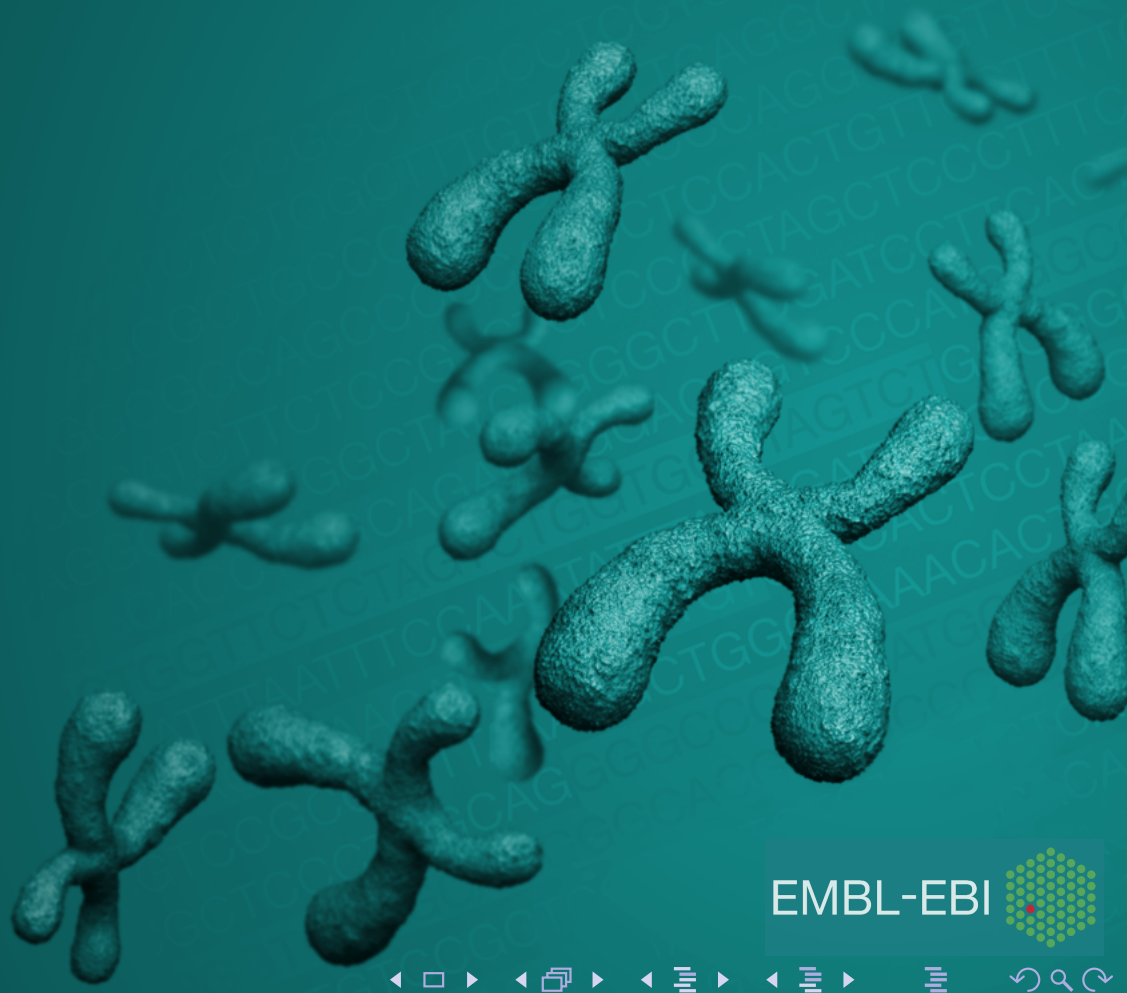
CellNOpt: a brief overview

Thomas Cokelaer
Aidan MacNamara
Julio Saez-Rodriguez

European Bioinformatics Institute
Hinxton (Cambridge) U.K
www.ebi.ac.uk/saezrodriguez



Lausanne, Switzerland, April 2014

EMBL-EBI

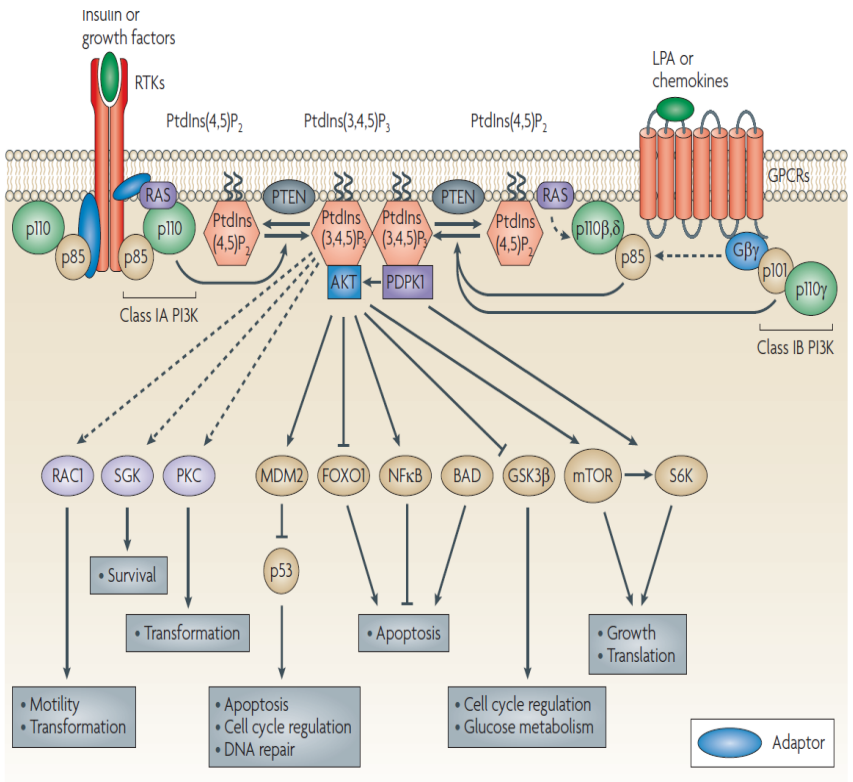
Outline

Outline

Deregulation of signalling networks

Context

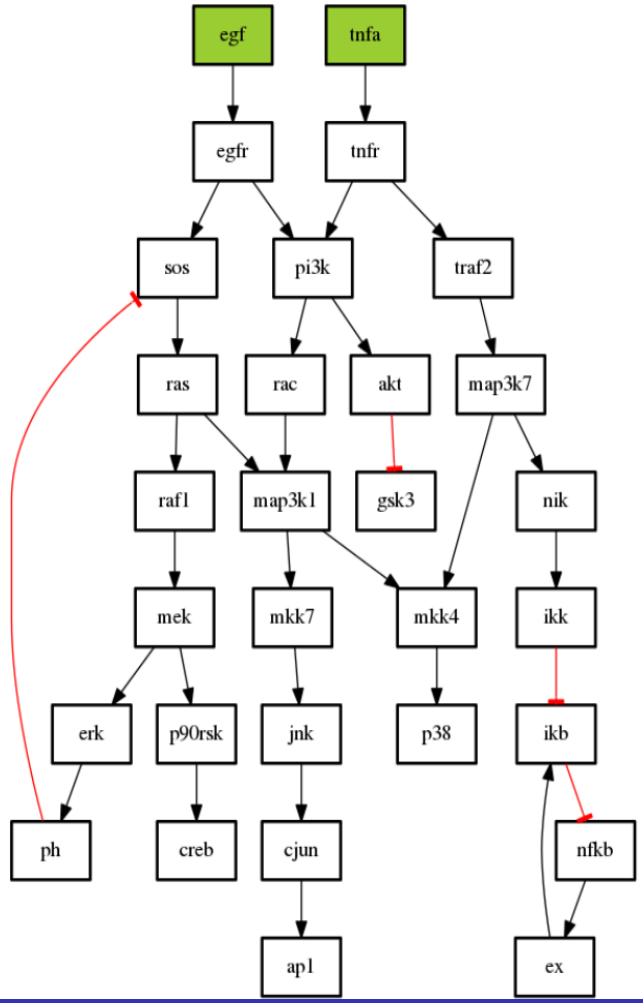
- Genomic alteration in cancers manifest themselves by altering the wiring and functionality of signalling networks.
- There are lots of resources that provide network that describe interactions between proteins.
- Networks are useful to explore signalling pathways but in general not cell-type specific and not computable models. May also have missing links.



Understand signalling networks

Why

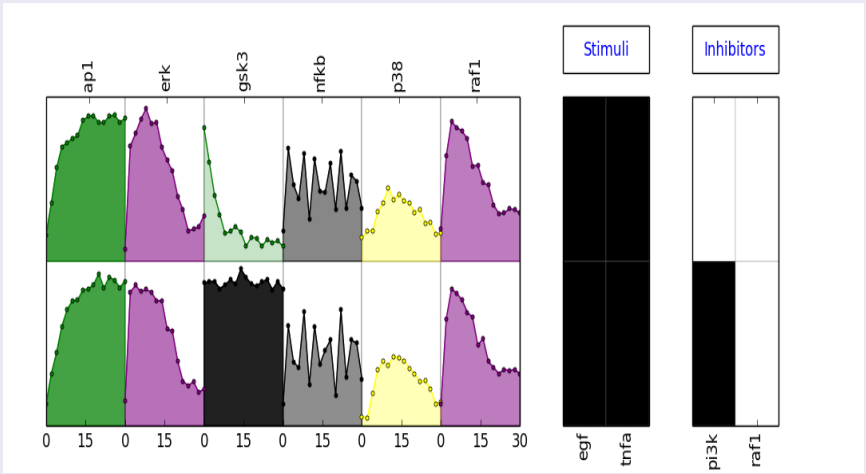
- Most therapeutic affect elements of these networks.
- Understanding the deregulation of signalling networks in cancer is a useful approach to understand drug mode of action and identify novel therapeutic opportunities.

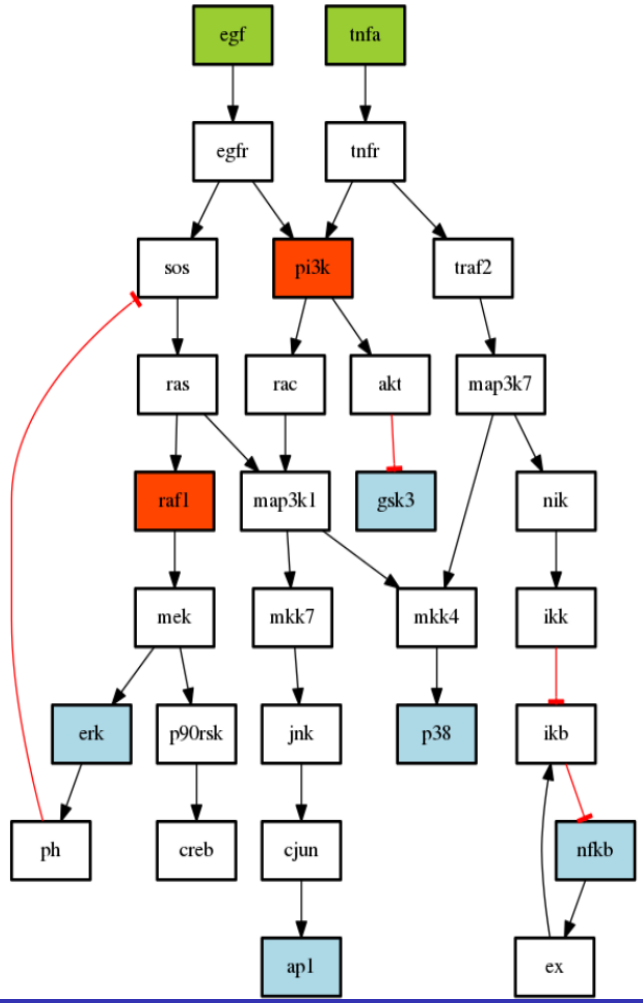


Data perturbation

How can we link large protein networks to signaling data (i.e., biochemical measurements upon perturbation) to create cell-specific explanatory and predictive models ?

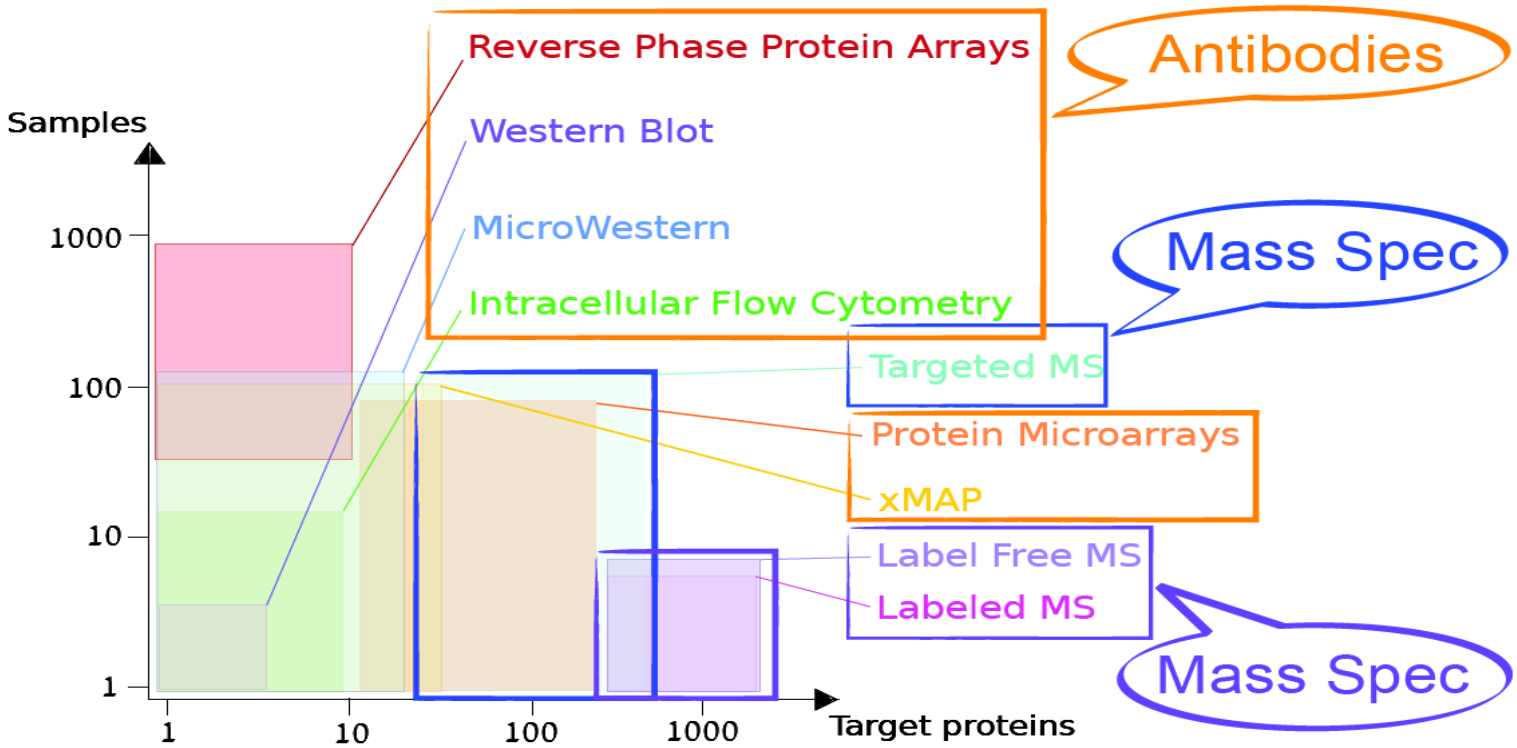
Example of perturbation data.





Data origin and complexity

To model signaling, proteomics is the natural source of data, but there are many platforms



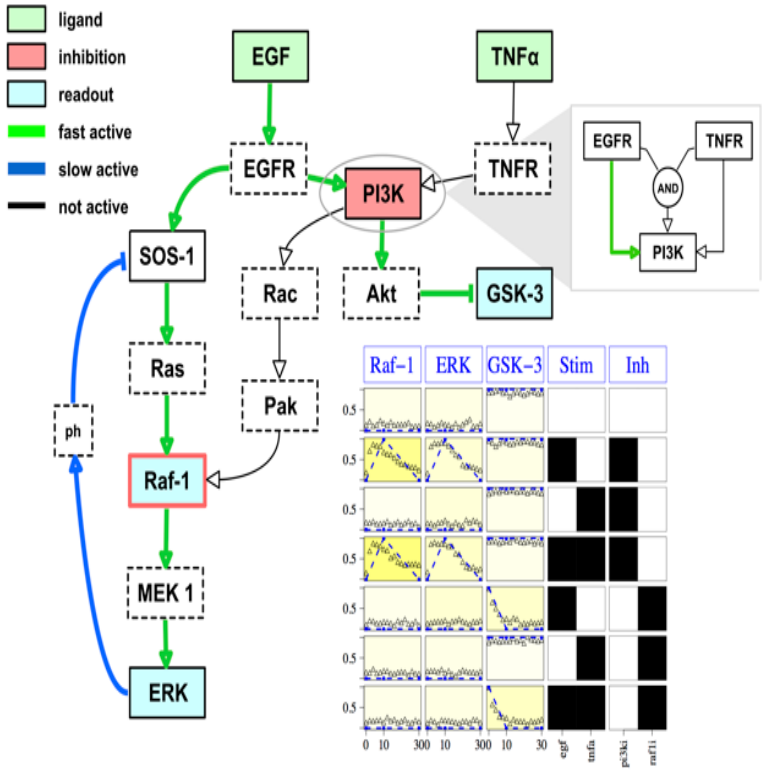
Terfve, C., Saez-Rodriguez, J. 2012, Modeling Signaling Networks Using High-throughput Phospho-proteomics. Advances in experimental medicine and biology.

Outline

CellNOpt

Goals

- Build a logic modelling to link protein signalling network with functional analysis of signal transduction. (i.e., training models to experimental data)
- Identify active/inative links and logic gates (AND/OR)
- Provide different mathematical logic formalisms to cover wide spectrum of modeling approaches (boolean, fuzzy, ode, ...)

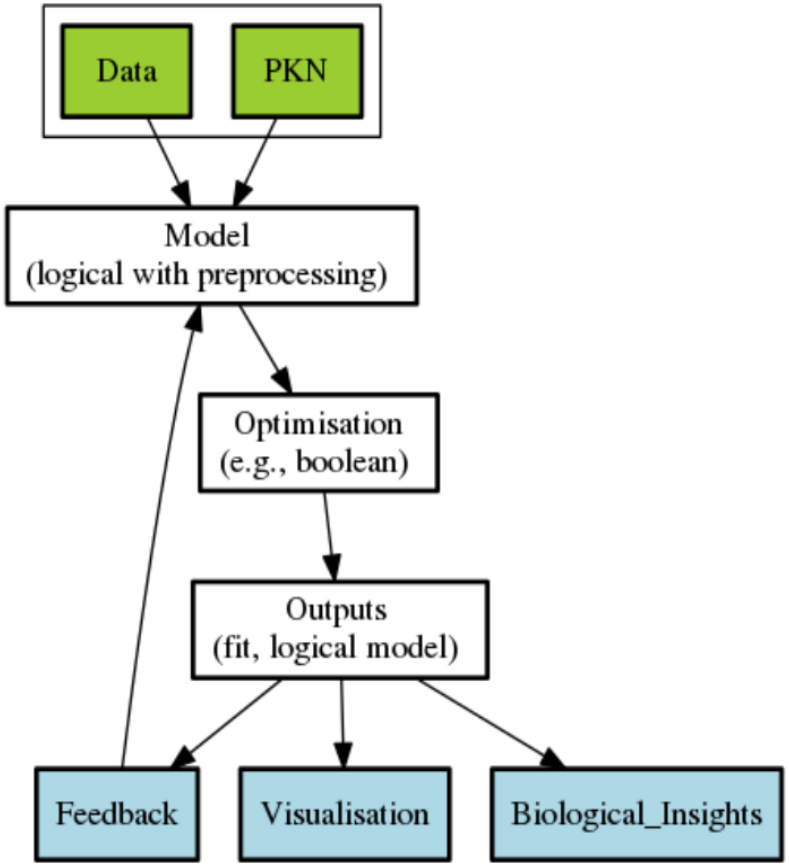


- MacNamara A, Terfve C, Henriques D, Penalver B, Saez-Rodriguez J, Phys Biol 9 045003, 2012
- Terfve C, Cokelaer T, MacNamara A, Henriques D, Gonçalves E, Morris MK, van Iersel M, Lauffenburger DA, Saez-Rodriguez J, BMC Syst Bio, 6:133, 2012

Typical pipeline

Pipeline flow

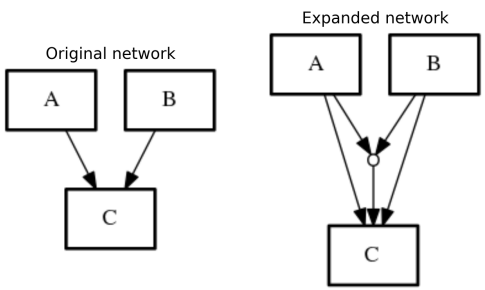
- Data is in MIDAS format. We have lots of non-official **codecs** for Luminex, Mass-Spec, RPPA formats.
- Inputs Prior Knowledge Networks are in SIF/SBMLqual. Output of the optimisation can be in SIF/SBMLqual.
- Optimisation can use various formalisms (boolean, ode... see followinf slides)
- Visualisation of the PKN/MIDAS, Fitness versus experiments.

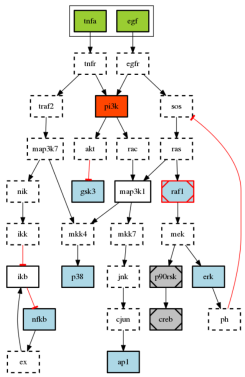


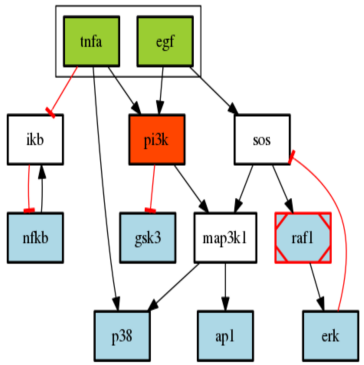
Chaouiya et al, SBML qualitative models: a model representation format and infrastructure to foster interactions between qualitative modelling formalisms and tools, BMS Systems Biology 2013, 7:135

preprocessing

- Prior Knowledge Network must be transformed into a computable model.
- Logical gates (AND) should be added assuming existing edges are ORs.
- Nodes/Species that are not measured could be compressed to simplify the model.
- A modelling approach should be chosen to accomodate the complexity of the data. See next slides
- An optimisation approach (e.g. Boolean) should be chosen depending on the modelling approach
- Finally, interpretation and feedback are required to improve the original Network

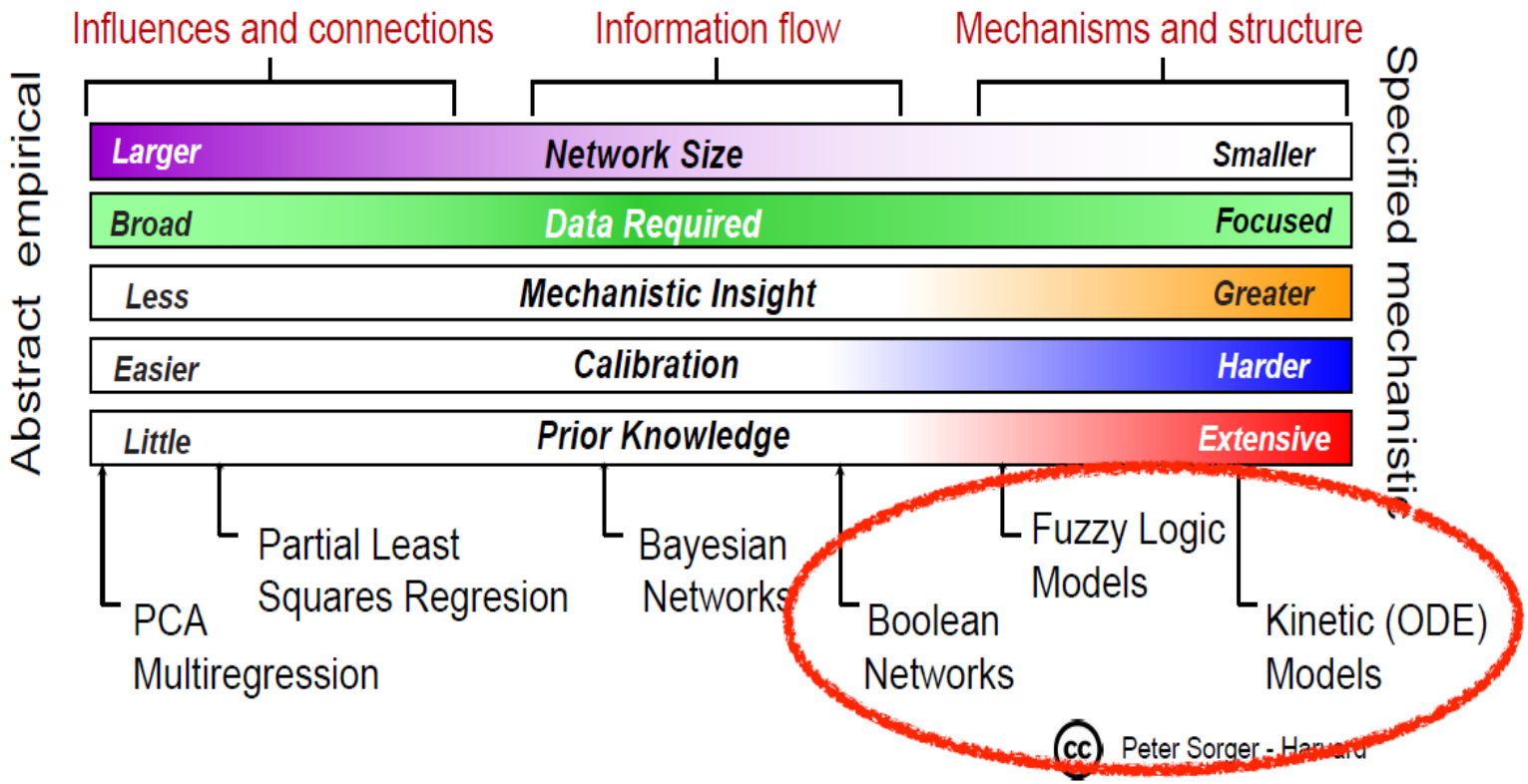






Modelling Approaches

Choice of method depends on the biological question, the prior network knowledge, the data (modeler's expertise) The choice can be seen as more an art than science



Boolean formalism (steady state)

CellNOptR, boolean case, fitness

The Boolean formalism implements a steady state simulator given a logic model with AND/OR logical gates. The metric to evaluate a model given the data is

$$\theta = \theta_f + \alpha.\theta_s$$

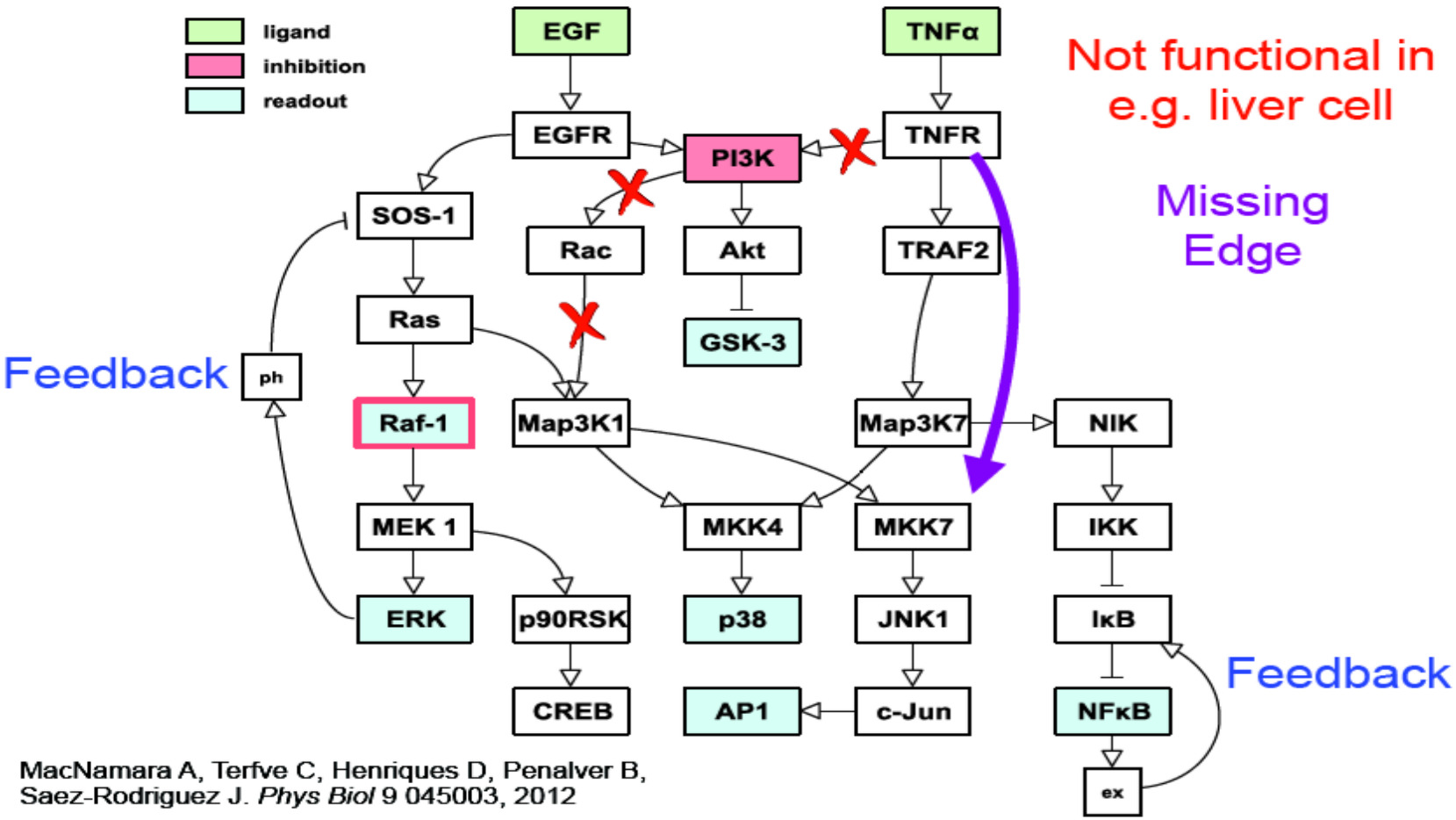
where

$$\theta_f = \sum_{i=1}^{S}\sum_{j=1}^{E}\left(X_{i,j} - \hat{X}_{i,j}\right)^2$$

is the fit of the data, θ_s the model size N and α the relative importance of the fit versus size, $X_{i,j}$ the measurements and $\hat{X}_{i,j}$ the simulated values.
Model parameters are the active/inactive edges.
Optimisation is performed via Genetic Algorithm.

Saez-Rodriguez, J. et al., 2009. Discrete logic modelling as a means to link protein signalling networks with functional analysis of mammalian signal transduction. Mol Syst Biol.

A toy model

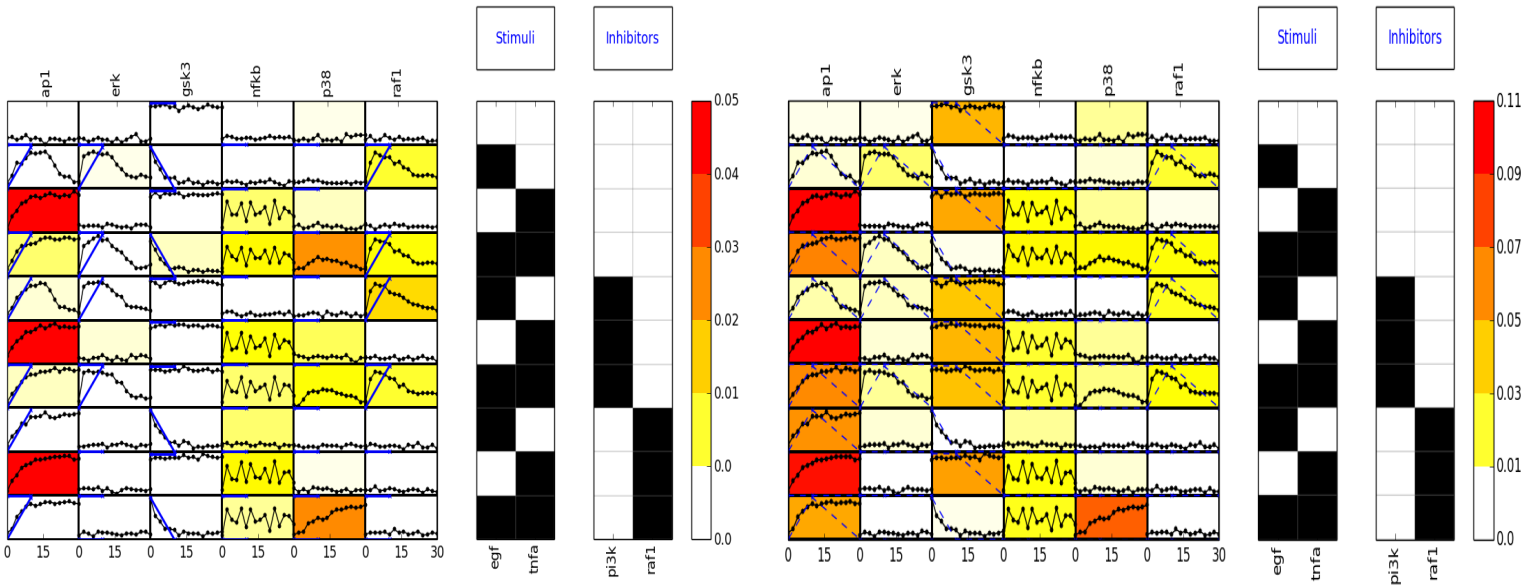


MacNamara A, Terfve C, Henriques D, Penalver B, Saez-Rodriguez J. *Phys Biol* 9 045003, 2012

Boolean formalism 1 and 2 steady state

In CellNOpt, the Boolean steady state approach can identify :

- strong active links (or strongly inactive); see e.g., LHS figure
- strong feedback effects when using the 2 steady states formalism (RHS figure)



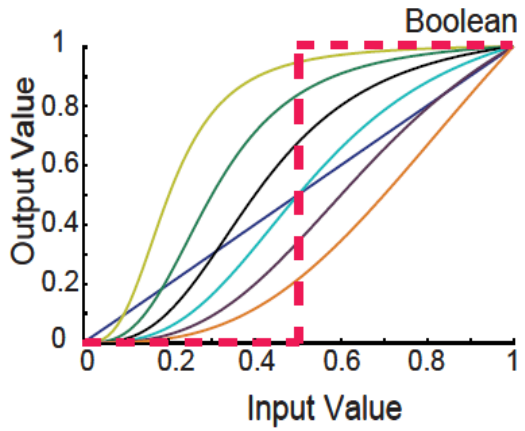
Terfve, C, Cokelaer T, MacNamara A, Henriques D, Gonçalves E, Morris MK, van Iersel M, Lauffenburger DA, Saez-Rodriguez J, BMC Syst Biol, 6:133, 2012

Discrete Fuzzy formalism

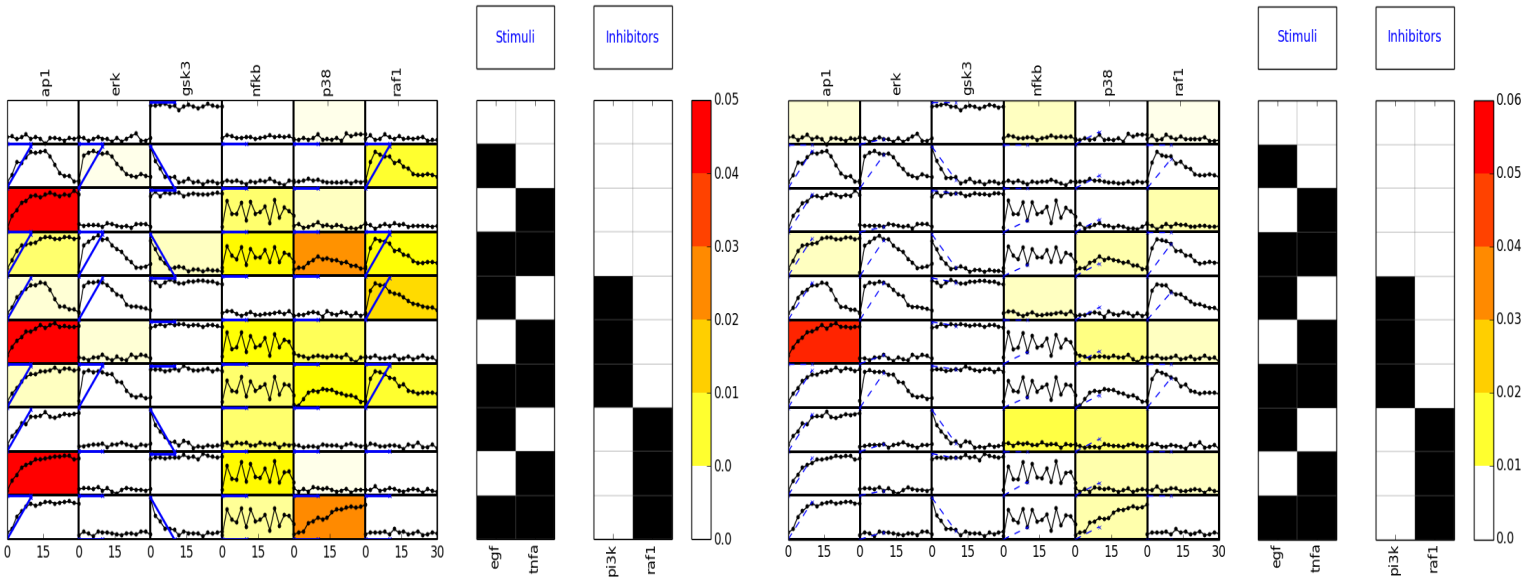
Boolean modeling can not describe quantitative aspects, unlike the Fuzzy approach that can identify :

- strong **and weak** active links
- feedback loops cannot be identified.

- 2 parameters per edge to express the transfer function.
- discrete variables to speed up optimisation.



Discrete Fuzzy formalism



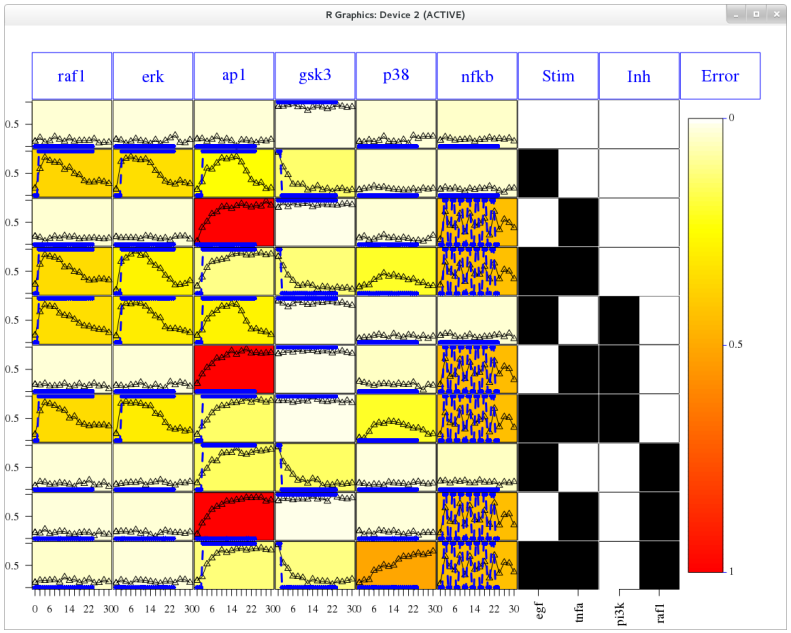
Morris MK, Saez-Rodriguez J, Clarke D, Sorger PK, Lauffenburger DA. PLoS Comp Bio 2011

Extends boolean to continuous time

Why not using Boolean at each time points ? This is the discrete-time formalism.

- feedback loops can be identified.
- issue: more parameters so optimisation is more difficult

Main difference with steady state boolean approach is that updates are synchronous.



MacNamara A, Terfve C, Henriques D, Penalver B, Saez-Rodriguez J. Phys Biol 9 045003, 2012

logic-based ODE approach

Convert Boolean update function into a continuous homologue.

- strong and weak active links and feedback loops can be identified.
- issue: more parameters so optimisation is more difficult

- Convert the Boolean update function *Bi* into a *continuous homologue* $\overline{Bi}$ using multivariate polynomial interpolation
  - **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 $f(\overline{x_i}) = \frac{\overline{x_i}^n}{(\overline{x_i}^n + k^n)}$
- Transform into differential equation

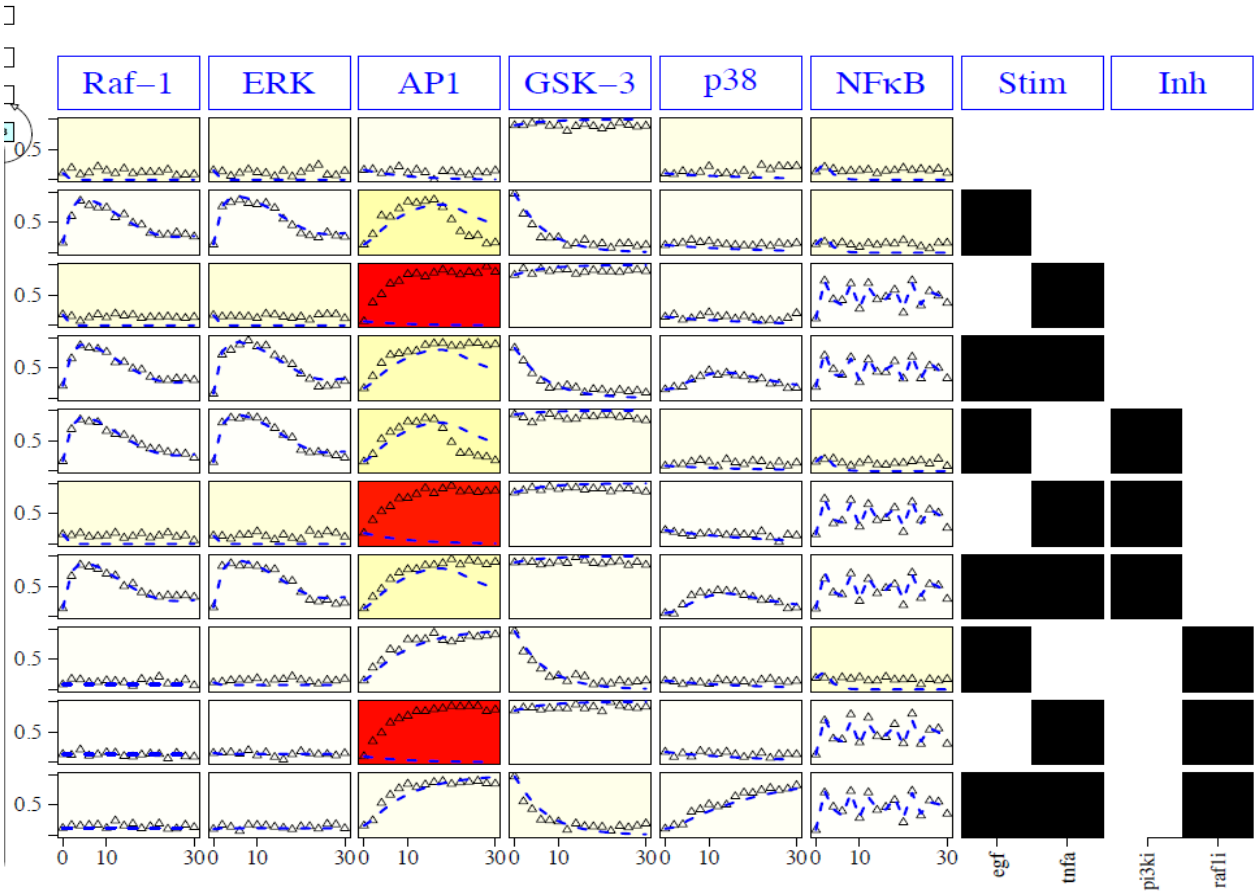
$$\overline{x}_i(t+1) = \overline{B}_i(\overline{x}_{i1}(t), \overline{x}_{i2}(t), \ldots, \overline{x}_{iN_i}(t)). \rightarrow \dot{\overline{x}}_i = \frac{1}{\tau_i} \cdot (\overline{B}_i(\overline{x}_{i1}, \overline{x}_{i2}, \ldots, \overline{x}_{iN}) - \overline{x}_i)$$

- **Example:** A AND B inactivate C

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

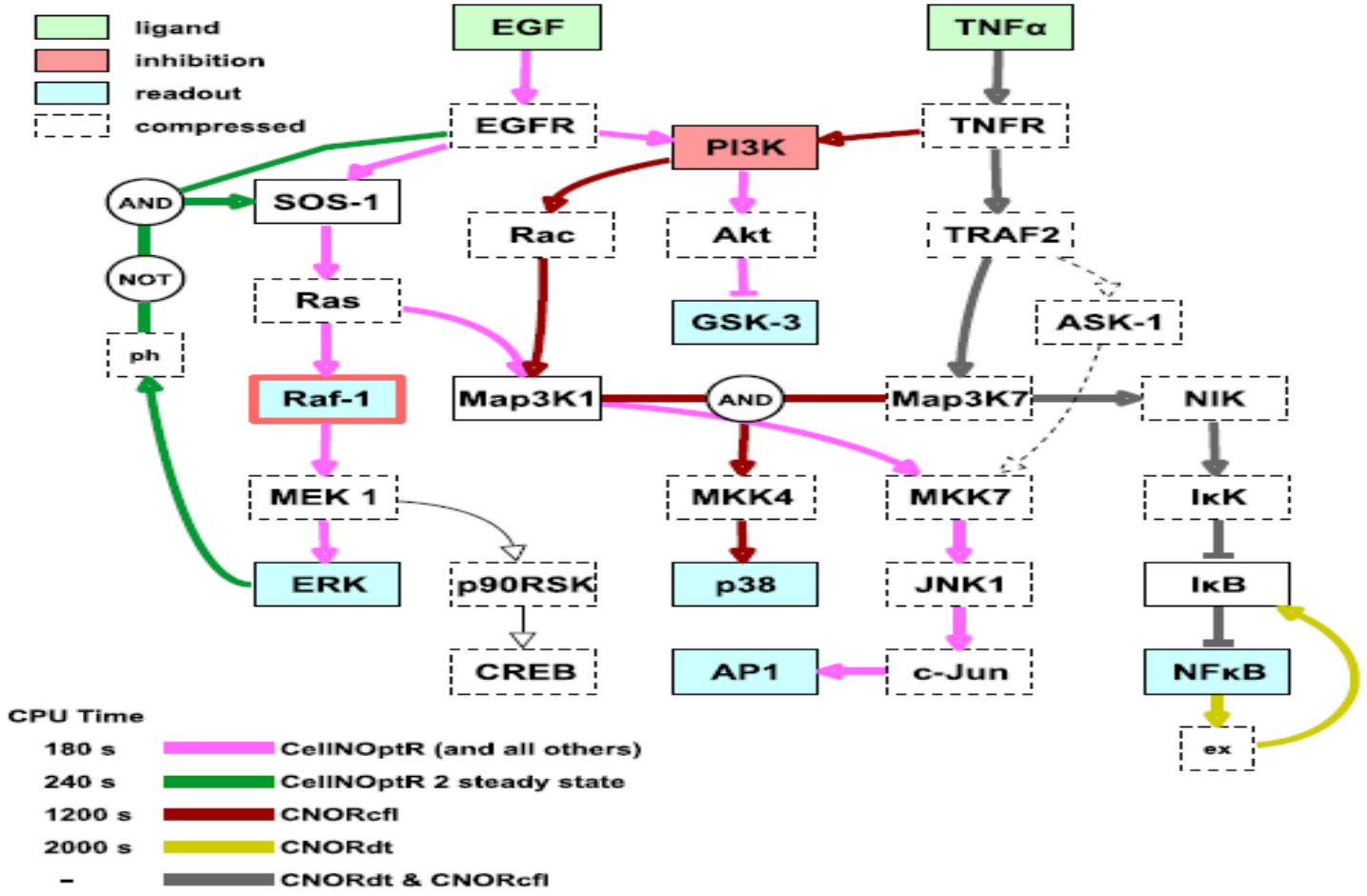
Wittman D, Krumsiek J, Saez-Rodriguez J, Lauffenburger DA, Klamt S, Theis FJ, *BMC Sys Bio* 2009

logic-based ODE approach



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

Summary of different approaches



MacNamara A, Terfve C, Henriques D, Penalver B, Saez-Rodriguez J. Phys Biol 9 045003, 2012

Other modules/aspects

- Feedback can be model using the ODE formalism but also the **discrete-time** (boolean + synchronous + all time points) . – *MacNamara A, Terfve C, Henriques D, Penalver B, Saez-Rodriguez J. Phys Biol 9 045003, 2012*
- CNORfeer (finds missing links) – *Eduati F, de las Rivas J, di Camilo B, Toffolo G, Saez-Rodriguez J Bioinformatics 10.1093/bts363, 2012*
- ASP (answer set programming) – *Carito Guziolowski, Santiago Videla, Federica Eduati Sven Thiele, Thomas Cokelaer, Anne Siegel, Julio Saez-Rodriguez, Exhaustively characterizing feasible logic models of a signaling network using Answer Set Programming* Bioinformatics (2013) 29 (18) 2320-2326
- Multi-Cell lines – *Saez-Rodriguez J, Alexopoulos LG, Zheng M, Lauffenburger DA, Sorger PK, Cancer Research 71(16) 1-12, 2011*
- Optimisation with MEIGOR tool (Metaheuristics for global optimization in systems biology and bioinformatics) – *Egea, J, Henriques, D, Cokelaer T, Villaverde A, Banga JR, Saez-Rodriguez J in preparation*

Outline

Availability

CellNOptR is a suite of R packages

Official releases every 6 months on BioConductor.

- CellNOptR : main IO functionalities + boolean formalism
- CNORode: the ODE optimisation/simulator
- CNORfuzzy: the discrete-fuzzy optimisation/simulator
- CNORdt: the discrete-time optimisation/simulator
- CNORfeeder: finds missing links

30,000 LOC (R code + manual), 5,000 LOC in C, + R tutorials.

GUI

CytoCopter
a Cytoscape plugin http://apps.cytoscape.org/apps/cytocopter
Author: Emanuel Gonçalves see also: http://www.cellnopt.org/cytocopter/index.html

Visibility

Website

- Hosted at http:// www.cellnopt.org
- Provide access to sets of PKN/MIDAS files published or that serve as toy examples.
- links to official releases (BioConductor) and intermediate releases
- mailing lists

Python packages

- the PKN/MIDAS files are also available independently as a Python package called **cellnopt.data** available on Pypi website.
- Access to the R packages in Python is possible via wrappers called **cellnopt.wrapper**.

Outline

Conclusions

CellNOpt Status

- One main R package called CellNOptR
- Provides common functionalities (e.g., IO such as MIDAS, SIF, SBMLqual)
- Various formalisms are available in CellNOptR or add-on packages (boolean, discrete time, ODE, boolean, fuzzy)
- Mailing lists, download pages, tutorials (www.cellnopt.org)

Future directions

- More tools for Parameter Estimation and Model fitting.
- Input Pathways (infered or web resources)
- Automatic name convertion to standard protein databases (e.g. UniProt)
- GUI
- Interoperability with other simulators.

More information

Please visit www.cellnopt.org website if you want to know more about CellNOpt.

Acknowlegments

All CellNOpt(R) developers, in particular:

- Federica Eduati (missing links with CNORfeeder)
- Emanuel Gonçalves (CytoCopteR)
- David Henriques (ODE package)
- Aidan MacNamara (Discrete Time)
- Melody Morris (Fuzzy package)
- Camille Terfve (CellNotpR boolean steady states)