CellNOpt:
modelling prior knowledge networks trained to experimental data

Julio Saez-Rodriguez

European Bioinformatics Institute

EMBL-EBI

Hinxton (Cambridge) UK

www.ebi.ac.uk/saezrodriguez

Logic modelling to link protein signalling networks with functional analysis of signal transduction

- **CellNOpt**: a flexible pipeline to model protein signalling networks
- (1) Converts protein signalling network into logic model
  - (Unless known) create all possible logic gates (AND/OR) compatible with the network
- (2) Find the combination of logic gates (i.e. the model) that best
  - Balance fit to data with model size
    describes the experimental data

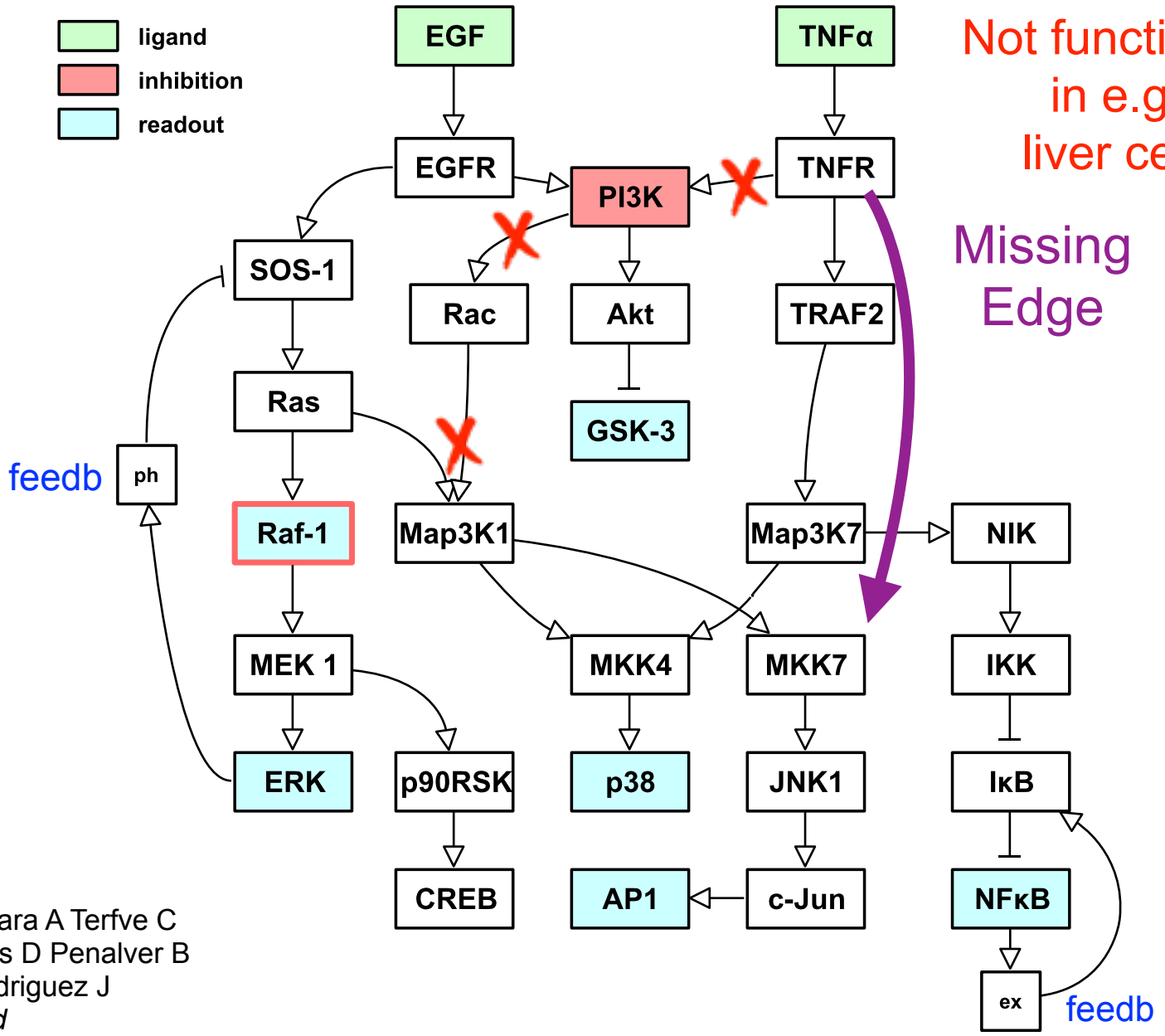
$$\theta = \theta_f + \alpha \cdot \theta_S$$

Bioconductor & Matlab,
available at www.ebi.ac.uk/saezrodriguez/software.html

Morris MK, Melas I, Saez-Rodriguez J, *Methods Mol. Biol, in press*

Saez-Rodriguez J, Alexopoulos LG, Epperlein J, Samaga R, Lauffenburger DA, Klamt S, Sorger PK *Mol Sys Bio* 5:331,2009

A Toy model

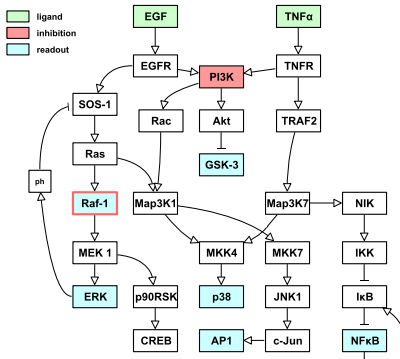


Not functional in e.g. liver cell

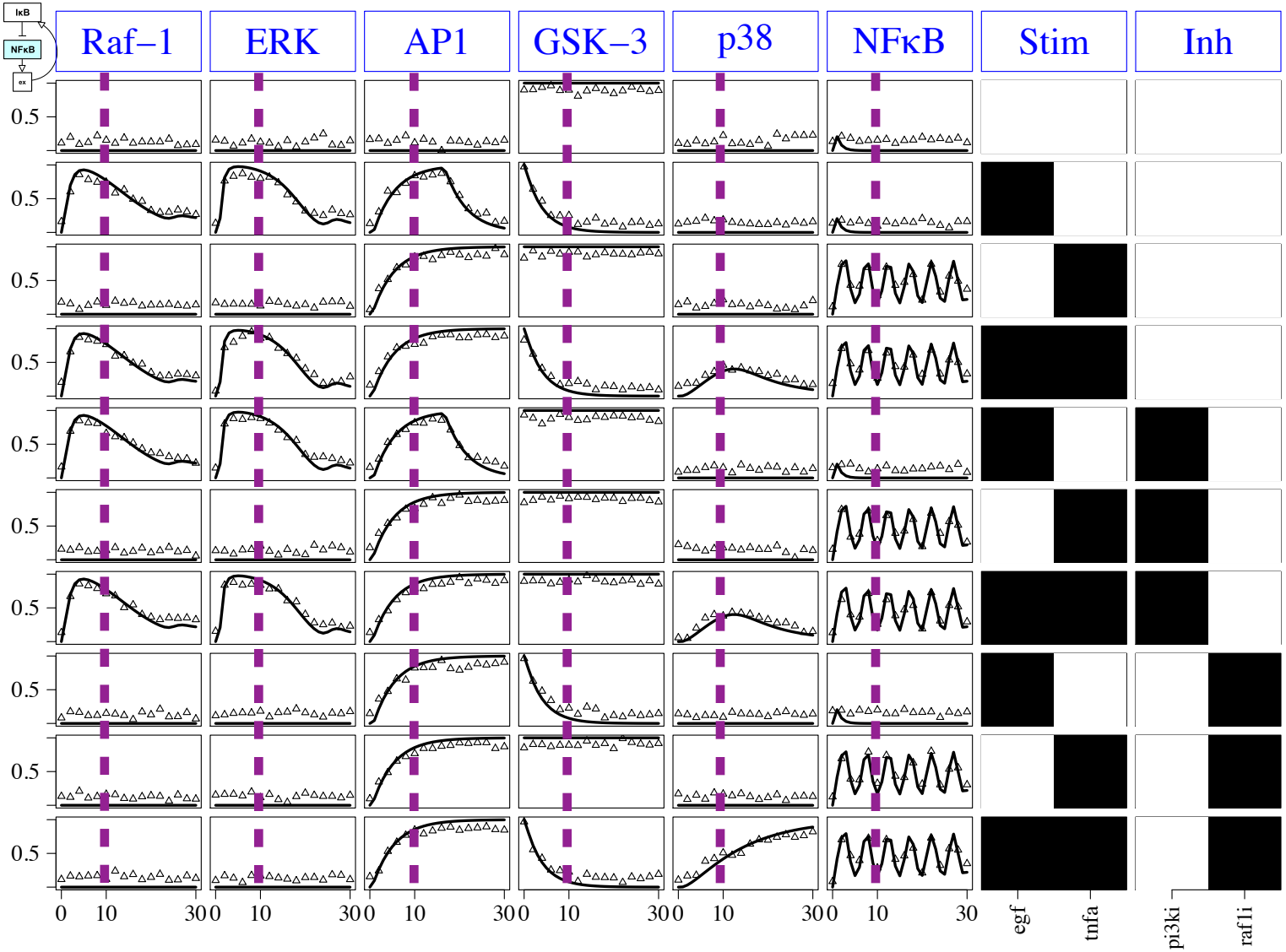
Missing Edge

MacNamara A Terfve C Henriques D Penalver B Saez-Rodriguez J *submitted*

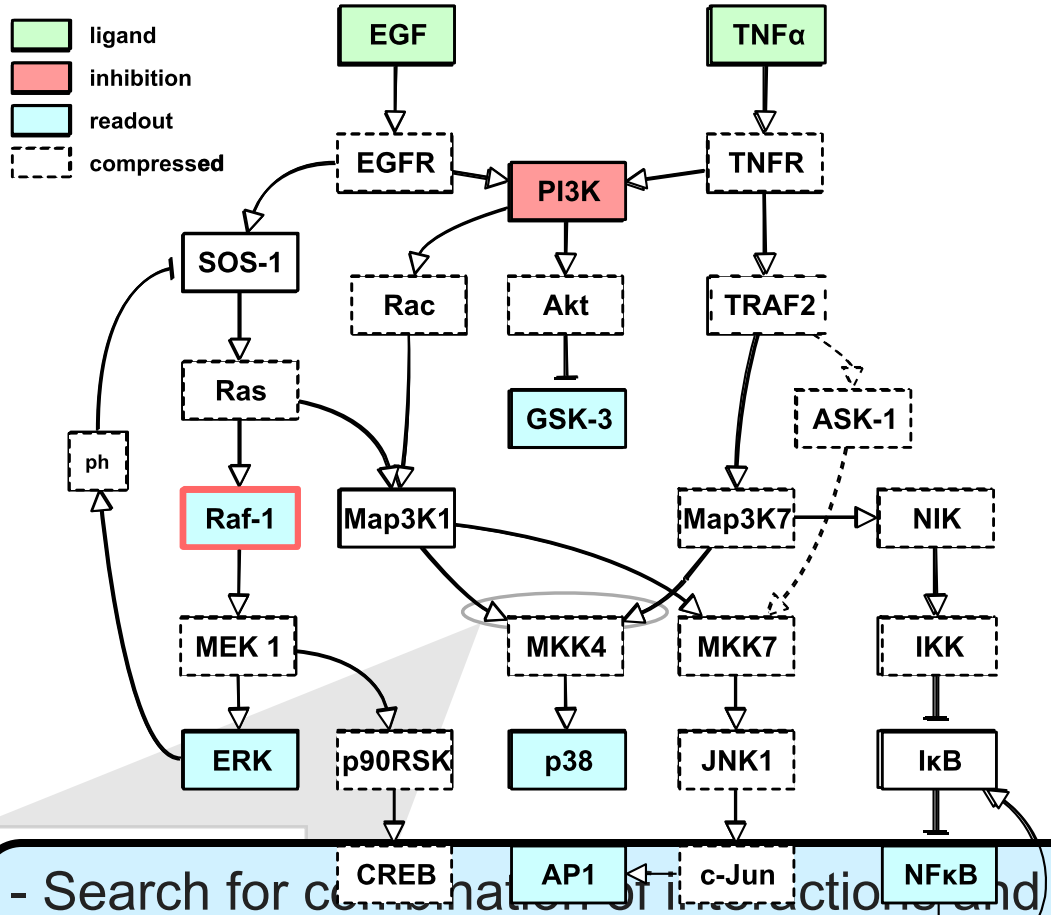
The ‘real’ data



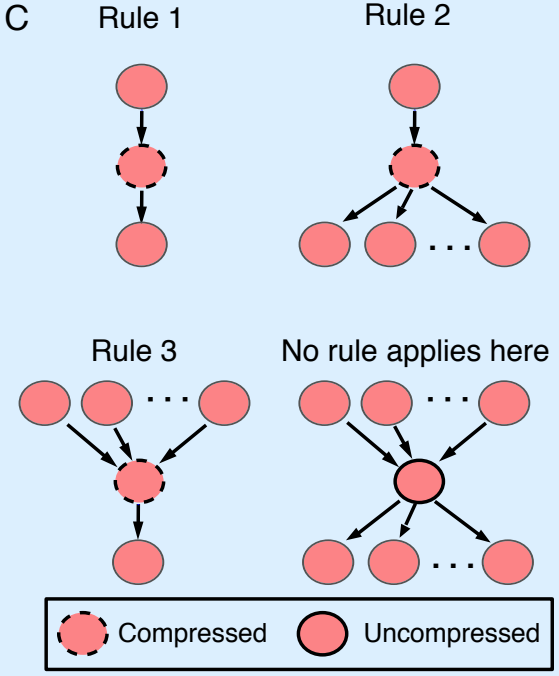
How to pick right time to measure?



Model preprocessing: compression and expansion of gates



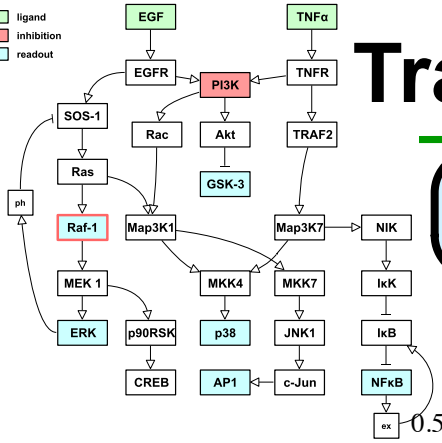
Model compression



Saez-Rodriguez J, et al., *Mol. Syst. Biol,* 2009

- Search for combination of interactions and gates that best describe the data using a pseudo-state simulation (signal propagation);
- Search is an optimization problem that ca be solved
  (1) Enumeration (2) Heuristic (e.g. genetic algorithm)
  (3) Integer Linear Programming (Mitsos et al . Plos Comp Bio 2009)

Training to data recovers structure



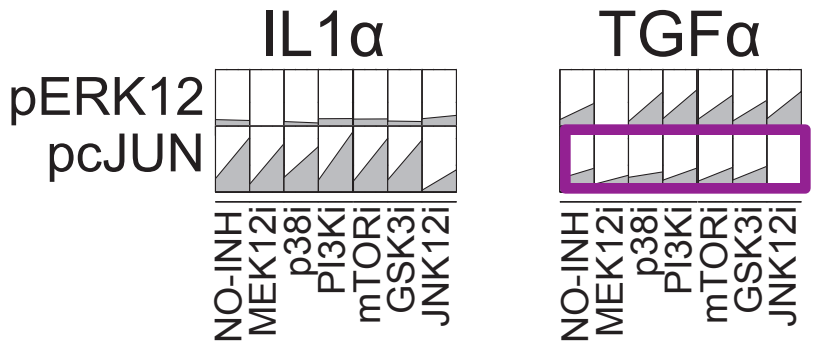
Identifies active/non-active links (except feedbacks)

Can not explain data due to missing links

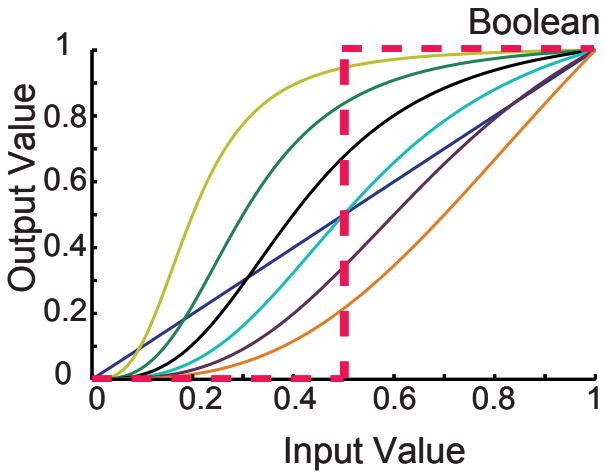
MacNamara A Terfve C Henriques D Penalver B Saez-Rodriguez J *submitted*

Constrained Fuzzy Logic can handle quantitative differences

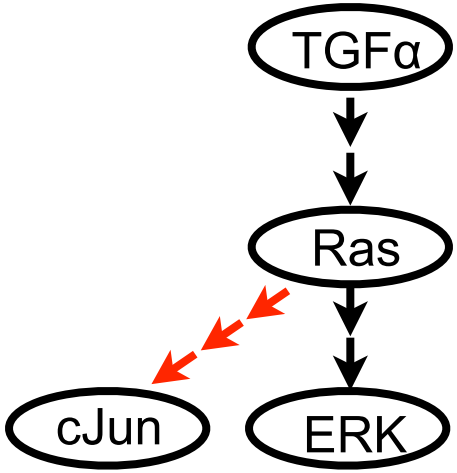
- Boolean modeling can **not** describe **quantitative** aspects (e.g. intermediate activation)



- Fuzzy logic can model quantitative signaling data (Aldridge et al. *Plos Comp. Bio.* 2009)

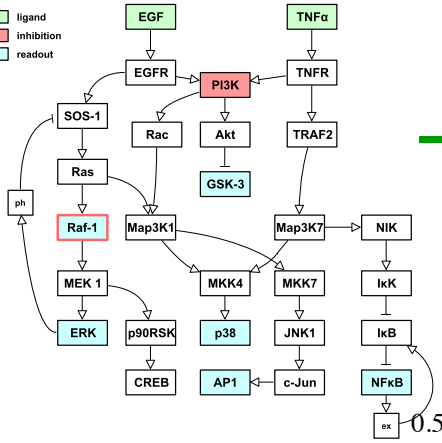


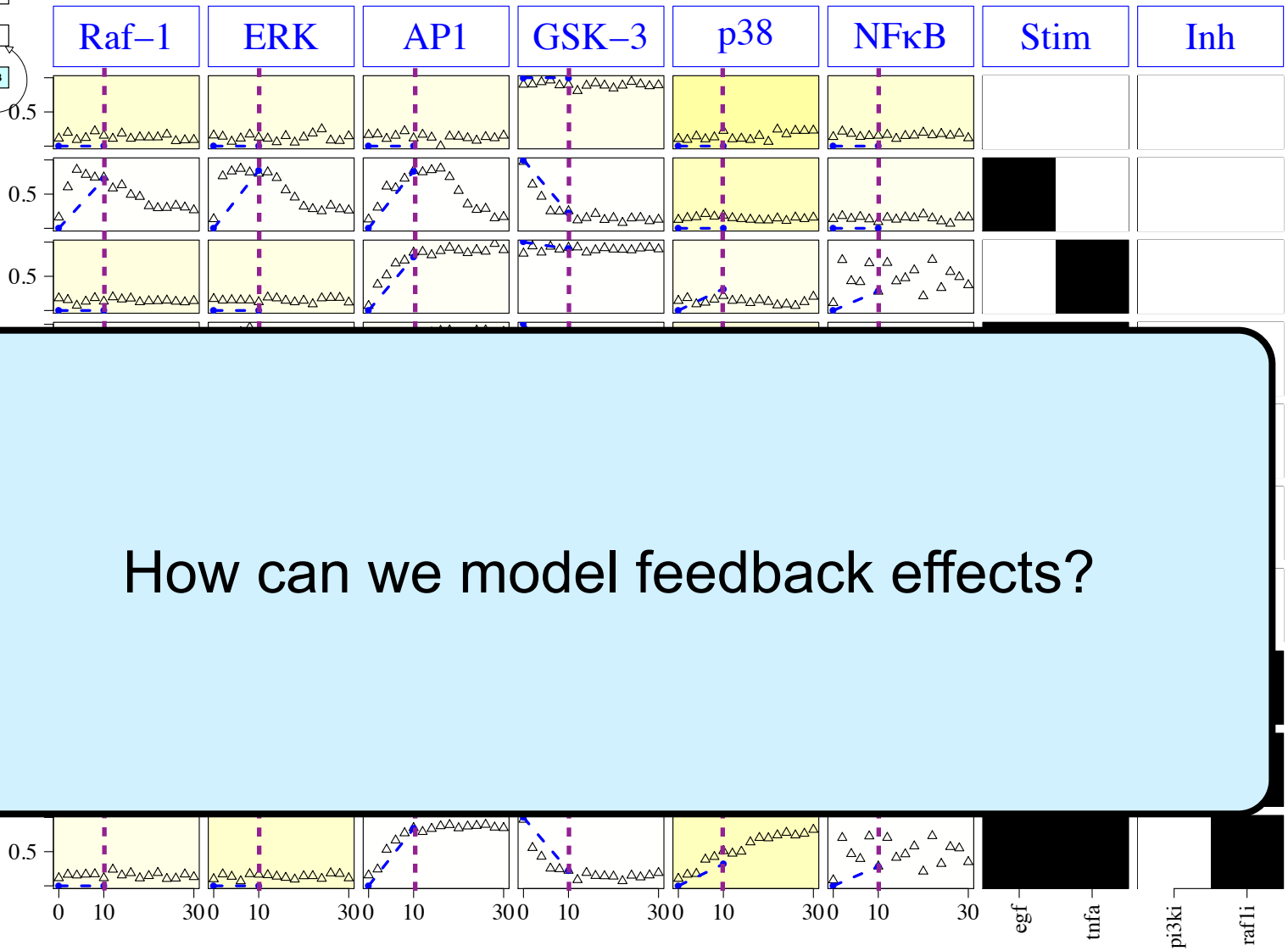
- Development of **Constrained Fuzzy Logic** & implementation within *CellNOpt*



Morris MK, Saez-Rodriguez J, Clarke D, Sorger PK, Lauffenburger DA, PloS Comp Bio 2011.

fit of fuzzy toy model

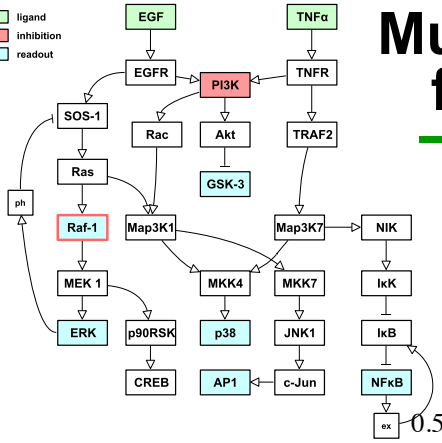


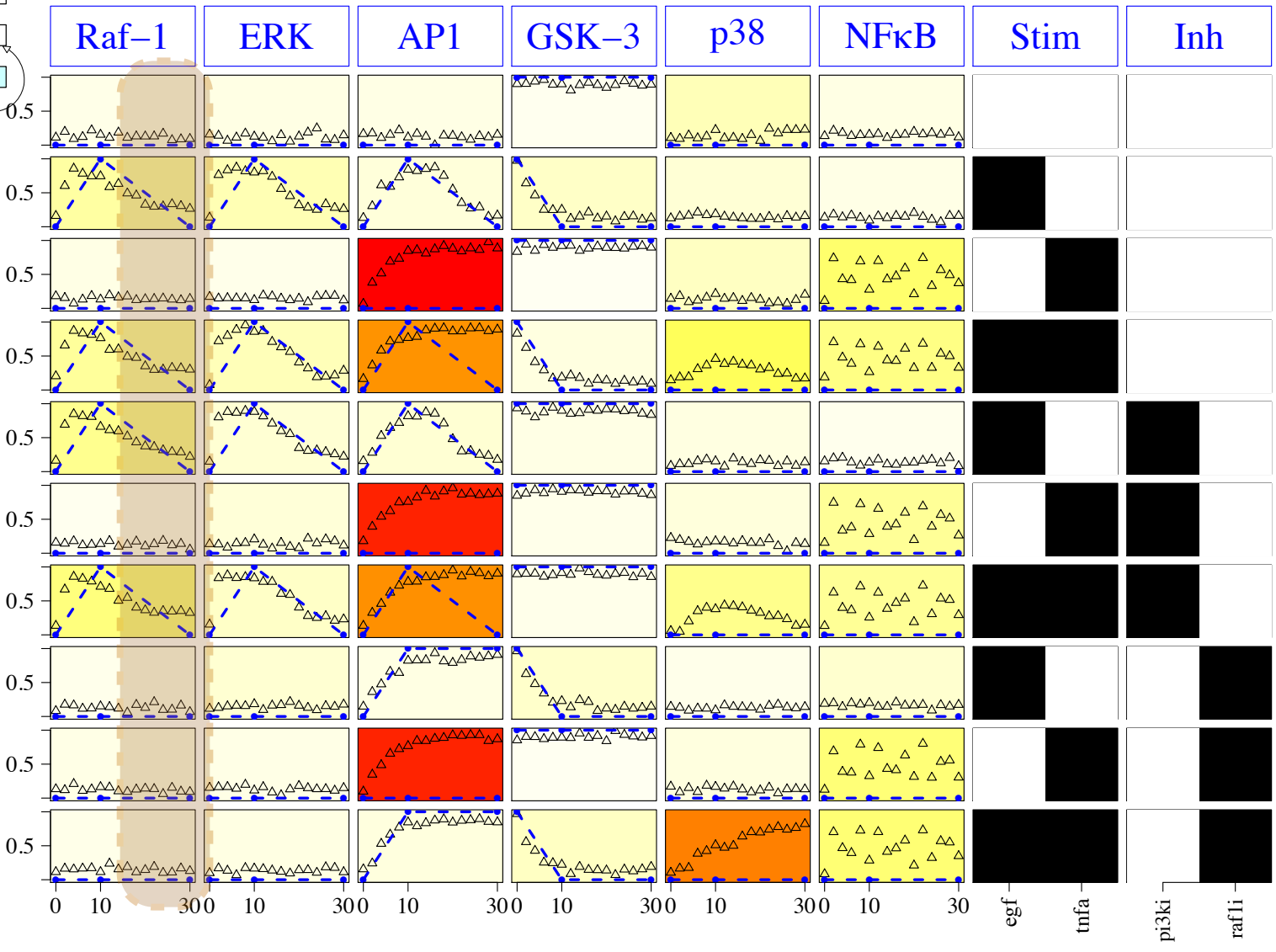


How can we model feedback effects?

MacNamara A Terfve C Henriques D Penalver B Saez-Rodriguez J *submitted*

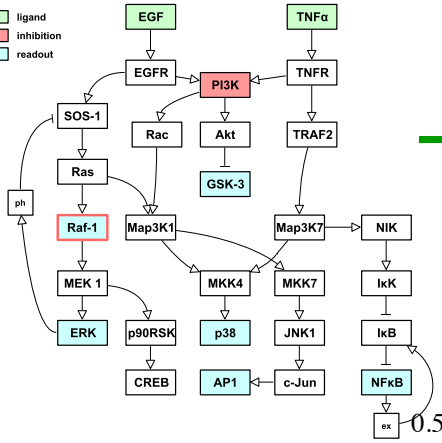
Multiple pseudo-steady-state simulations capture feedback effects that lead to transient signals

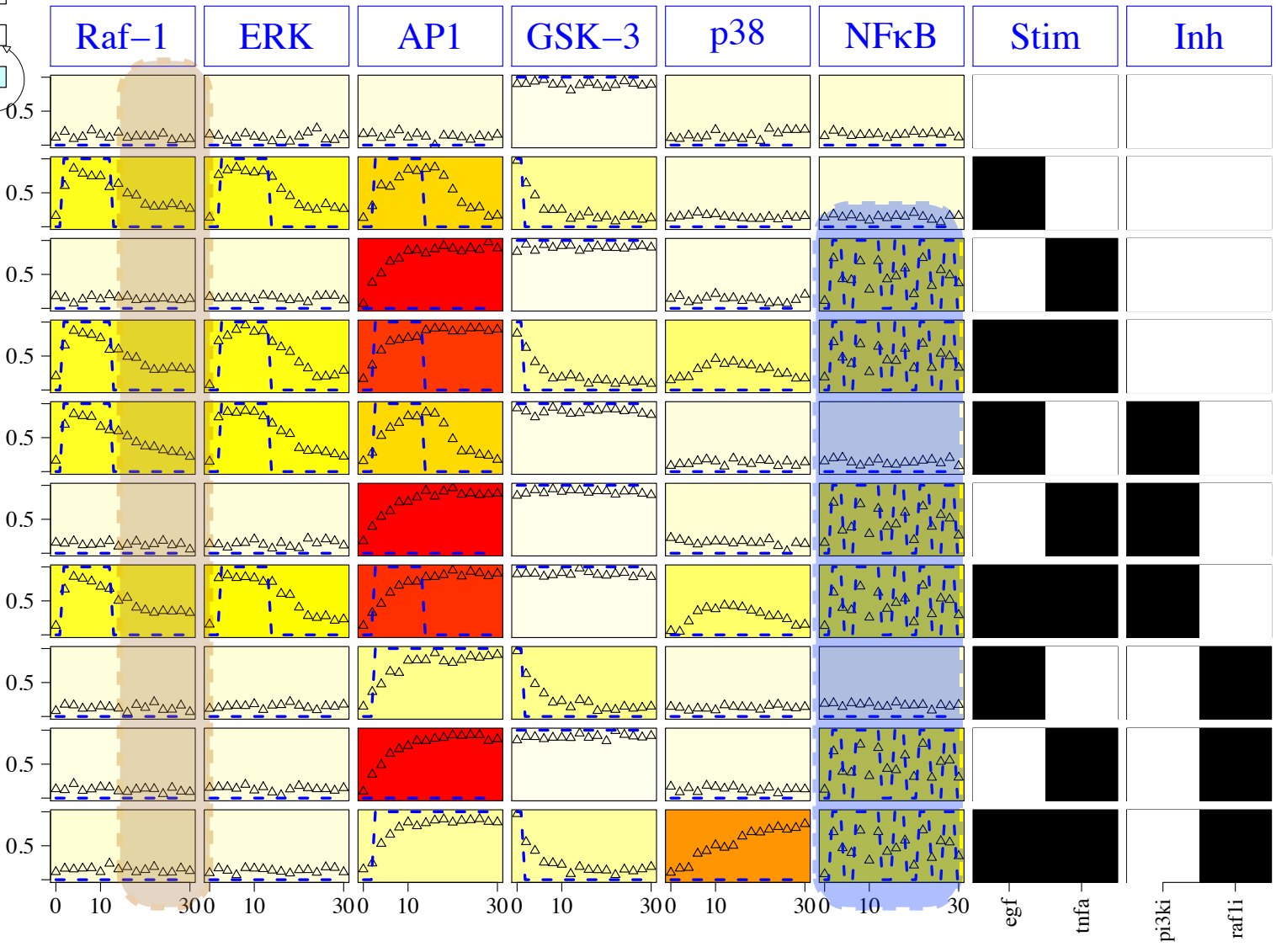




MacNamara A Terfve C Henriques D Penalver B Saez-Rodriguez J *submitted*

Synchronous simulation captures oscillatory behaviour

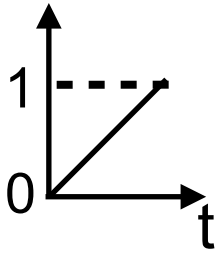




MacNamara A Terfve C Henriques D Penalver B Saez-Rodriguez J *submitted*

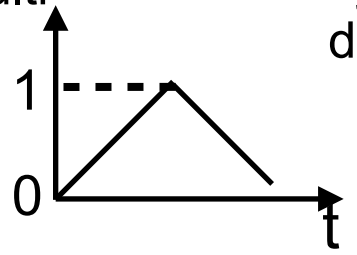
From Boolean to continuous and dynamic models within *CellNOpt*

Boolean (binary)
logic steady state

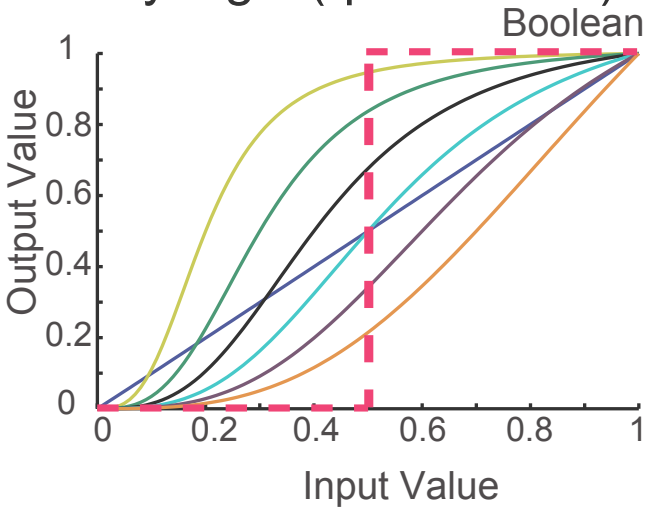


Boolean multi
time-scale

sync.
dynamics



Fuzzy logic (quantitative)



Morris et al., PloS Comp Bio 2011

Logic-based ODEs

- Convert Boolean update function *Bi* into a *continuous homologue* $\overline{Bi}$ using multivariate polynomial interpolation (Odefy: Wittman et al)
  - **Accuracy** (same behavior as *Bi* for 0/1
    → same monotony & steady state behavior)
  - Good **analytical** properties (smoothness)
  - **Minimal** and **unique**

- Make non linear replacing variable with Hill function

$$f(\bar{x}_i) = \frac{\bar{x}_i^{\,n}}{(\bar{x}_i^{\,n} + k^n)}$$

- Transform into differential equation

$$\bar{x}_i(t+1) = \bar{B}_i(\bar{x}_{i1}(t), \bar{x}_{i2}(t), \ldots, \bar{x}_{iN_i}(t)) \rightarrow \dot{\bar{x}}_i = \frac{1}{\tau_i} \cdot (\bar{B}_i(\bar{x}_{i1}, \bar{x}_{i2}, \ldots, \bar{x}_{iN}) - \bar{x}_i)$$

- E.g. a AND b inactivate C

$$\frac{d}{dt}c = \frac{1}{\tau}\left(\frac{a^{n_a} * (1 + k_a^{\,n_a}) * (1 - b^{n_b}) * (1 + k_b^{\,n_b})}{(a^{n_a} + k_a^{\,n_a}) * (b^{n_b} + k_b^{\,n_b})} + \frac{(1 - a^{n_a}) * (1 + k_a^{\,n_a}) * b^{n_b} * (1 + k_b^{\,n_b})}{(a^{n_a} + k_a^{\,n_a}) * (b^{n_b} + k_b^{\,n_b})} + \frac{a^{n_a} * (1 + k_a^{\,n_a}) * b^{n_b} * (1 + k_b^{\,n_b})}{(a^{n_a} + k_a^{\,n_a}) * (b^{n_b} + k_b^{\,n_b})} - c\right)$$

Wittman D, Krumsiek J, Saez-Rodriguez J, Lauffenburger DA, Klamt S, Theis FJ, *BMC Sys Bio* 2009

ODEs can be automatically generated from Boolean model (Odefy)

d/dt(tnfa) = 0*(1-tnfa_inh) **%Note that this implies a continuous stimulus**

d/dt(tgfa) = 0*(

d/dt(raf) = ((eg

d/dt(pi3k) = ((e
(1-pi3k_inh)

d/dt(ikb) = ((tn
(pi3k^ikb_n_pi
+ikb_k_tnfa^ik
(1+ikb_k_pi3k
*pi3k^ikb_n_p

d/dt(gsk3) = ((
gsk3_tauinv)*(

d/dt(erk12) = (
ikb^erk12_n_i
(raf^erk12_n_
+erk12_k_ikb^
(1+erk12_k_ra
erk12) * erk12

Problem: even if structure is known need Identify parameters, difficult optimisation problem (collab with J. Banga group)

$$\frac{d}{dt}c = \frac{1}{\tau}\left(\frac{a^{n_a} * (1+k_a{}^{n_a}) * (1-b^{n_b}) * (1+k_b{}^{n_b})}{(a^{n_a}+k_a{}^{n_a}) * (b^{n_b}+k_b{}^{n_b})} + \frac{(1-a^{n_a}) * (1+k_a{}^{n_a}) * b^{n_b} * (1+k_b{}^{n_b})}{(a^{n_a}+k_a{}^{n_a}) * (b^{n_b}+k_b{}^{n_b})} + \frac{a^{n_a} * (1+k_a{}^{n_a}) * b^{n_b} * (1+k_b{}^{n_b})}{(a^{n_a}+k_a{}^{n_a}) * (b^{n_b}+k_b{}^{n_b})} - c\right)$$

_inh)

_tauinv)*

af)*

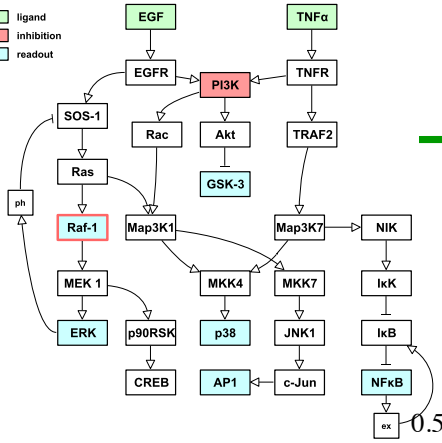
_ikb)-

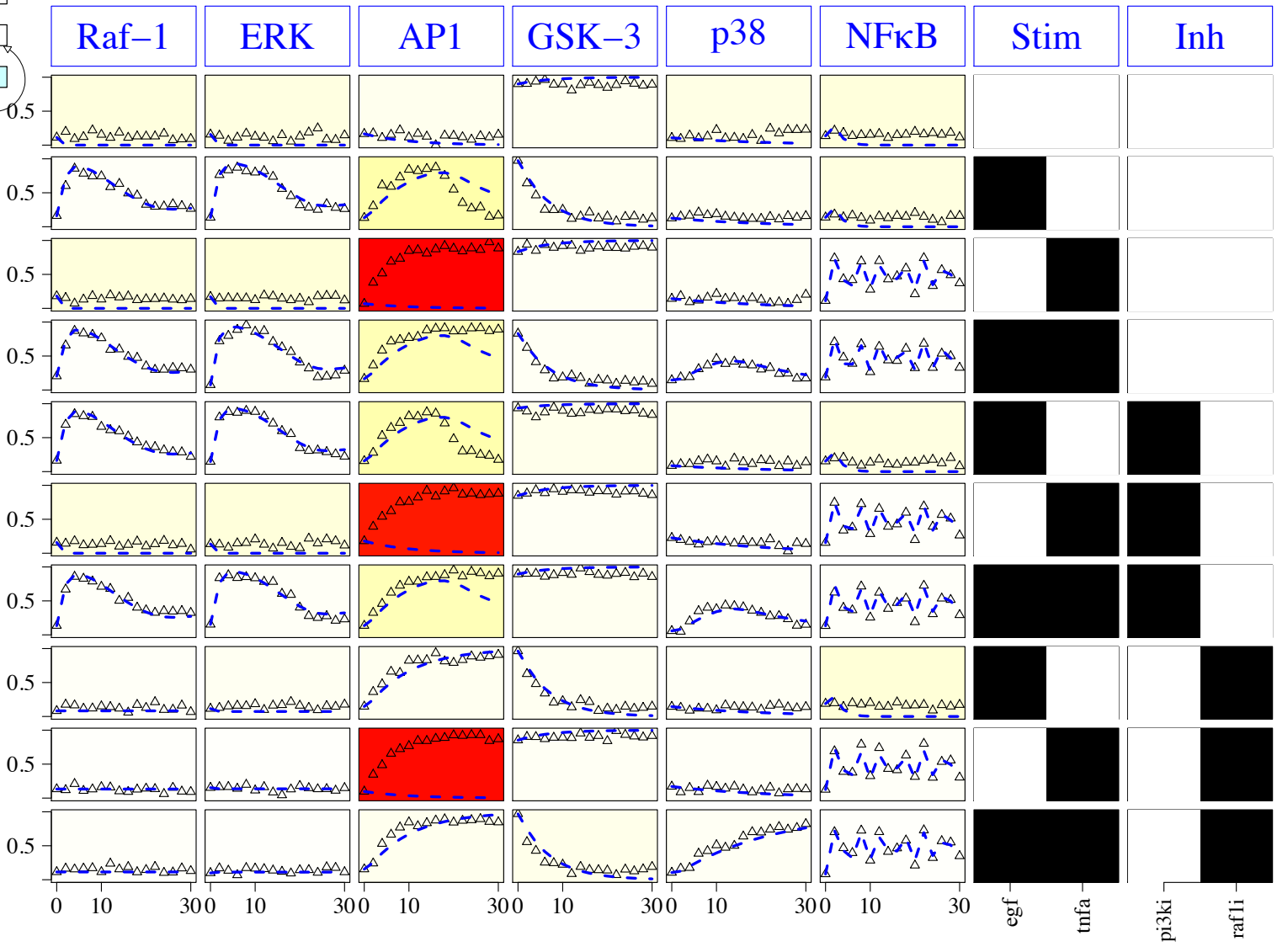
d/dt(egfr) = ((tgfa^egfr_n_tgfa/(tgfa^egfr_n_tgfa+egfr_k_tgfa^egfr_n_tgfa)*(1+egfr_k_tgfa^egfr_n_tgfa)-egfr) * egfr_tauinv)*(1-egfr_inh)

d/dt(casp8) = ((tnfa^casp8_n_tnfa/(tnfa^casp8_n_tnfa+casp8_k_tnfa^casp8_n_tnfa)*(1+casp8_k_tnfa^casp8_n_tnfa)-casp8) * casp8_tauinv)*(1-casp8_inh)

d/dt(akt) = ((pi3k^akt_n_pi3k/(pi3k^akt_n_pi3k+akt_k_pi3k^akt_n_pi3k)*(1+akt_k_pi3k^akt_n_pi3k)-akt) * akt_tauinv)*(1-akt_inh)

Fit of ODE model

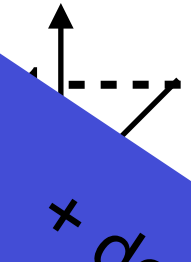




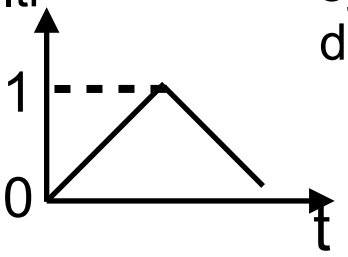
MacNamara A Terfve C Henriques D Penalver B Saez-Rodriguez J
submitted

From Boolean to continuous and dynamic models within *CellNOpt*

Boolean (bina… logic steady



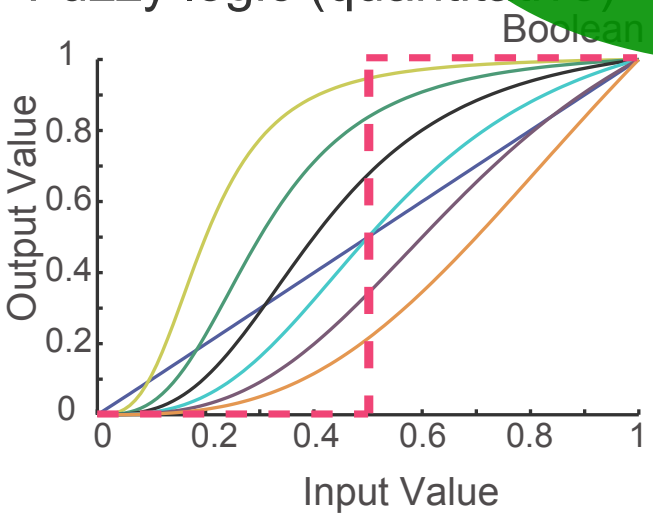
Boolean multi time-scale sync. dynamics



+ detail

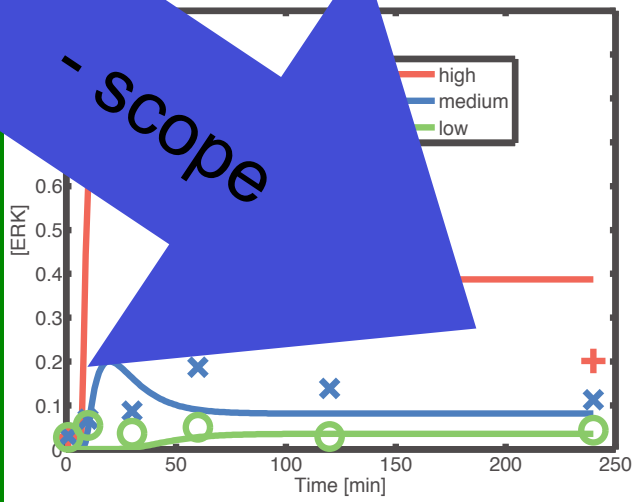
- scope

Fuzzy logic (quantitative)



Morris et al., PloS Comp Bio 2011

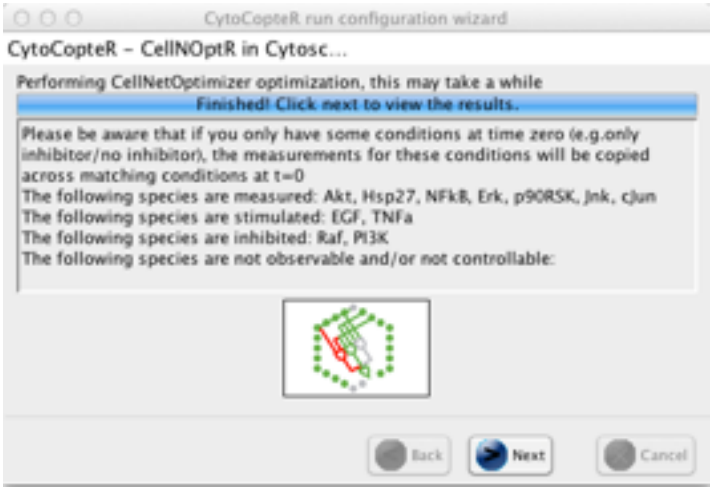
…ogic ODEs (dynamic)

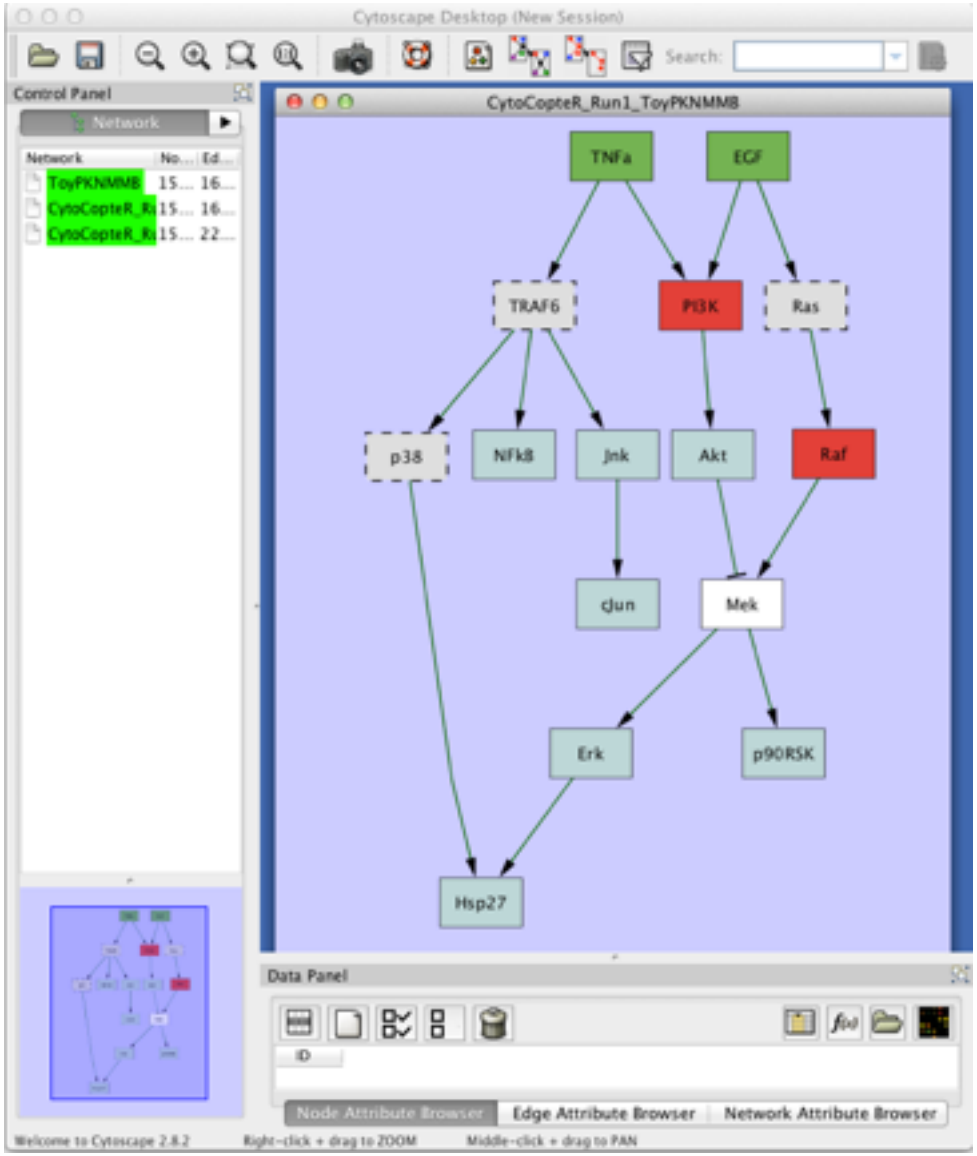


w. J Banga & J. Egea, B. Penalver

David Henriques

Cytoscape plugin (Cytocopter)





Martijn van Iersel

Emanuel Gonçalves

Explore sources of prior knowledge

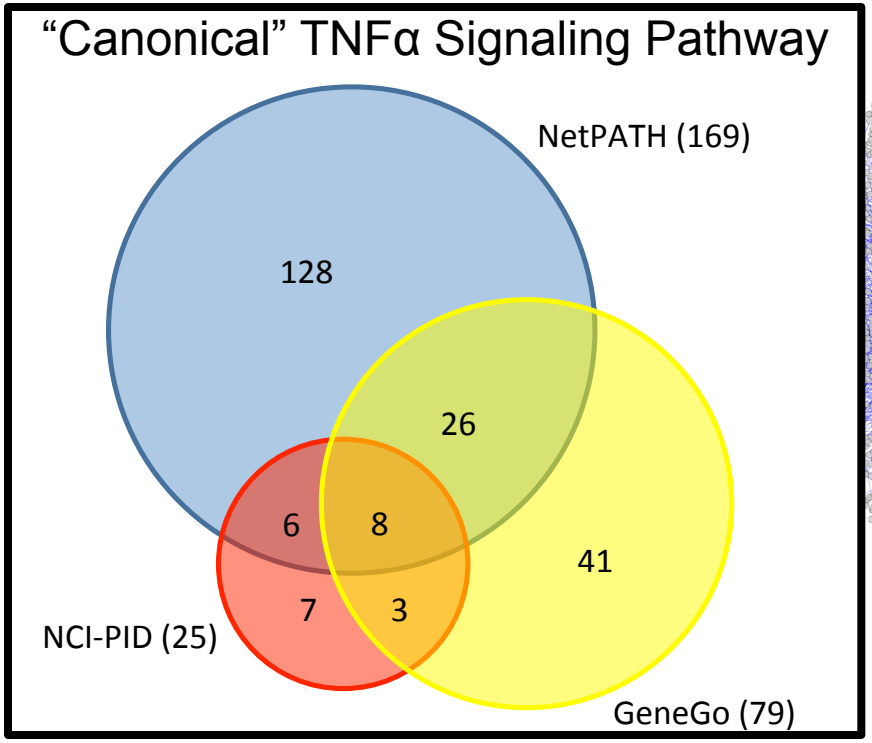
- **Databases** of curated **pathways** (Reactome, KEGG, Wikipathways, ...) incomplete, low overlap, different qualities
➜ Path2Models (standarized pathway resouces; w. LeNovere)

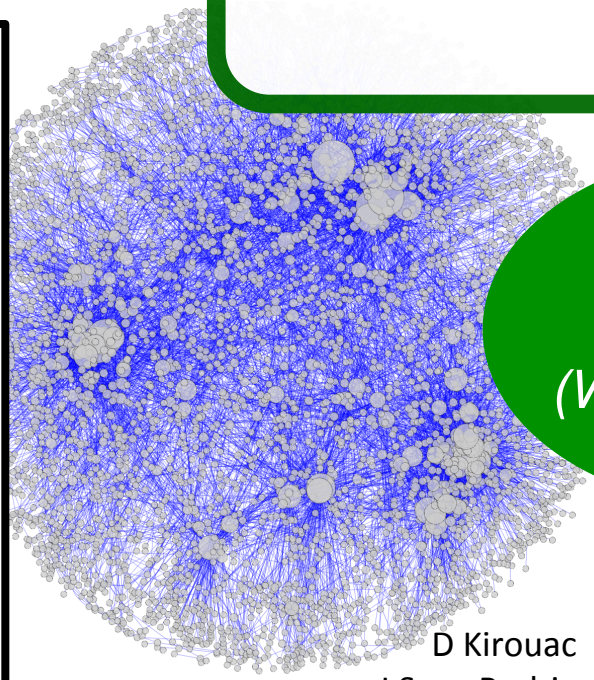
- Protein **Interaction** Networks

- Literature mining

Link CellNOpt
to methods
to infer new links

Federica Eduati





collaborative efforts (WikiPathways)?

D Kirouac
J Saez-Rodriguez
et al, *submitted*

Martijn van Iersel

Acknowledgments

Systems Biomedicine group
www.ebi.ac.uk/saezrodriguez

EMBL EIPOD
EBI-Sanger ESPOD
EU-7FP BiopreDyn
Sanofi-Aventis

Camille Terfve Francesco Iorio Michael Menden Aidan MacNamara Martijn van Iersel Thomas Cokelaer Clare Pacini Federica Eduati Emanuel Gonçalves

Nicolas Le Novère (EBI)

Claudine Chaouya (Gulbenkian)

Jose Egea(U. Cartagena)
David Henriques Julio Banga (CSIC)

Beatriz Penalver (Northwest. Uni)

Jeremy Muhlich Jona Epperlein
Ming Zhang Peter Sorger (HMS)

Melody Morris Alexander Mitsos
Doug Lauffenburger (MIT)

Ioannis Melas Leo Alexopoulos(NTUA)

Regina Samaga Steffen Klamt (MPI)