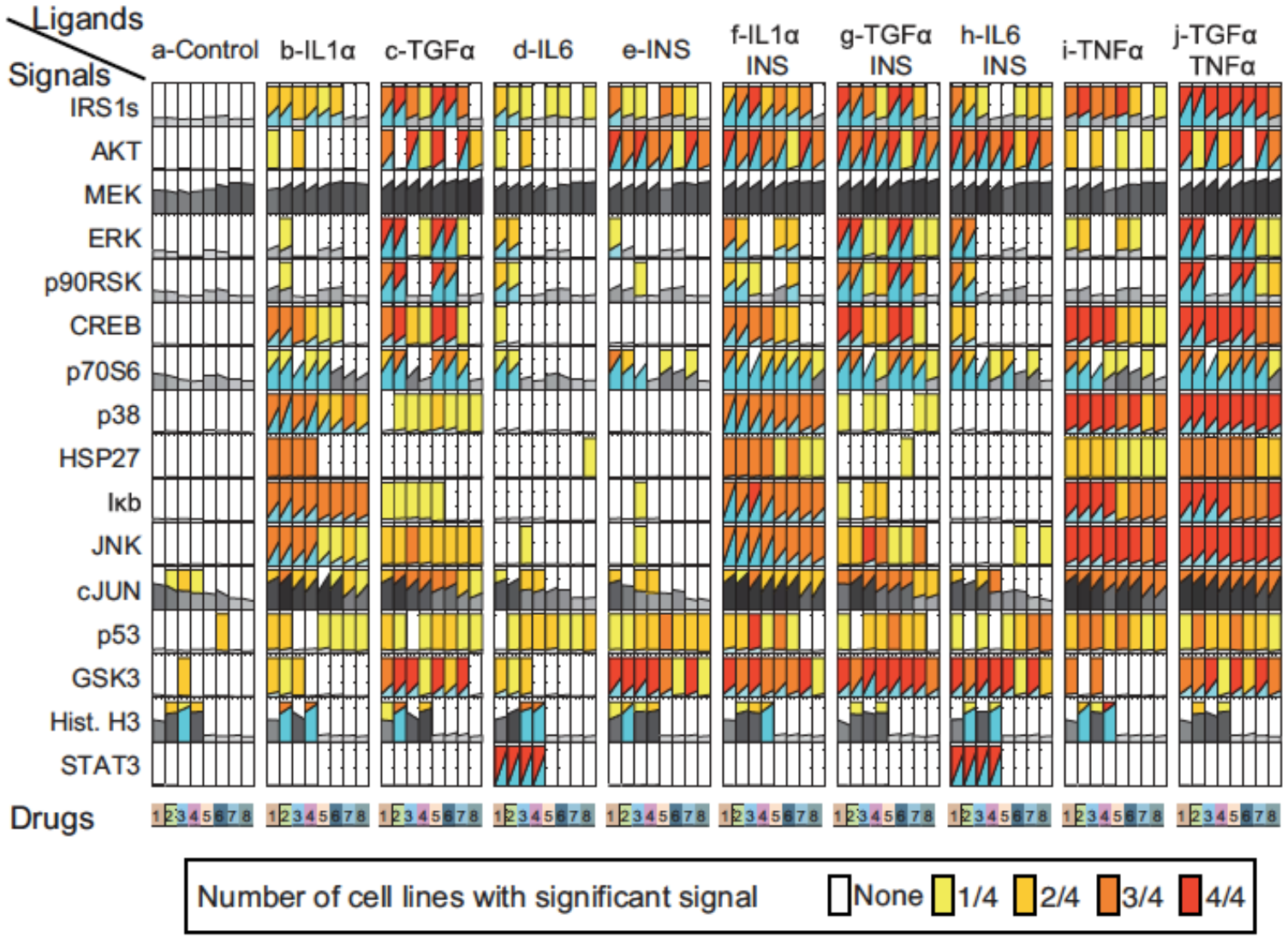




Model Validation -- successful a priori *predictions of new test data*

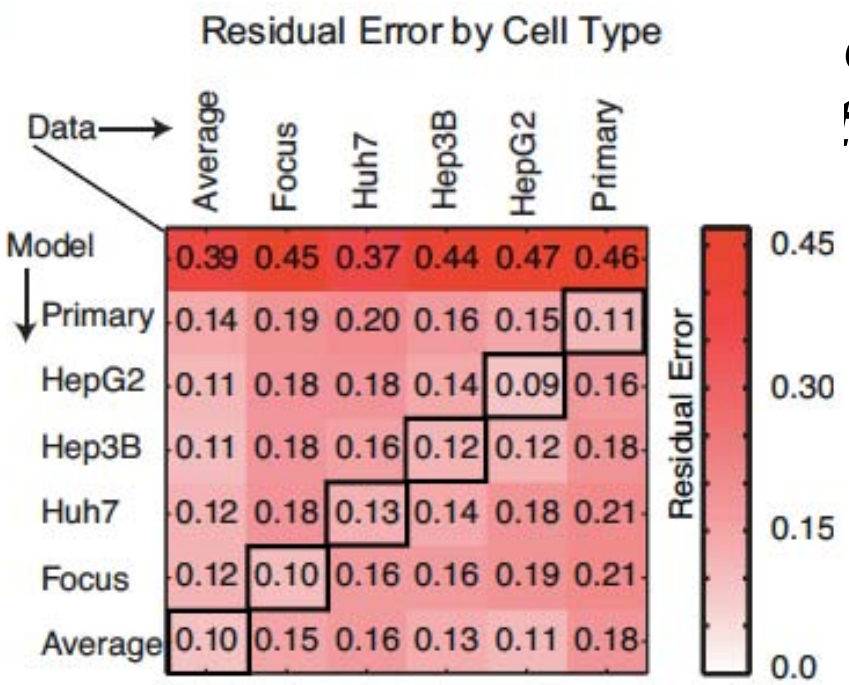
- **Used trained model to *a priori* predict effects of ligand combinations, additional inhibitors, and inhibitor combinations**
- **New test data predicted to within ~11% error, comparable to ~9% for original training data**
- **Can identify loci needing more detailed inquiry**

Extension to comparison among hepatocellular lines -- phosphoproteomic data

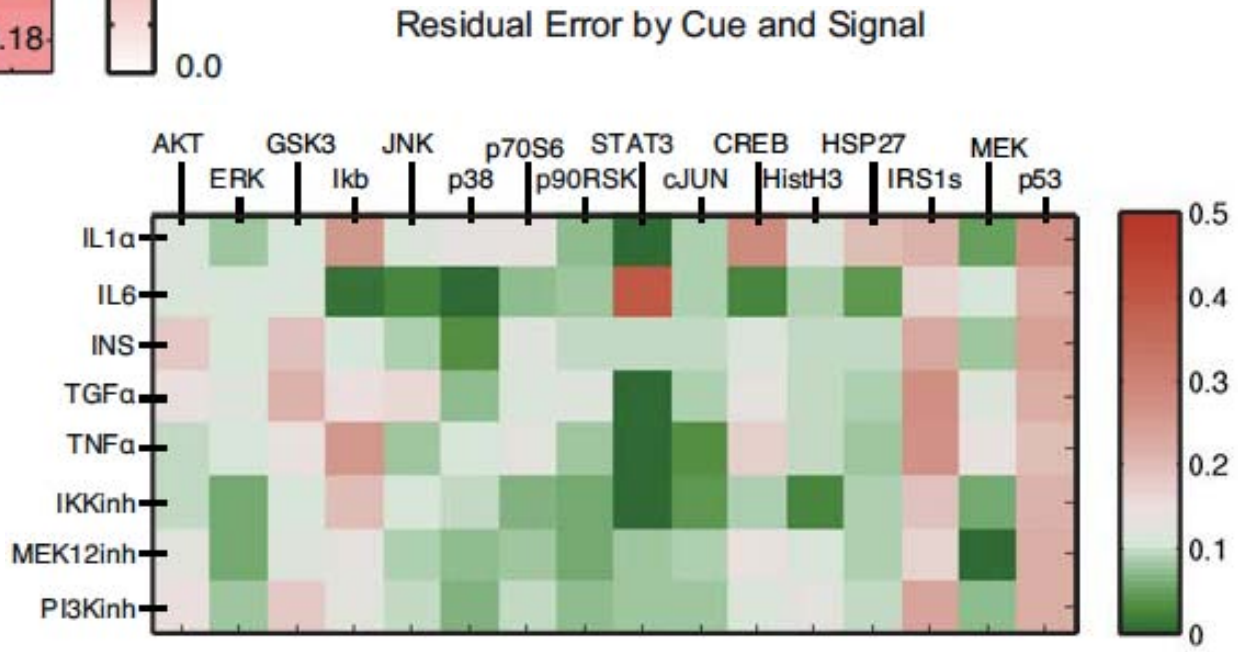


Source: Saez-Rodriguez, Julio, Leonidas G. Alexopoulos, et al. "Comparing Signaling Networks Between Normal and Transformed Hepatocytes Using Discrete Logical Models." *Cancer Research* 71, no. 16 (2011): 5400-11.

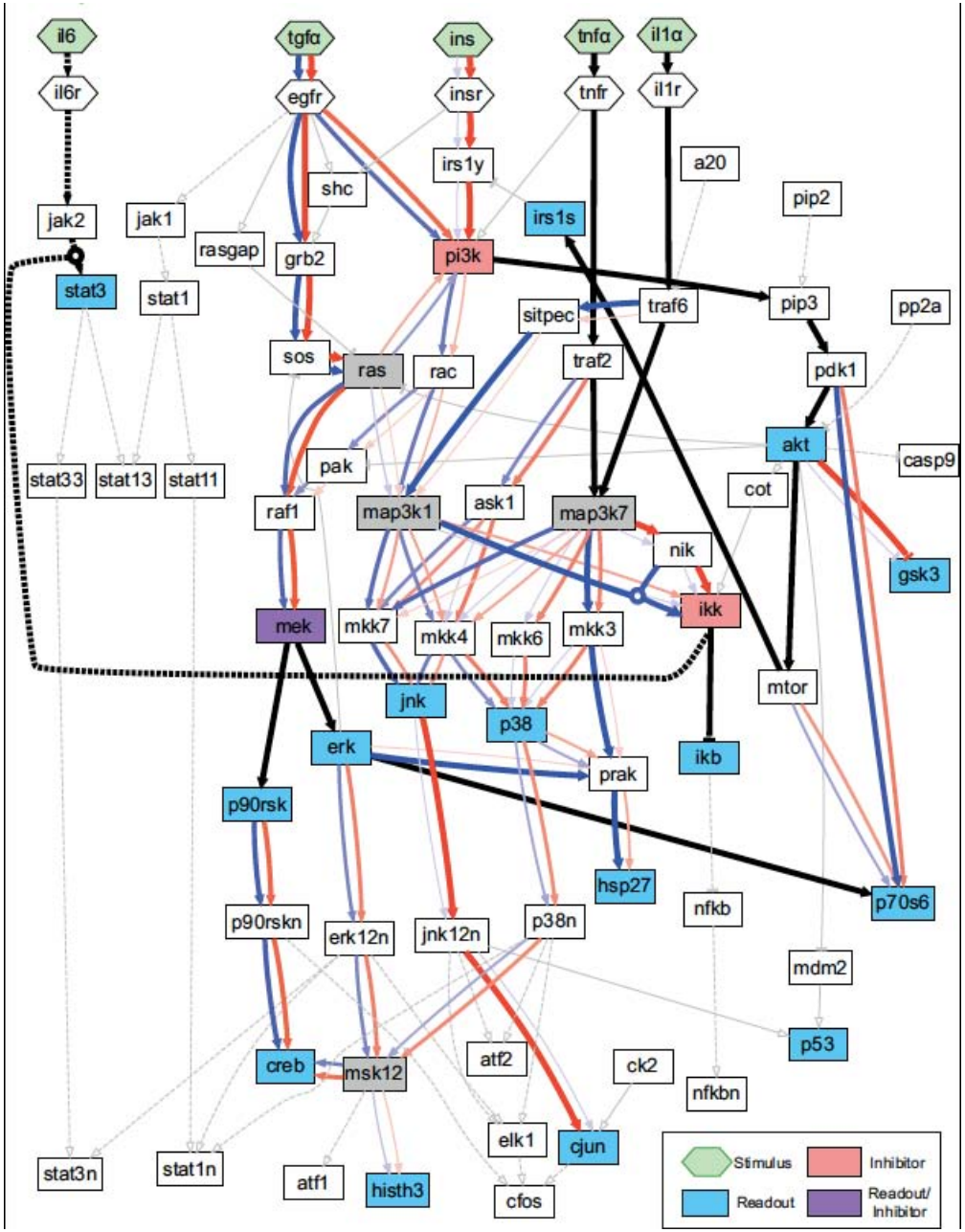
onstration of benefit of ype-specific models



Demonstration of capability to identify particular points inviting further study



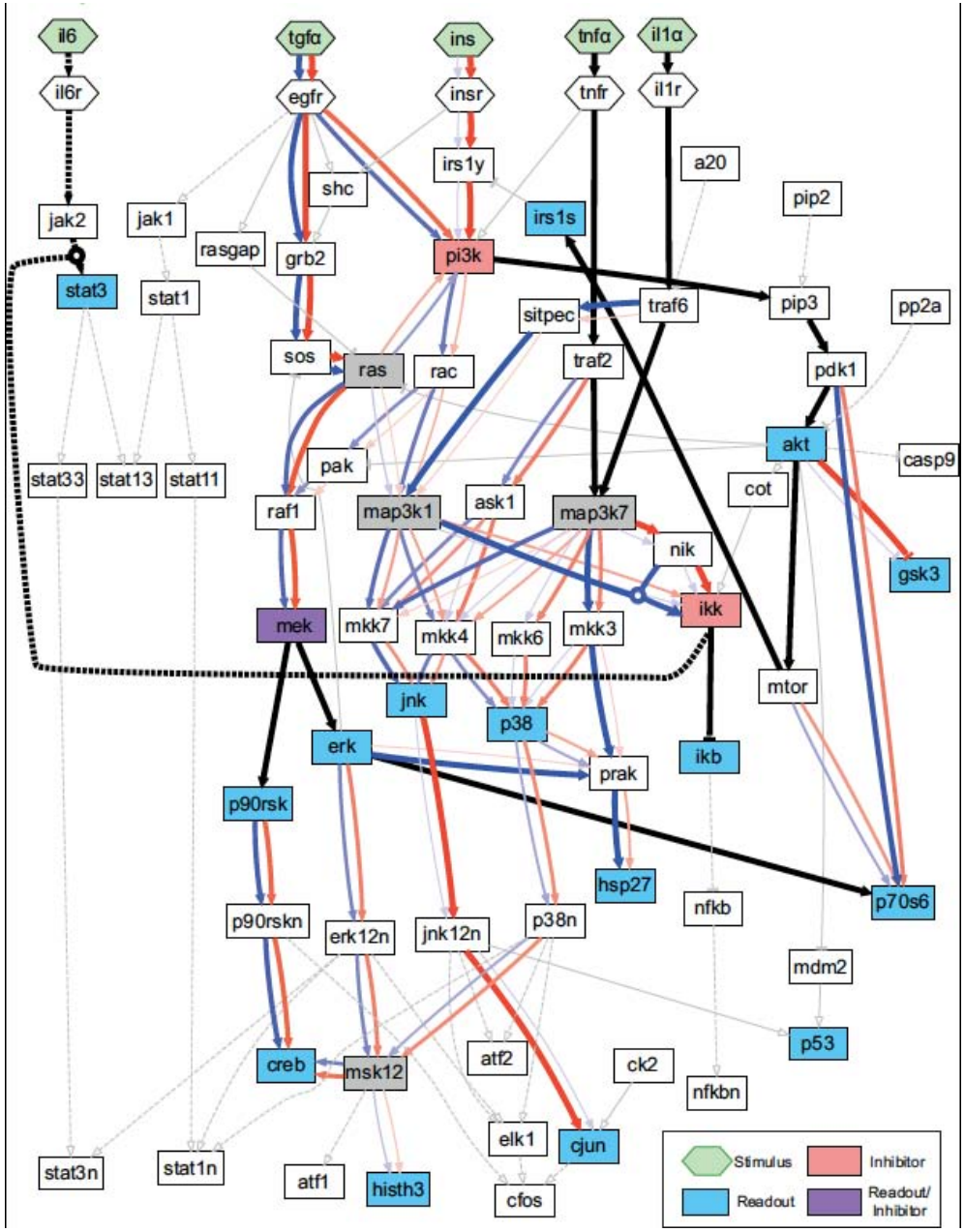
Source: Saez-Rodriguez, Julio, Leonidas G. Alexopoulos, et al. "Comparing Signaling Networks Between Normal and Transformed Hepatocytes Using Discrete Logical Models." *Cancer Research* 71, no. 16 (2011): 5400-11.



Best-Fit Boolean Logic Model Families for Primaries versus Lines

- **Arc width corresponds to proportion of best-fit models bearing it**
- **Black arcs – all models in both primaries and HCC lines**
- **Blue arcs – most or all primary models**
- **Red arcs – most or all HCC line models**
- **Gray arcs deleted from original scaffold**
- **Dashed arc added to account for especially recalcitrant data**

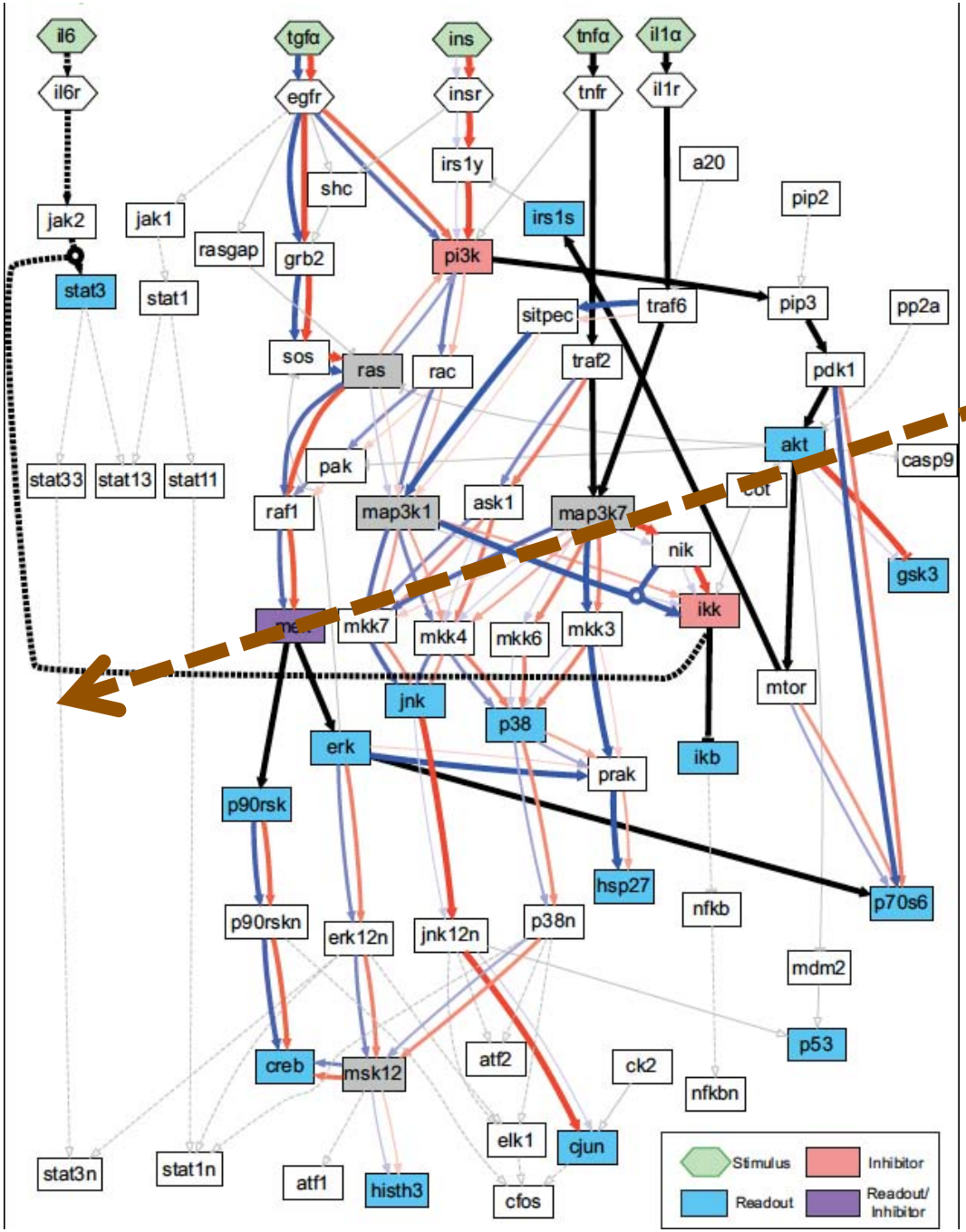
Source: Saez-Rodriguez, Julio, Leonidas G. Alexopoulos, et al. "Comparing Signaling Networks Between Normal and Transformed Hepatocytes Using Discrete Logical Models." *Cancer Research* 71, no. 16 (2011): 5400-11.



Best-Fit Boolean Logic Model Families for Primaries versus Lines

- **~90% of original scaffold interactions were found in at least one best-fit model across families for all cell types**
- **but only <10% were found both in most primary cell models and cell line models**
- **multiple pathways are identifiable as dysregulated from normal to tumor lines**

Source: Saez-Rodriguez, Julio, Leonidas G. Alexopoulos, et al. "Comparing Signaling Networks Between Normal and Transformed Hepatocytes Using Discrete Logical Models." Cancer Research 71, no. 16 (2011): 5400-11.



Model permits novel insights concerning drug actions

Dashed arc added to fit data generated in presence of IKK inhibitor TPCA1 – two potential explanations:

- **IKK activity suppresses STAT3 activity downstream of JAK2;**

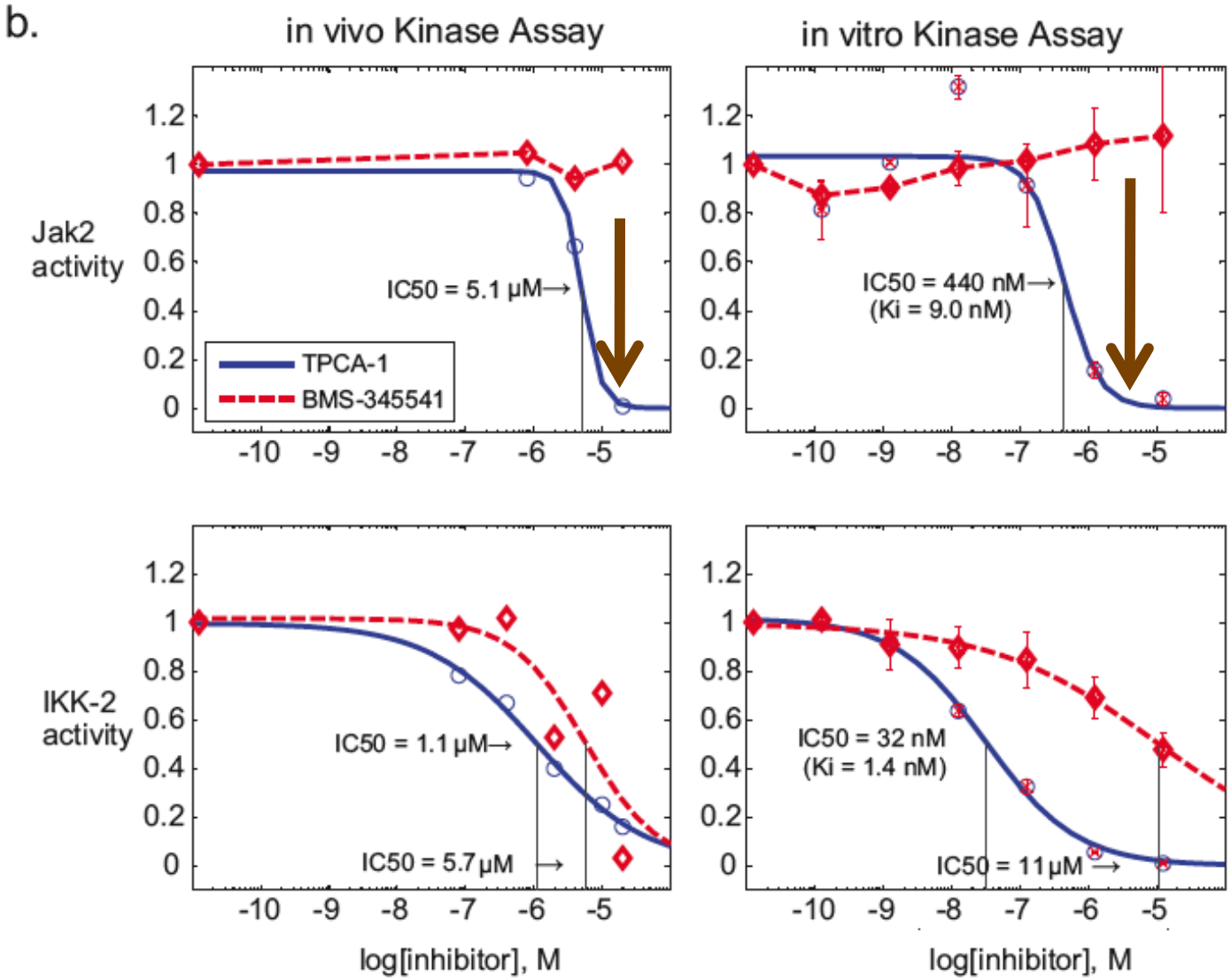
or

- **TPCA1 has off-target effect on JAK2**

Source: Saez-Rodriguez, Julio, Leonidas G. Alexopoulos, et al. "Comparing Signaling Networks Between Normal and Transformed Hepatocytes Using Discrete Logical Models." *Cancer Research* 71, no. 16 (2011): 5400-11.

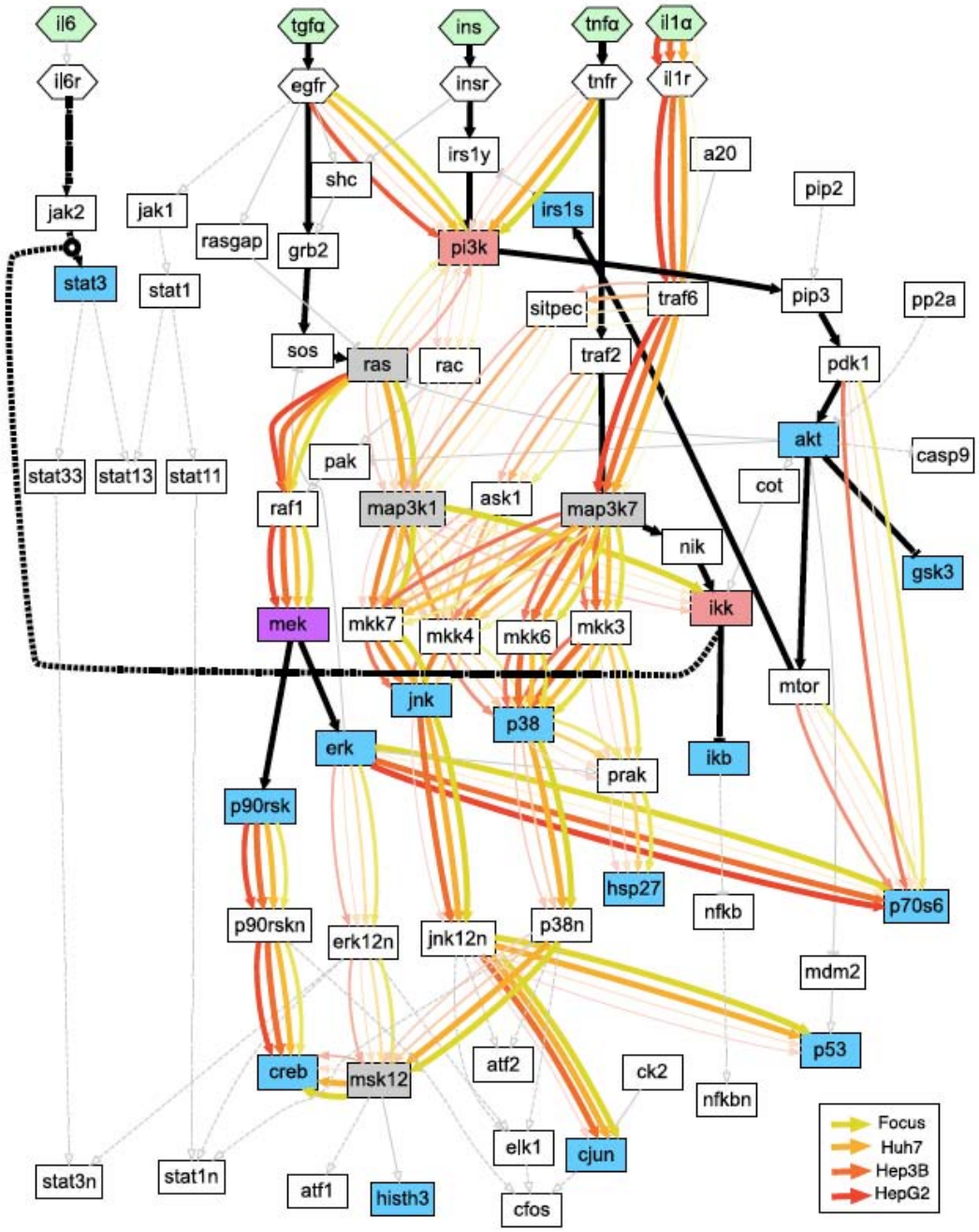
Experimental validation of model prediction that putative IKK inhibitor TPCA1 hits JAK2 as an off-target substrate

(whereas BMS-345541 does not)



…perhaps providing an explanation for why TPCA1 has been found to be more efficacious for airway inflammation treatment than other IKK inhibitors

Source: Saez-Rodriguez, Julio, Leonidas G. Alexopoulos, et al. "Comparing Signaling Networks Between Normal and Transformed Hepatocytes Using Discrete Logical Models." *Cancer Research* 71, no. 16 (2011): 5400-11.



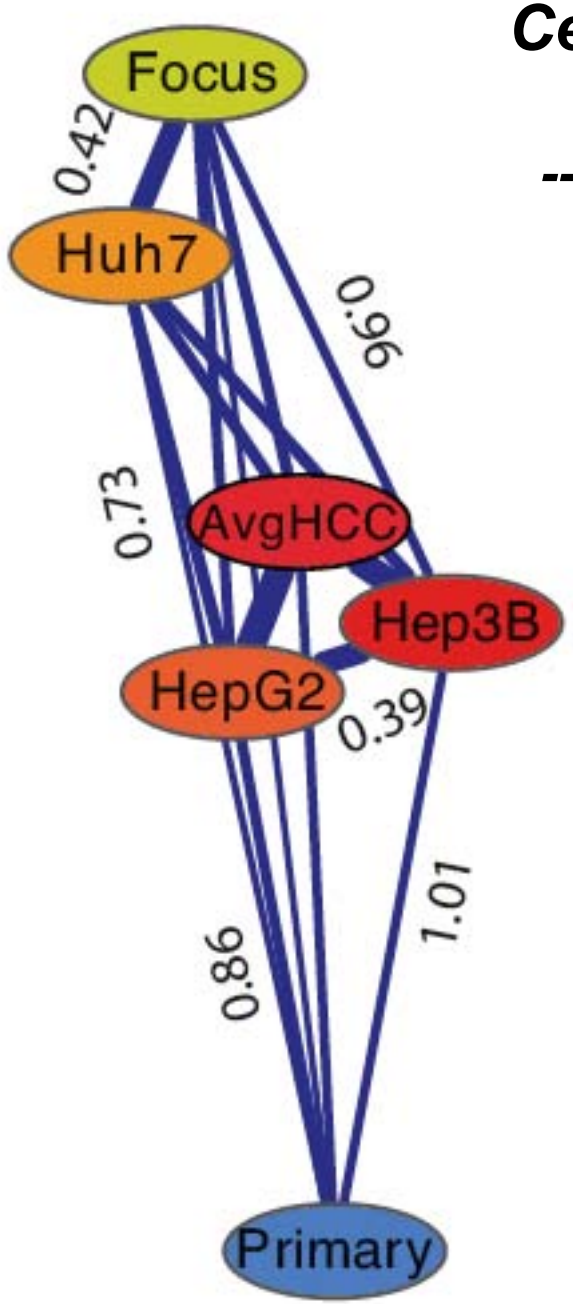
Best-Fit Boolean Logic Model Families -- comparison among HCC Lines

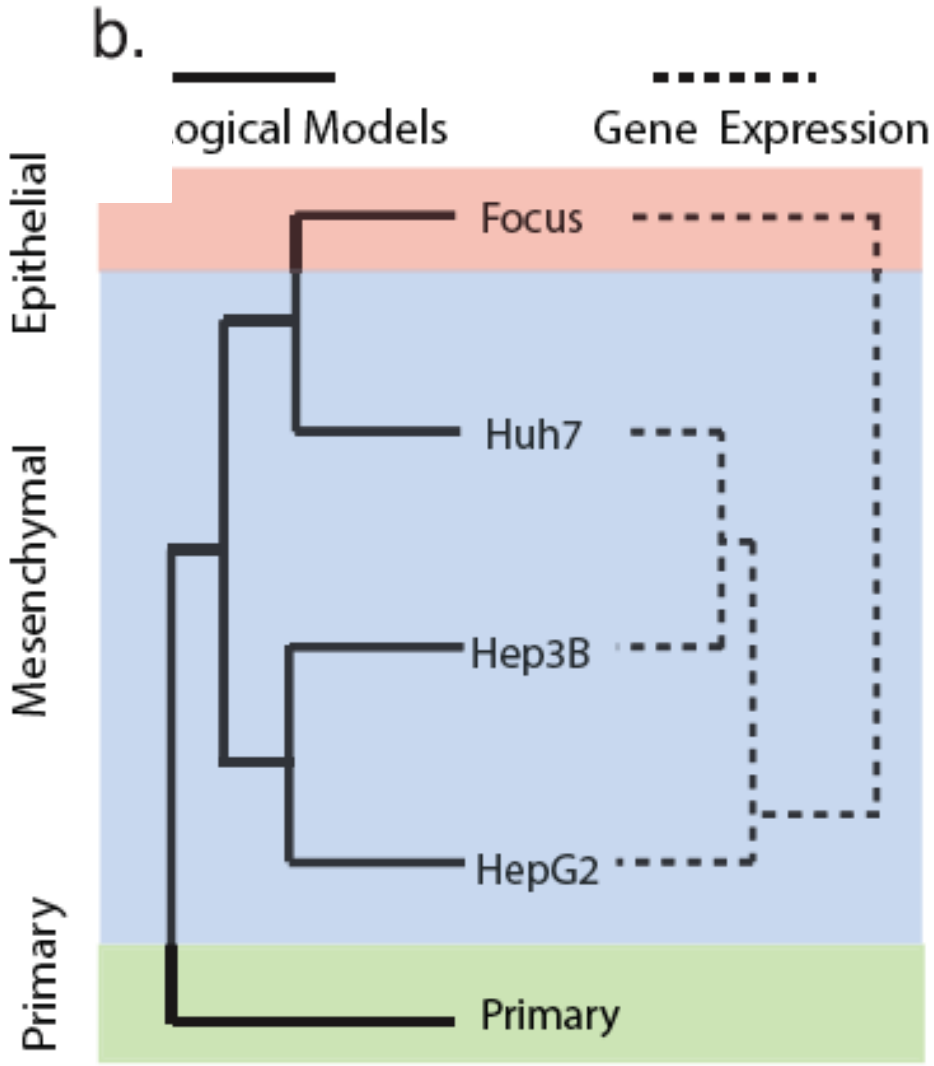
- **cell-type specificity of network operation is thus explicitly characterized – not only contrasting primaries to tumor lines but also disparities between different tumor lines**

Source: Saez-Rodriguez, Julio, Leonidas G. Alexopoulos, et al. "Comparing Signaling Networks Between Normal and Transformed Hepatocytes Using Discrete Logical Models." *Cancer Research* 71, no. 16 (2011): 5400-11.

Cell types can be quantitatively clustered with respect to common edges

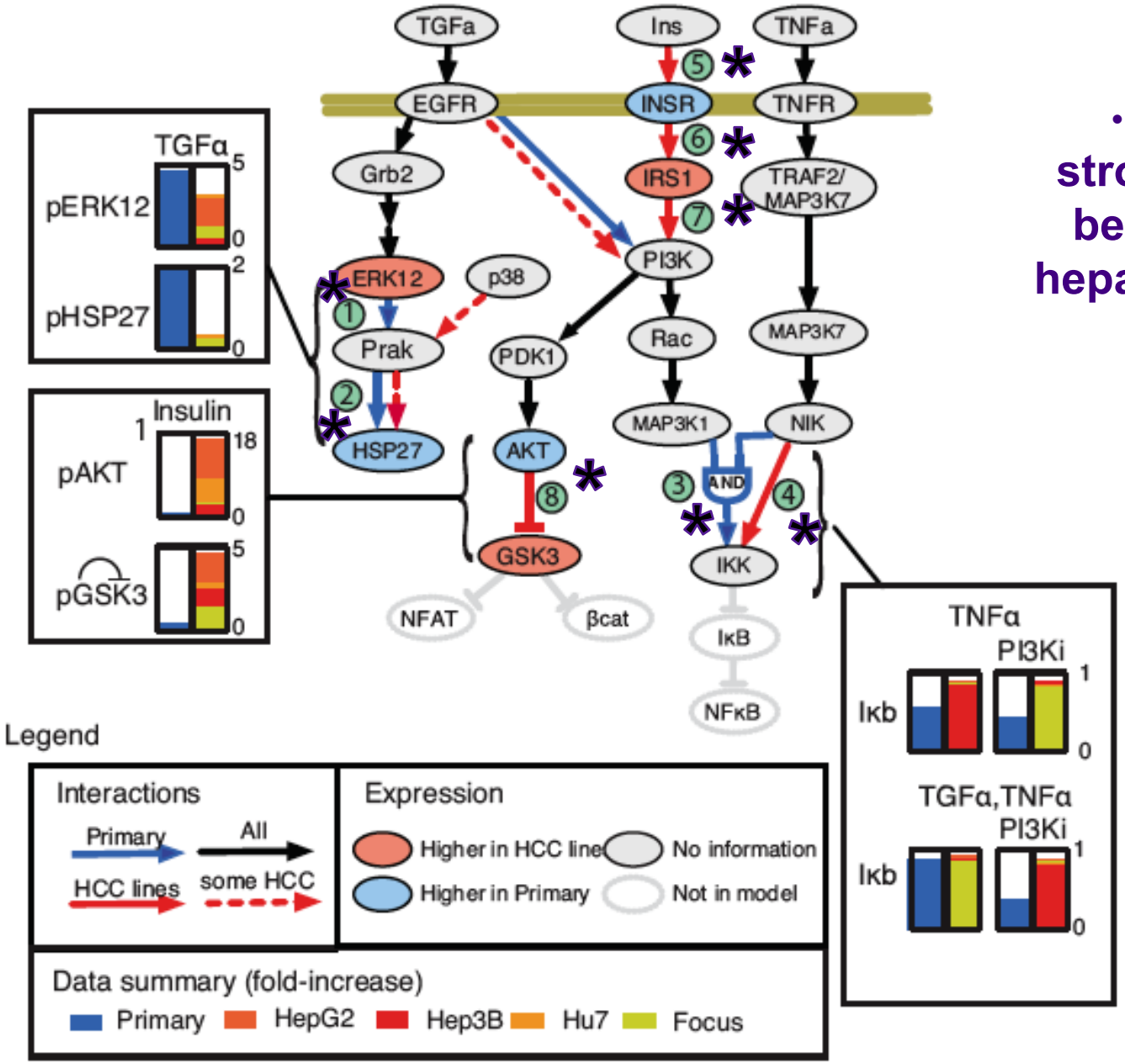
-- reasonable similarity to transcriptomic result





Source: Saez-Rodriguez, Julio, Leonidas G. Alexopoulos, et al. "Comparing Signaling Networks Between Normal and Transformed Hepatocytes Using Discrete Logical Models." *Cancer Research* 71, no. 16 (2011): 5400-11.

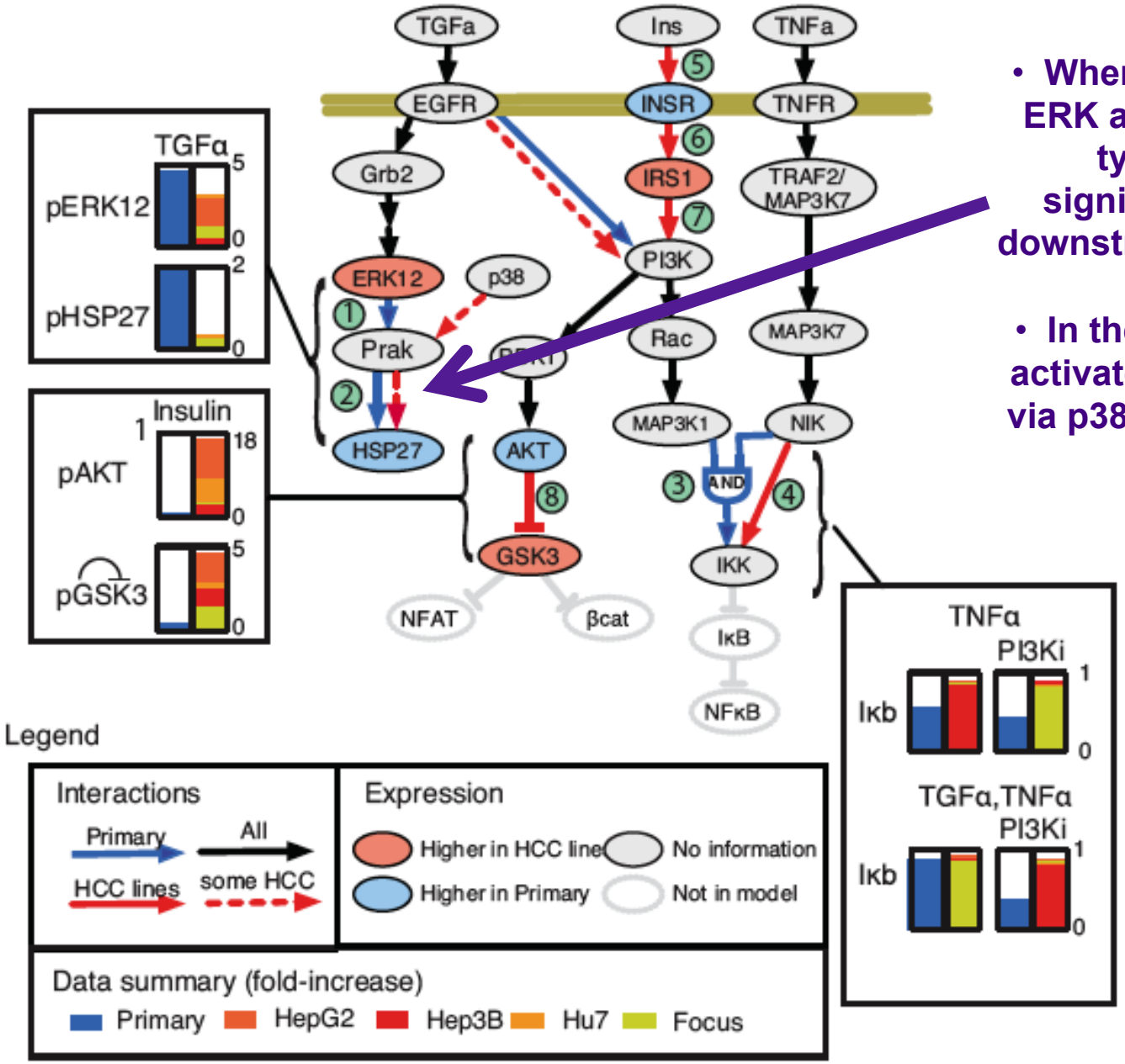
Detailed Primary-vs-Lines Comparison



- 8 edges are strongly disparate between primary hepatocytes and the HCC lines

Source: Saez-Rodriguez, Julio, Leonidas G. Alexopoulos, et al. "Comparing Signaling Networks Between Normal and Transformed Hepatocytes Using Discrete Logical Models." *Cancer Research* 71, no. 16 (2011): 5400-11.

Detailed Primary-vs-Lines Comparison – insights gained

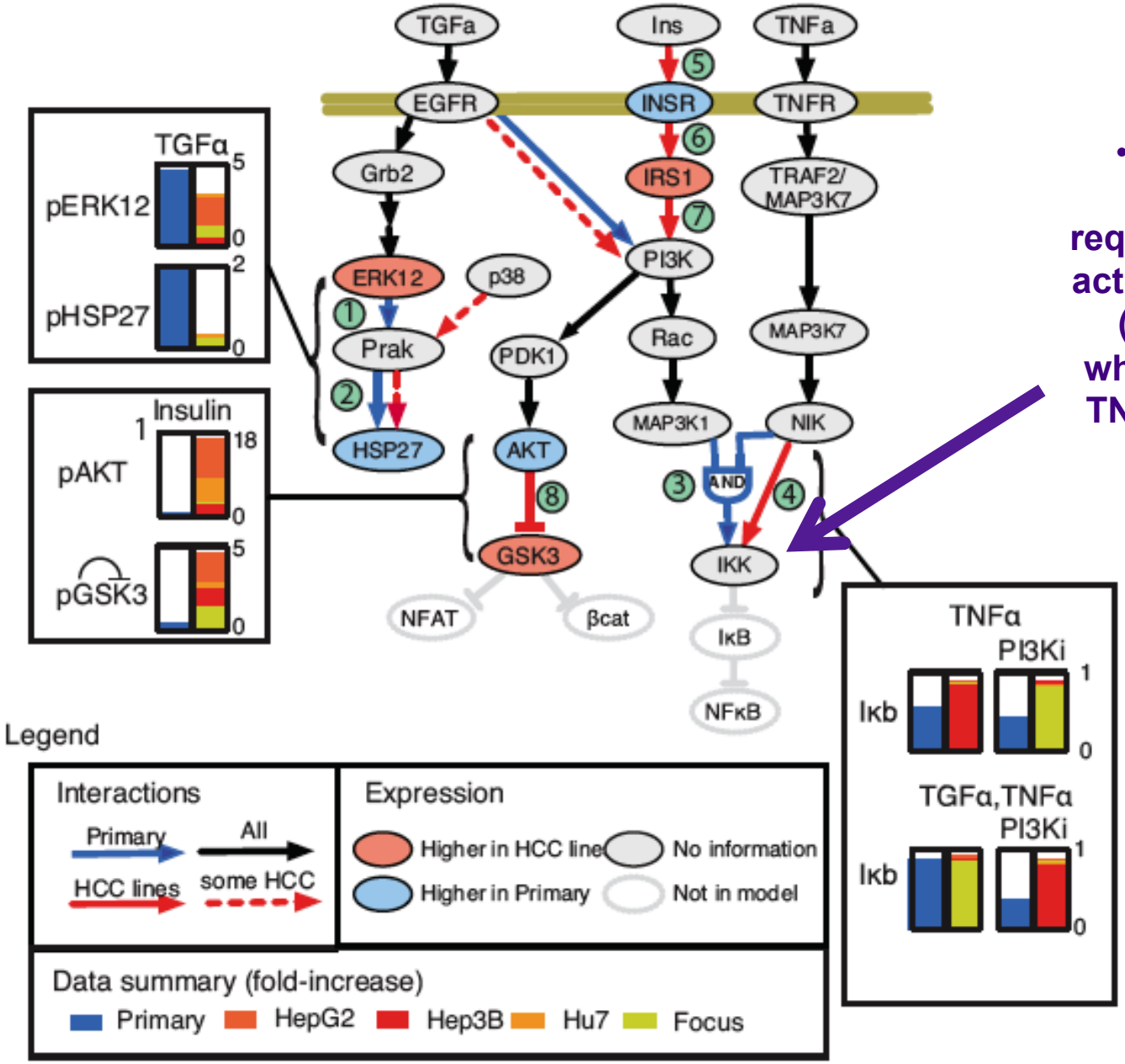


- **Whereas EGFR leads to ERK activation in all cell types, HSP27 is significantly activated downstream of ERK only in primaries**
- **In the lines, HSP27 was activated more mildly and via p38 instead of via ERK**

(Literature: HCC tumor progression is associated with decreased HSP27 activation -- despite HSP27 over-expression)

Source: Saez-Rodriguez, Julio, Leonidas G. Alexopoulos, et al. "Comparing Signaling Networks Between Normal and Transformed Hepatocytes Using Discrete Logical Models." *Cancer Research* 71, no. 16 (2011): 5400-11.

Detailed Primary-vs-Lines Comparison – insights gained

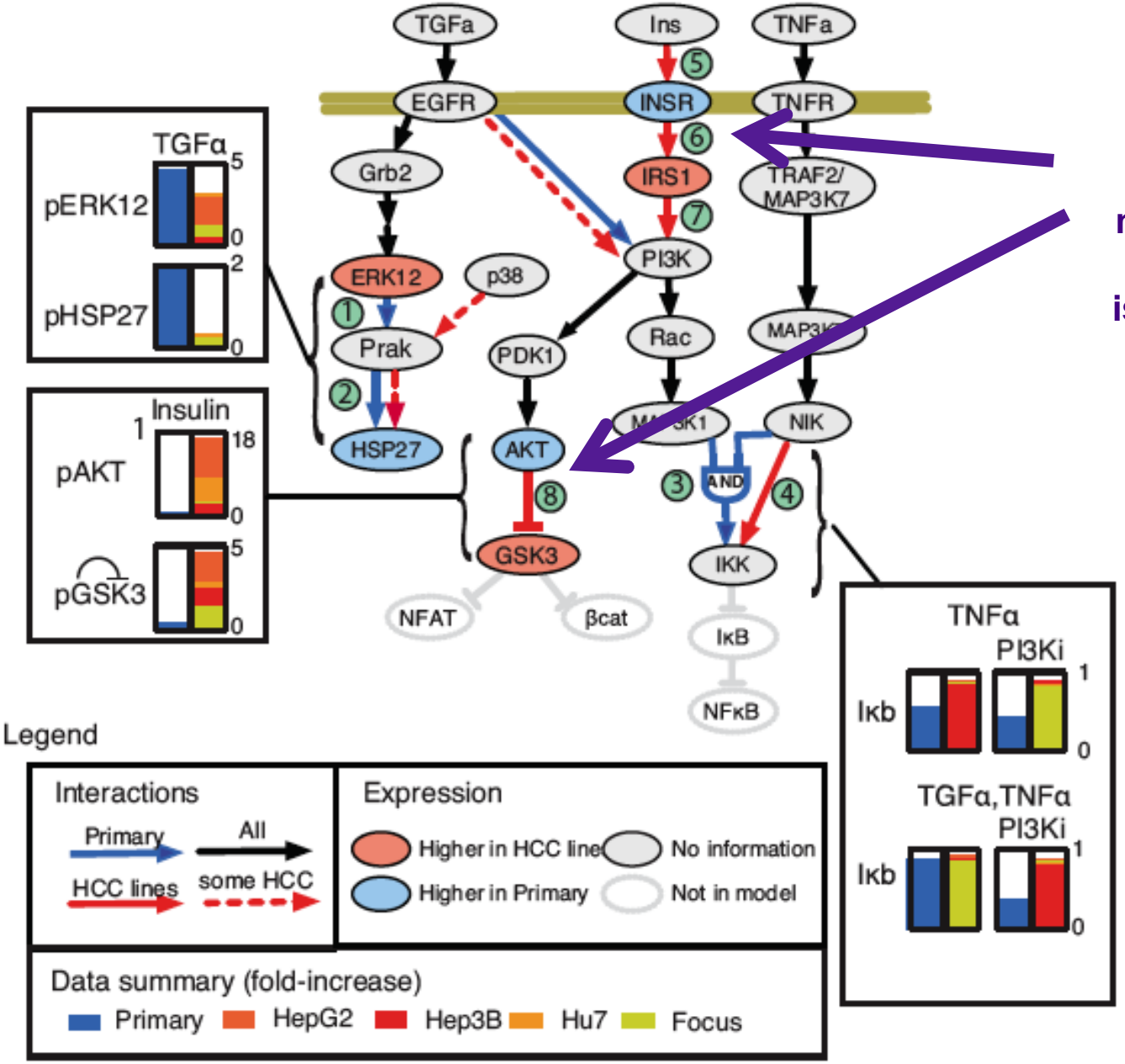


- **In primaries Ikb phosphorylation requires TNFa-NIK and activation of PI3K-JNK (via TGFa or Ins), whereas in lines only TNFa-NIK is required**

(Literature: HCC tumor progression is associated with looser control over NFkB-mediated survival signals)

Source: Saez-Rodriguez, Julio, Leonidas G. Alexopoulos, et al. "Comparing Signaling Networks Between Normal and Transformed Hepatocytes Using Discrete Logical Models." *Cancer Research* 71, no. 16 (2011): 5400-11.

Detailed Primary-vs-Lines Comparison – insights gained



- **GSK3 phosphorylation by Akt (leading to nuclear activation of pro-mitotic factors) is induced by Insulin in lines but not in primaries**

(Literature: IRS1 is over-expressed in HCC, potentially shifting Insulin-induced signaling from IRS2-mediated metabolism to proliferation)

Source: Saez-Rodriguez, Julio, Leonidas G. Alexopoulos, et al. "Comparing Signaling Networks Between Normal and Transformed Hepatocytes Using Discrete Logical Models." *Cancer Research* 71, no. 16 (2011): 5400-11.

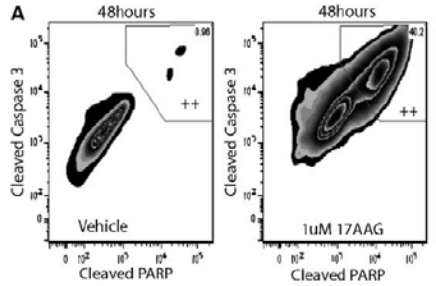
These same three pathways have been implicated in combination kinase therapy for HCC

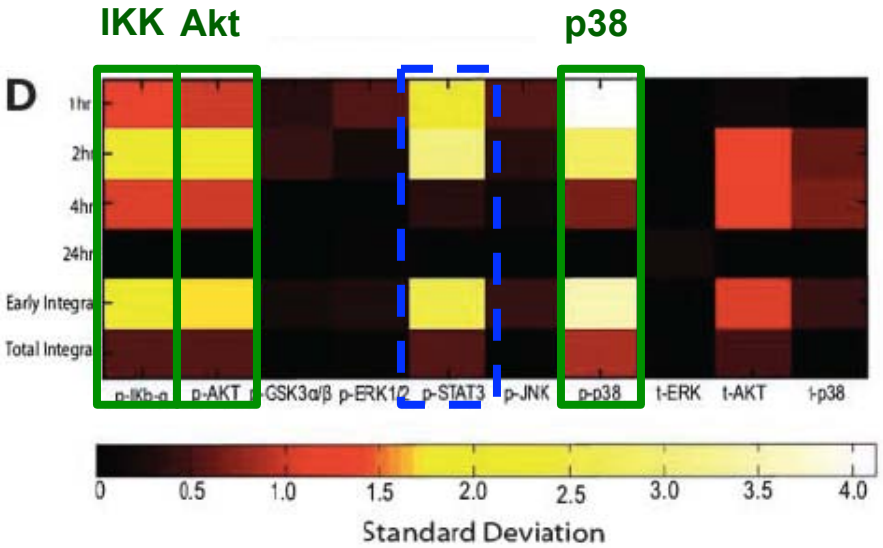
Three-kinase inhibitor combination recreates multipathway effects of a geldanamycin analogue on hepatocellular carcinoma cell death

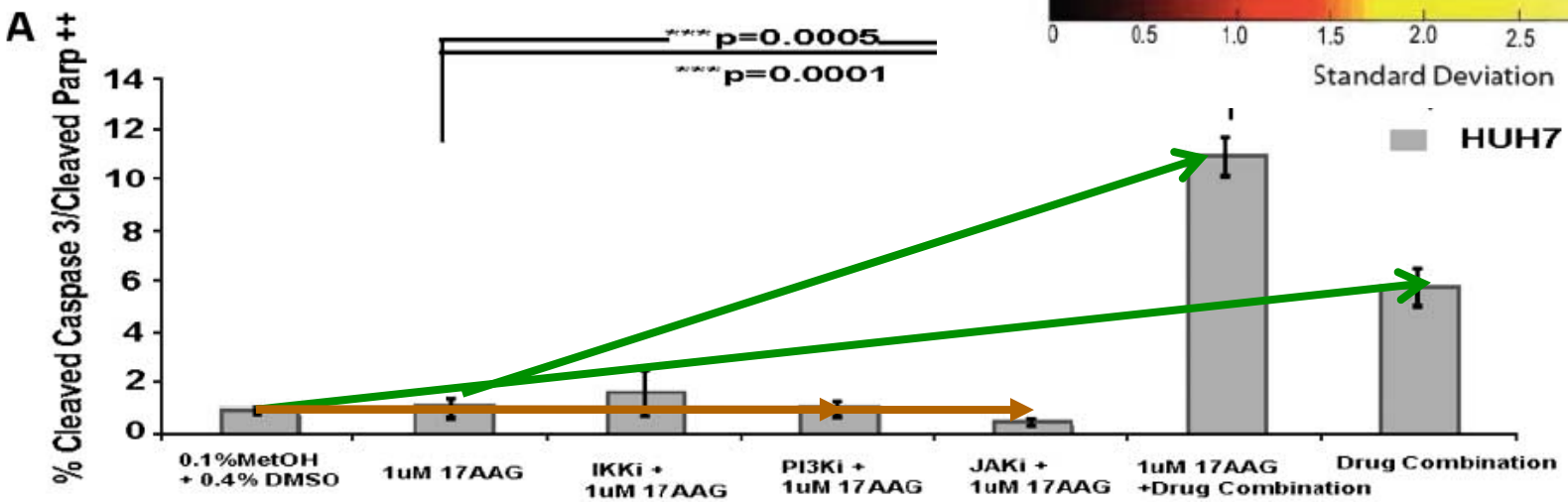
Justin R. Pritchard,[1] Benjamin D. Cosgrove,[2]
Michael T. Hemann,[1] Linda G. Griffith,[2]
Jack R. Wands,[3] and Douglas A. Lauffenburger[1,2]

Departments of [1]Biology and [2]Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts and [3]The Warren Alpert Medical School of Brown University, Providence, Rhode Island

[Mol Cancer Ther 2009;8(8):OF1–10]







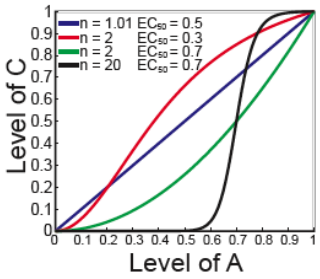
Source: Pritchard, Justin R., Benjamin D. Cosgrove, et al. "Three-kinase Inhibitor Combination Recreates Multipathway Effects of a Geldanamycin Analogue on Hepatocellular Carcinoma Cell Death." *Molecular Cancer Therapeutics* 8, no. 8 (2009): 2183-92.

One-input interactions: How does C depend on A?

where n is the hill coefficient, k specifies the EC_50 for each gate and a and c are the quantitative levels of their respective species (A and C)

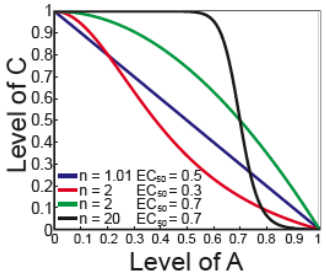
A → C

$$c = (k^n + 1)\frac{a^n}{k^n + a^n}$$



A ⊣ C

$$c = 1 - (k^n + 1)\frac{a^n}{k^n + a^n}$$



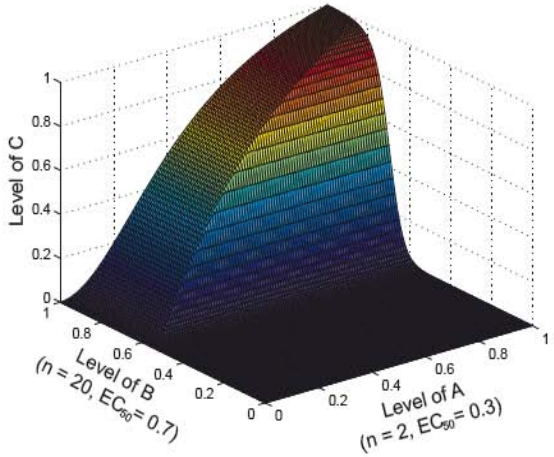
'Constrained Fuzzy Logic' Framework allows Analog Model rather than Digital

Two-input interactions: How is C evaluated when both A and B affect it?

where subscript 1 and 2 indicate the gate-specific parameters of the A-to-C and B-to C interactions, respectively.

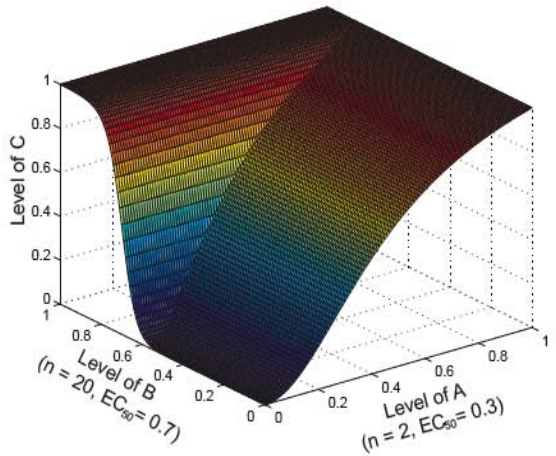
A AND B → C

$$c = \min\left((k_1^{n_1} + 1)\frac{a^{n_1}}{k_1^{n_1} + a^{n_1}}, (k_2^{n_2} + 1)\frac{b^{n_2}}{k_2^{n_2} + b^{n_2}}\right)$$



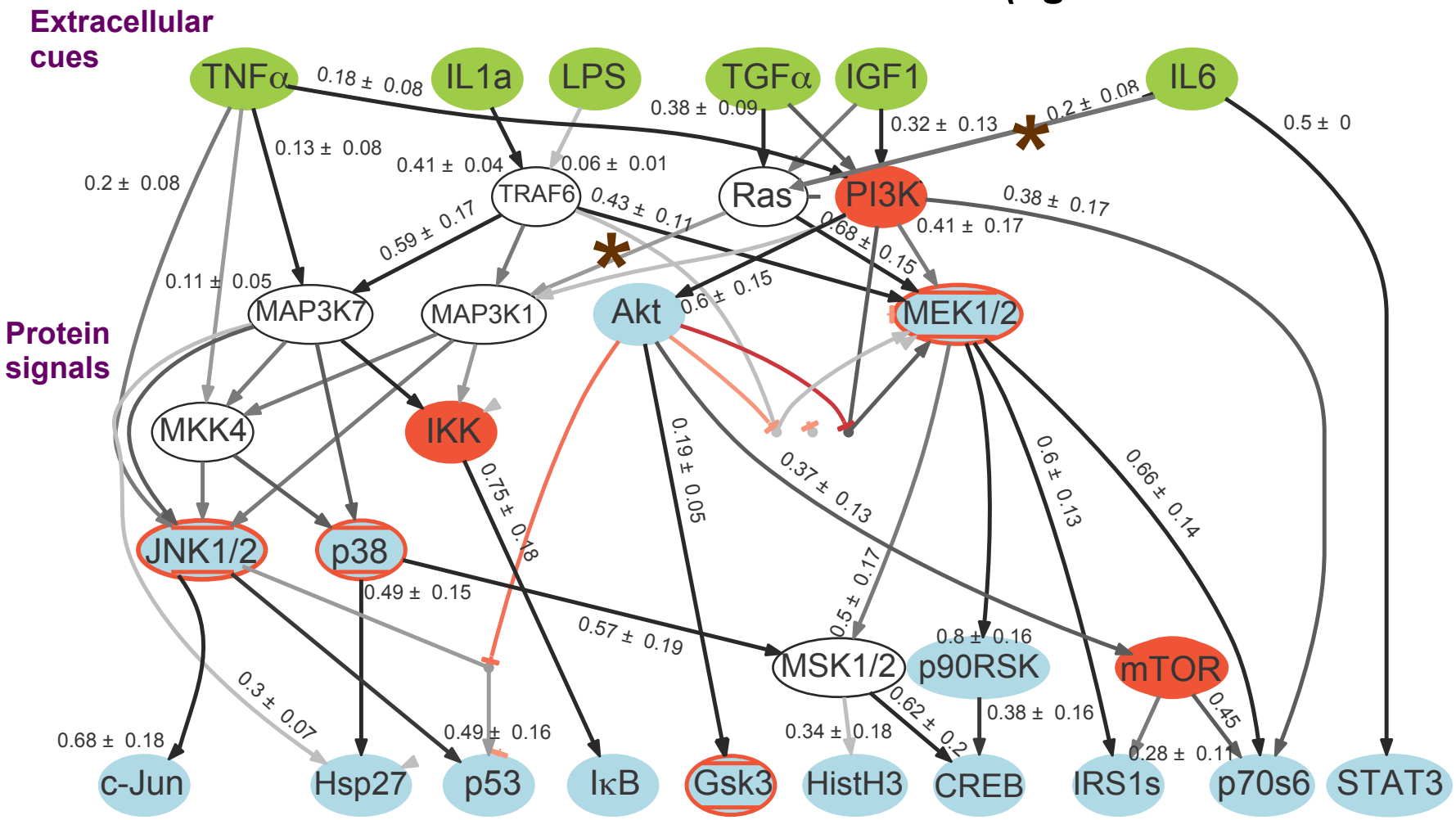
A OR B → C

$$c = \max\left((k_1^{n_1} + 1)\frac{a^{n_1}}{k_1^{n_1} + a^{n_1}}, (k_2^{n_2} + 1)\frac{b^{n_2}}{k_2^{n_2} + b^{n_2}}\right)$$



Source: Morris, Melody K., Julio Saez-Rodriguez, et al. "Training Signaling Pathway Maps to Biochemical Data with Constrained Fuzzy Logic: Quantitative Analysis of Liver Cell Responses to Inflammatory Stimuli." *PLoS Computational Biology* 7, no. 3 (2011): e1001099.

HepG2 Constrained Fuzzy Logic Network Model
(again consensus family)



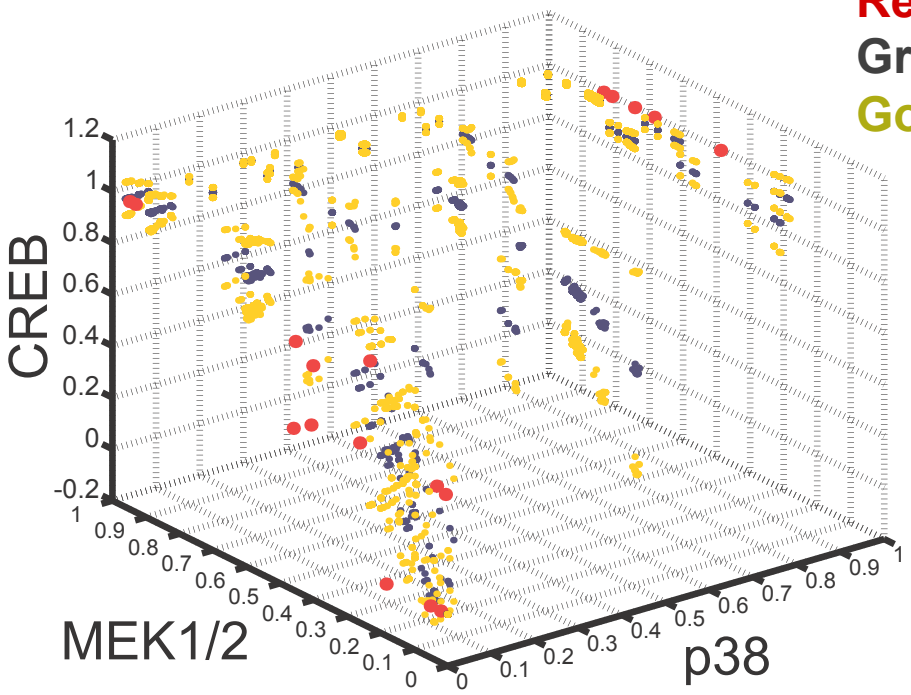
Source: Morris, Melody K., Julio Saez-Rodriguez, et al. "Training Signaling Pathway Maps to Biochemical Data with Constrained Fuzzy Logic: Quantitative Analysis of Liver Cell Responses to Inflammatory Stimuli." *PLoS Computational Biology* 7, no. 3 (2011): e1001099.

Intensity of arc = likelihood of connection

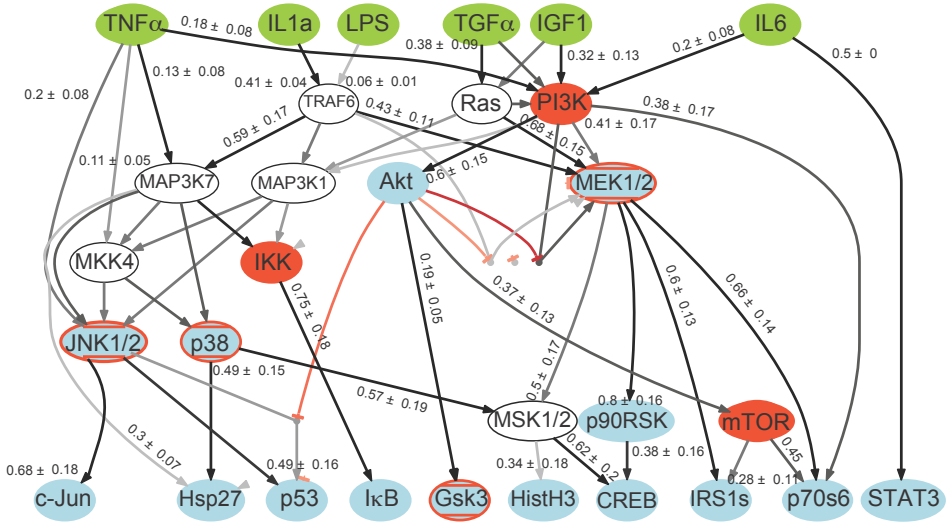
Numerical descriptor = upstream-downstream effect strength

**** = new arcs not identified by Boolean model***

Example Results for Quantitative Cell Circuit Logic
-- downstream 'child' node versus upstream 'parent' nodes



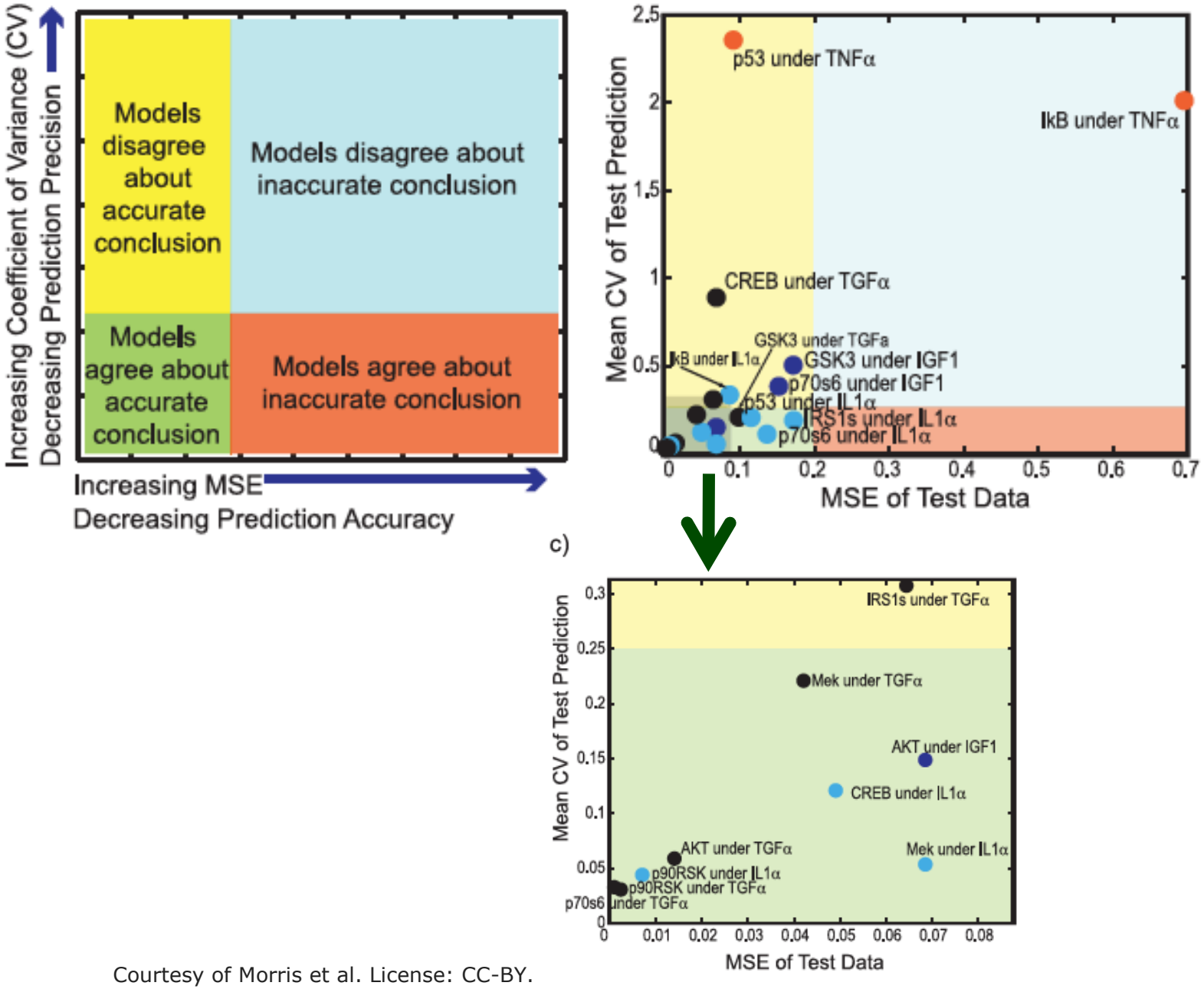
Red points: experimental values
Gray points: averaged-model predictions
Gold points: individual model predictions



Source: Morris, Melody K., Julio Saez-Rodriguez, et al. "Training Signaling Pathway Maps to Biochemical Data with Constrained Fuzzy Logic: Quantitative Analysis of Liver Cell Responses toInflammatory Stimuli." *PLoS Computational Biology* 7, no. 3 (2011): e1001099.

New test data fell within one standard deviation of predictions across all conditions

Model family precision generally presages accuracy



Source: Morris, Melody K., Julio Saez-Rodriguez, et al. "Training Signaling Pathway Maps to Biochemical Data with Constrained Fuzzy Logic: Quantitative Analysis of Liver Cell Responses to Inflammatory Stimuli." *PLoS Computational Biology* 7, no. 3 (2011): e1001099.

7.91J / 20.490J / 20.390J / 7.36J / 6.802J / 6.874J / HST.506J Foundations of Computational and Systems Biology
Spring 2014